The Paradox of Information Access: On Modeling Polarization in the Age of Information

Chao Xu, Jinyang Li, Dachun Sun, Jinning Li, Tarek Abdelzaher, Jesse Graham, Michael Macy, Christian Lebiere, and Boleslaw Szymanski

***Abstract*— The paper derives a new nonlinear stochastic model of evolution of human beliefs that demonstrates how an increase in democratized information production and sharing, combined with consumers' confirmation bias and natural bias for outlying content, result in increased polarization. The model shows that the evolution of human beliefs can be approximated by a nonlinear diffusion-drift equation in which systematic psychological biases contribute to *drift*, whereas other random influences contribute to *diffusion*. The nonlinear formulation predicts a growth in *polarization that is attributable to increasing information production and sharing*. While the core contribution is analytical, an anecdotal model parameter fitting to empirical data is also presented. Specifically, we show that our model closely predicts the changing and increasingly polarized distribution of ideology of members of the US Congress over the last quarter century (taken as an approximate proxy for shifts in the US population ideology), when we take the mobile phone penetration curve as a proxy for democratization of information access. The model suggests that escaping the polarizing forces in the age of information access may be an uphill battle.**

***Index Terms*— Social networks; dynamic models; polarization; paradox of information access.**

I. Introduction

In this paper, we ask the question: how do increasing information production and sharing relate to societal polarization? A model is derived that shows that human beliefs follow a diffusion-drift equation in which ingrained systematic psychological biases contribute to belief drift, whereas other random factors and influences contribute to diffusion. The diffusion-drift equation predicts a steady-state belief distribution in which *increased access to information production and sharing contributes to increased levels of polarization*. The extent of this effect depends on the relative strength of drift versus diffusion terms. Anecdotal empirical evidence is presented that at least some societies may indeed be operating in a regime consistent with a non-trivial information-access-facilitated polarization growth. Specifically, for the US, the model accurately predicts the growing polarization of the US Congress, taking as input the technology penetration curve for mobile phones (as a proxy for democratized information access and sharing) in the last 25 years.

The work is motivated by the historic change in information access patterns in the 21st century. Over the course of most of human history, information *broadcast* has been prohibitively expensive. It required significant investments (e.g., having a radio station or a publishing house). With the invention of the Internet, the barrier to making content available for potentially global consumption was significantly reduced. We say that "information broadcast" (both access and sharing) has become *democratized*. While the benefits of democratizing information broadcast are undeniable, it is interesting to model the impact of this change on societal polarization (as such models are a prerequisite to the design of proper mitigation policies for any undesirable side-effects).

The idea that increased access can facilitate polarization is not new. For example, evidence suggests that the interstate highway system in the US may have contributed to socio-economic disparity and geographic polarization in metropolitan areas [1]. Highways allowed individuals to live further away from where they worked, facilitating urban sprawl, and allowing communities to self-segregate into more homogeneous and separated geographic neighborhoods (in an analogy with social echo-chambers) of significantly different character.

Are similar forces at play in the case of *information* access? Our drift and diffusion terms roughly correspond to the idea of "thinking fast and slow" [2], referring to the tension between our fast, intuitive reactions and more deliberate thinking (known in psychology as "system 1" and "system 2"). Fitting empirical data to the model sheds light on the regime that a society operates in, regarding the balance between the two.

Figure 1 notionally demonstrates the paradox of access. Given a bias for content that already agrees with one's prior beliefs [3], [4], when the globally available information

Resubmitted on Jan 27, 2023. This work was conducted in part under DARPA award HR001121C0165, and in part under DoD Basic Research Office award HQ00342110002.

Chao Xu is with the Kavli Institute for Theoretical Sciences, University of Chinese Academy of Sciences, Beijing (e-mail: chx035@ucsd.edu).
Jinyang Li is with the Department of Computer Science, University of Illinois at Urbana Champaign (e-mail: jinyang7@illinois.edu).
Dachun Sun is with the Department of Computer Science, University of Illinois at Urbana Champaign (e-mail: dsun18@illinois.edu).
Jinning Li is with the Department of Computer Science, University of Illinois at Urbana Champaign (e-mail: jinning4@illinois.edu).
Tarek Abdelzaher is with the Department of Computer Science, University of Illinois at Urbana Champaign (e-mail: zaher@illinois.edu).
Jesse Graham is with the Eccles School of Business, University of Utah (e-mail: jesse.graham@eccles.utah.edu).
Michael Macy is with the Department of Sociology, Cornell University (e-mail: mwm14@cornell.edu).
Christian Lebiere is with the School of Computer Science, Carnegie Mellon University (e-mail: cl@cmu.edu).
Boleslaw Szymanski is with the Department of Computer Science, Rensselaer Polytechnic Institute (e-mail: szymab@rpi.edu).

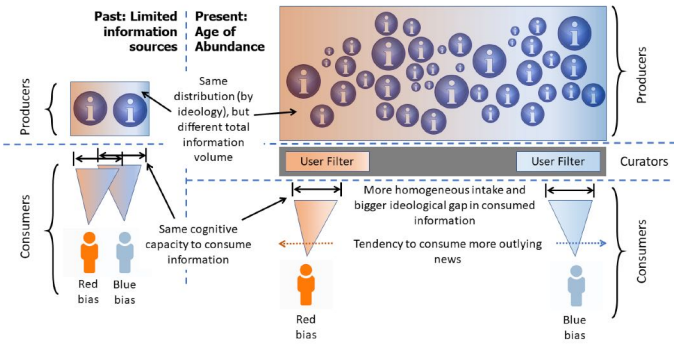


Fig. 1: A notional illustration of the paradox of information access, showing the impact of information volume on polarization under the constraints of fixed cognitive capacity, confirmation bias, and preference for outlying content.

sources are plentiful, the attained *coverage* by an individual (specifically, coverage of alternative ideological views, perused under cognitive capacity constraints) is small. Bias for outlying content [5]–[8] (and indeed competition on producing it as a means to capture attention) leaves consumers with biased views of reality that inch gradually at normalizing extremes, driving consumers away from the center. Larger ideological gaps emerge between information consumed by different parties, leading to polarization. To prevent circular reasoning, where we conclude what we assumed, this paper starts with well-agreed-upon biases that are well-validated in social psychology and proves the implications on emergent belief distribution. It also postulates a model with *tunable weights* that can be empirically estimated, thus remaining valid over a wide spectrum of extents to which social and psychological biases drive behavior. We acknowledge that the accuracy of the model is generally difficult to validate without further large-scale studies that may constitute opportunities for future work. In addition, algorithms for closed-loop control (for mitigation) are delegated to future work.

The rest of this paper is organized as follows. Section II reviews related work. Section III introduces our model and assumptions on human biases, borrowed from social psychology. Section IV derives the resulting overall nonlinear dynamics, solves for the equilibrium belief distribution under this model, and proves key model properties, including the growth of polarization with volume. Section V provides anecdotal evidence of model validity based on empirical observations. Section VI discusses implications and future work. The paper concludes with Section VII.

II. Related Work

The paper addresses a challenge brought about by the broad problem of modern information overload, discussed by the authors in earlier work [9]. A two-page extended abstract, published in an earlier workshop [10], framed the polarization analysis problem addressed in this work. For earlier arXiv preprints with partial results, please see [11], [12].

While neural-network models might conceivably do better at matching empirical data, a main advantage of our model is that it is derived from first principles and is kept intentionally minimal in its assumptions. Thus, it has a potentially greater explanatory value. The model leverages the well-known fact that homophilly [13] and confirmation bias [3] cause individuals to gravitate to other like-minded sources, an effect that may be exacerbated by a tendency to prioritize exploitation over exploration [14]. Prior work also demonstrated the attention-capturing effects of more outlying content [5]–[8]. We show that the above behaviors, if followed, will create polarization that increases with the degree of content access and sharing. By espousing simplicity, we hope the stylized model leads to better generalizability, as fewer assumptions are needed to produce the effect.

From a methodology perspective, the work reported in this paper falls in the general area of *continuous opinion dynamics*. Our contribution lies in adapting the methodologies used for the study of (continuous) opinion dynamics to the problem of modeling the impact of content volume on the emergence of polarization in the age of democratized content production and access. Work on opinion dynamics has roots in consensus formation models, such as DeGroot's pooling scheme [15] and its extensions [16], as well as the Friedkin-Johnsen model [17]–[19]. The latter significantly advanced the study of polarization by incorporating exogenous influence or bias into opinion formation, thereby producing steady state behaviors where stable disagreements emerge. Notable extensions include (i) the notion of bounded confidence [4], [20] that postulates that agents influence each other only when sufficiently close in the belief space, (ii) the related notion of stubborn agents [21] whose opinion does not change over time, and (iii) extensions to multidimensional opinions [22], [23]. While these models generally assume that agents move towards a weighted average of beliefs of their influencers (typically their neighbors in the belief space), some models also introduce a notion of random noise. For example, the noisy Hegselmann-Krause model [24] assumes that, besides converging on the weighted average of their neighbors, individuals experience random jumps. The magnitude of allowed jumps was shown to have a significant impact on steady state system behavior [24].

Special attention was also paid (in continuous opinion dynamics literature) to the way *interactions among agents* influence their beliefs. Of particular interest was to explain the emergence of *extremism*. For example, the Deffuant-Weisbuch model [25] extended the bounded confidence models [4], [20] by postulating that the influence of contacts between neighboring agents in the belief space is weighted by the confidence of the source. Thus, according to that model, an extremist might have a higher influence than a moderate, which is a different way of framing influence of more extreme content. An alternative explanation of the disproportionate influence of extremism comes from nonlinear extensions of the Friedkin-Johnsen models, where source *susceptibility* to opinion change is set proportionally to their degree of conviction [26]. Thus, agents with more stubborn beliefs are less susceptible to change, making those beliefs more "sticky".

The continuous model studied in this paper is novel in investigating the impact of information volume. We show that an increase in volume contributes to increased polariza-

tion when coupled with bias for more extreme or outlying content [5]–[8]. Once upon a time, this bias gave humans an evolutionary advantage [6]; those who paid preferential attention to news of trouble (e.g., "wolves at the gate") were better able to survive the threat. We derive the population belief distribution prompted by such content biases, reducing it to a well-known diffusion-drift equation for which we can study the equilibrium [27].

Opinion dynamics are also studied under the framework of mean-field game (MFG) theory [28]–[30], which is a non-linear model of constrained Brownian motion of many agents while the control strategy to each agent is governed by Hamilton-Jacobi dynamics (or equivalently Euler-Lagrange dynamics). The non-linearity of such theory comes from the control process to each agent. MFG is also applied to study the polarization of opinion distribution [31], where a reward mechanism along with cost functions is designed, and polarization occurs in some cases. In our model, a novel non-linearity is included by introducing a content influence weight function. We then study the impact of information overload.

Finally, we acknowledge the parallel literature on *discrete* opinion models, where individuals are forced to choose between *finite alternatives* (as opposed to gradually evolving their beliefs on a continuum in the belief space). An example is the Naming Game (NG) model where a person converges to a single opinion from repeated interactions with neighbors [32]. Committed agents who consistently proselytize their opinion (and are immune to influence) were introduced into the Naming Game and shown to create a tipping point in opinion spread [33]; with a single small community of committed agents, the majority opinion in a population is rapidly replaced by the opinion of a small fraction once this fraction reaches the tipping point ($\approx 10\%$). This work was extended to two competing committed agent groups [34], and its steady-state behavior analyzed under different conditions [35]. Additional research has focused on empowering agents with individualized behavior [36], and considering properties of the networks [37] connecting them.

III. A Model of Modern Belief Dynamics

Let set, $\mathcal{X}$, denote a society of individuals in an age of democratized information access and sharing, where the total population size $|\mathcal{X}| = N$. Democratized information access and sharing (in our model) means that all content published by anyone in $\mathcal{X}$ is accessible to everyone in $\mathcal{X}$. Thus, the key factor influencing the topology of actual information transfer is *consumer choices* (as opposed to say, physical travel distance between producer and consumer). These choices depend on individual beliefs. Let the position of each individual $a_i \in \mathcal{X}$ at time t be represented by a scalar value $x_i(t) \in \mathbb{R}$ (where $\mathbb{R}$ is the set of real numbers). The origin of the real-valued axis represents neutrality. Deviations from the origin in either direction represent ideological bias, such as political left versus right, or such as conservative versus liberal outlook. Note that, while some political systems represent political positions in a two dimensional space, such as the Swiss system,[1] we restrict our discussion to a single axis. We assume that individuals produce content that reflects their positions. Thus, $x_i(t)$ refers to both the position espoused by an individual a_i and the position espoused by content they produce at time, t. Below, we describe the belief dynamics model addressed in this paper. We call it a *belief influence-field model* because we show that our ingrained content consumption biases (shaped by social psychology) act as a *force field* that causes drift in our ideological positions (beliefs) towards increased polarization, the way physical forces cause drift of impacted particles. In the meantime, other factors (such as sampling random content outside of our ideological neighborhood) contribute to diffusion that mitigates the drift. The model studies how the relative strengths of the two effects impact the relation between information volume and polarization.

A. Bounded Confidence

We assume that an individual, a_i, of position $x_i(t)$ at time t, will engage *in part* with a subset of sources, $\mathcal{X}^{(i)}(t) \subset \mathcal{X}$, that match the individual's own belief; a phenomenon known in opinion dynamics literature as *bounded confidence* [4]. Bounded confidence can be thought of as a manifestation of *confirmation bias* [3]; opinions supporting one's belief are acted upon, whereas those far from one's belief are ignored. Note that, since sources in $\mathcal{X}$ are (by definition) *globally accessible*, the subset $\mathcal{X}^{(i)}(t)$ depends only on ideological positions (non-withstanding any geographic or other boundaries). Accordingly, for each consumer, a_i, we assume that an ideological visibility radius, ϵ_i, determines how ideologically distant the neighbors they engage with might be, when acting under the influence of bounded confidence. Thus:

$$\mathcal{X}^{(i)}(t) = \{a_j || x_i(t) - x_j(t)| \leq \epsilon_i\} \tag{1}$$

In an age of democratized sharing and access, it is easy to find enough like-minded individuals who match one's own beliefs very closely. Thus, we assume that ϵ_i is small compared to the entire range of beliefs being represented in society. This allows linearization-based approximations within radius, ϵ_i. We call $\mathcal{X}^{(i)}(t)$ consumer a_i's *neighborhood set*.[2] Note that although the entire belief space is the range of real numbers, the majority of beliefs fall within an interval. Exceptionally extreme beliefs fall into the long tail in the distribution which is only a negligible small fraction.

B. Random Influence

Some cultural values (such as tolerance, inclusion, diversity, and "worldliness"), as well as elements of random chance and acts of exploration in lieu of exploitation, introduce an additional component of influence in belief dynamics that breaks out of confirmation bias and bounded confidence. Since, by definition, this component captures factors that are orthogonal and free of bias, we model its effect by a random

[1] https://en.wikipedia.org/wiki/List_of_political_parties_in_Switzerland

[2] One can argue that the ideological radius, ϵ_i, of an individual's neighborhood set will continue to decrease with increased information production/access, since an individual of finite cognitive capacity will not be able to consume the increasing amount of content generated within a fixed ideological radius, ϵ_i. The paper proves a stronger result that depends only on ϵ_i being small, and does not require it to continue to decrease with increased information production/access.

walk, or *Brownian motion*, scaled by some coefficient, σ, proportional to the strength of such additional influences.

C. Belief Updates

As a result of an individual's interactions with others, the individual's beliefs are updated. The considerations described above suggest a belief update form similar to the Friedkin and Johnsen model [38]. Namely, an individual of position $x_i(t)$ in the belief space at time, t, will move to position $x_i(t+\Delta t)$ at time $t+\Delta t$, given by:

$$x_i(t+\Delta t) = (1-\alpha_{\Delta t})x_i(t) + \alpha_{\Delta t} f(\mathcal{X}^{(i)}(t)) + \sigma \Delta W$$

or, rearranging:

$$x_i(t+\Delta t) = x_i(t) + \alpha_{\Delta t}(f(\mathcal{X}^{(i)}(t)) - x_i(t)) + \sigma \Delta W\,. \quad (2)$$

The above equation features three components:

- *Susceptibility to local influence*, $\alpha_{\Delta t}$: The value $\alpha_{\Delta t}$ is a constant, $0 \leq \alpha_{\Delta t} \leq 1$, that represents how susceptible individuals generally are to external influence due to content in their neighborhood sets (in other words, $1-\alpha_{\Delta t}$ is essentially stubbornness). A larger $\alpha_{\Delta t}$ leads to a larger belief update during time Δt. Our stylized model views susceptibility as a global parameter in order to facilitate the study of society at large. While, in reality, different individuals can have different susceptibility values, it is not our goal to study a specific population mix. Thus, we defer such individualization to future work.
- *Center of bias*, $f(\mathcal{X}^{(i)}(t))$: This is the center of gravity of systematic forces describing the influence of those in one's neighborhood set on one's beliefs.
- *Other influences*: Finally, σ is a constant that scales the influence of remaining factors on the belief update. For example, marriages, relocation, ideological exploration, new job environments, and cultural values emphasizing new experiences might influence evolution of one's beliefs. These events might have independent causes and thus are modeled by a stochastic term, ΔW. Specifically, ΔW, is the integral of white noise over Δt, which is given by a Wiener process (i.e., Brownian motion).

D. The Components of Influence

The steady state distribution of beliefs in the above model depends, in part, on the shape of the function $f(\mathcal{X}^{(i)}(t))$, representing the influence of sources in one's belief neighborhood on their beliefs. Below, we elaborate two key factors that shape this function; namely, bias for outlying content and nonlinear social influence.

1) Consumer Bias for Outlying Content

A key novel element of the model represents the fact that we increasingly seek (and spread) more sensational and surprising news, as confirmed in prior studies [5]–[8]. This asymmetric interest pattern arguably biases our collective attention towards more extreme content. Thus, consistently with prior literature, we take into account that, of the content consumed by an agent, more outlying news have a deeper influence. Specifically, a source in $\mathcal{X}^{(i)}(t)$, that espouses position x in the belief space, has an influence weight, $\eta_0(x) = \eta_0(|x|)$, that generally increases with distance $|x|$ from the (neutral) origin. However, beyond a certain $|x|$, influence decreases again, when the espoused beliefs become "too extreme". Later in Section V, we show an example of $\eta_0(x)$ derived from empirical observations (in Figure 4b).

Although the consumer bias for outlying content is well studied in psychology, its exact form requires further study to determine. Fortunately, the exact equations do not affect our derivation as long as $\eta_0(x)$ reflects a bias for more extreme content and thus has a minimum at the origin and two maxima on the two sides. Eventually, biases drive beliefs as people seek information that reinforces those beliefs as truth [39].

2) Nonlinear Social Influence

Social influence makes a position more desirable if it is adopted more frequently in one's neighborhood set. We model social influence related to position, x, by a nonlinear function of the number of points in the immediate neighborhood of x. Thus, we assume that a consumed source in $\mathcal{X}^{(i)}(t)$, that lies at position, x, in the belief space, has a component of influence that increases with the density of points around x. Let the *density* of sources around position, x, and time, t, be denoted by $\rho_s(x,t)$. We assume that content influence increases with $e^{\kappa\rho_s(x,t)}$. The exponential form has the advantage of approximating a family of polynomial functions, depending on the value of κ. For example, if $\kappa = 0$, the exponential term becomes 1. In this case, each item or source contributes an independent influence regardless of its agreement with other items; the influence of item collections grows *linearly* with collection size. Otherwise, if $\kappa > 0$, the higher the κ, the more rapid the (super-linear) escalation of influence of items (around location x) with density of adoption of x. (As we show later, polarization emerges even with $\kappa = 0$, but it increases with κ.)

3) Putting it together

Taking both (i) bias for outlying content (from Section III-D.1) and (ii) social influence (from Section III-D.2) into account, the influence weight of a source at location x in the belief space, within consumer a_i's neighborhood set, $\mathcal{X}^{(i)}(t)$, is given by:

$$\eta(x,t) = \eta_0(x) e^{\kappa\rho_s(x,t)} \quad (3)$$

We are now ready to define the center of gravity, $f(\mathcal{X}^{(i)}(t))$ in Equation (2). Specifically, the center of gravity of the influence forces exerted by content in the neighborhood set of user a_i at time, t, is given by:

$$f(\mathcal{X}^{(i)}(t)) = \frac{\sum_{a_j \in \mathcal{X}^{(i)}(t)} x_j(t)\ \eta(x_j(t),t)}{\sum_{a_j \in \mathcal{X}^{(i)}(t)} \eta(x_j(t),t)} \quad (4)$$

Upon consuming information from the neighborhood set, per Equation (2), a consumer is attracted towards the (ideological) center of gravity (given by Equation (4)) of the nearby information items, each weighted by their influence upon the consumer. Note that, the weight, $\eta(x_j(t),t)$ favors more extreme positions, per preference for outlying content, thus the center of gravity is skewed towards the extreme. Equation (2) further scales the resulting change in consumer belief (in the direction towards the center of gravity of the neighborhood set)

by consumer susceptibility to local influence, $\alpha_{\Delta t}$, where, as mentioned earlier, $0 \leq \alpha_{\Delta t} \leq 1$.

We can now explain the role of model parameter κ better with the help of Equation (3) and Equation (4). When $\kappa = 0$, the center of gravity, $f(\mathcal{X}^{(i)}(t))$, is simply the weighted sum of all positions in the neighborhood set, weighted by consumer bias for outlying content, $\eta_0(x)$. As κ increases, more popular positions gain a disproportionately larger weight in impacting the center of gravity. For very large κ, the model becomes approximately winner-takes-all (a softmax). The center of gravity gets dominated by the most popular position.

In this paper, we are interested in a situation when the population, $N = |\mathcal{X}|$ is large. By *large* population, we refer to one where summations over finite fractions of members are well-approximated by integrals. In other words, a fluid model applies. Thus, the summation in Equation (4) can be replaced by an integral over source density. The influence, $f(\mathcal{X}^{(i)}(t))$ is estimated by integrating the density of sources (weighted by their influence) over the neighborhood ϵ around the consumer's position, $x_i(t)$. Thus, Equation (3) becomes:

$$f(\mathcal{X}^{(i)}(t)) = \frac{\int_{x_i(t)-\epsilon}^{x_i(t)+\epsilon} x\ \rho_s(x,t)\eta(x,t)\,dx}{\int_{x_i(t)-\epsilon}^{x_i(t)+\epsilon} \rho_s(x,t)\eta(x,t)\,dx} \tag{5}$$

E. Recap: The Belief Influence-Field Model

To summarize, the belief influence-field model discussed in this paper is characterized by Equations (2), (3), and (5), and given by the definition below.

Definition: *The belief influence-field model is a belief update model that postulates that an individual at position, $x_i(t)$, in the belief space will update their beliefs after time Δt according to the following dynamic model:*

$$x_i(t+\Delta t) = x_i(t) + \alpha_{\Delta t}(f(\mathcal{X}^{(i)}(t)) - x_i(t)) + \sigma \Delta W$$

where:

$$f(\mathcal{X}^{(i)}(t)) = \frac{\int_{x_i(t)-\epsilon}^{x_i(t)+\epsilon} x\ \rho_s(x,t)\eta(x,t)\,dx}{\int_{x_i(t)-\epsilon}^{x_i(t)+\epsilon} \rho_s(x,t)\eta(x,t)\,dx}$$

and

$$\eta(x,t) = \eta_0(x) e^{\kappa \rho_s(x,t)}$$

The last equation implies an asymmetry in influence depending on the magnitude of content departure, $|x|$, from neutrality. Note that, the assumptions introduced above (e.g., bounded confidence and confirmation bias) are not a contribution of the authors, but rather are borrowed from prior work on social psychology. Below, we derive some implications.

IV. The Paradox of Information Access

In this section, we compute the steady state density of population beliefs, $\rho(x)$, as a function of position, x, in the belief space. By definition, the integral of population density over the belief space is equal to the total population. Thus:

$$\int_x \rho(x)dx = N \tag{6}$$

We show that, at steady state, population density in the belief space is given by the solution to a diffusion-drift equation that is well-studied in physical systems, with a clear mapping between social model parameters and physical variables.

A. Deriving the Density of Steady-State Beliefs

The following theorem describes the relation between model variables (including number of globally accessible information sources) and the steady state belief distribution.

Theorem 1: In a large population, $\mathcal{X}$, that follows the belief influence-field model, the equilibrium distribution of population density, $\rho(x)$, in the belief space is well-approximated by:

$$\rho(x) = \Lambda \eta_0^{\mu/D}(x) e^{\frac{\mu\kappa\rho(x)}{D}} \tag{7}$$

where Λ is a constant such that $\int_x \rho(x)dx = N$, μ is a constant proportional to susceptibility to local influence, $\lim_{\Delta t \to 0} \frac{\alpha_{\Delta t}}{\Delta t}$, and $D = \sigma^2/2 - \mu$.

Proof: Our proof is inspired by a methodology common in fluid dynamics, where one first models the dynamics of individual particles, then derives properties of populations, such as (dynamics of) overall flow and particle density distribution. Steady state expressions can then be obtained. We will first describe a simplest diffusion-drift mechanism in the belief space and then derive Theorem 1 from there.

Let us examine consumer, a_i. Since their visibility radius ϵ is relatively small, compared to the overall belief space, we can simplify Equation (5) by linearizing the functions $\rho(y,t)$ and $\eta(y,t)$ in the neighborhood of $x_i(t)$. Thus:

$$\rho(x,t) \simeq \rho(x_i,t) + \rho'(x_i(t),t)(x - x_i) \tag{8}$$
$$\eta(x,t) \simeq \eta(x_i,t) + \eta'(x_i(t),t)(x - x_i) \tag{9}$$

where $\eta'(x_i(t),t) = \partial\eta(x,t)/\partial x$ computed at $x = x_i(t)$, and $\rho'(x_i(t),t) = \partial\rho(x,t)/\partial x$ computed at $x = x_i(t)$. Substituting with these linearized expressions into Equation (5), after some simplifications (see Appendix A), we get:

$$f(\mathcal{X}^{(i)}(t)) \simeq x_i(t) + \frac{\epsilon^2}{3}\Big(\frac{\eta'(x_i(t),t)}{\eta(x_i(t),t)} + \frac{\rho'(x_i(t),t)}{\rho(x_i(t),t)}\Big) \tag{10}$$

Recall that the center of gravity is weighted by the influence weight. There are two forces impacting each individual above; one depends on the normalized influence gradient $\eta'(x)/\eta(x)$ that attracts each individual towards more extreme opinions, and the other depends on the normalized population density gradient $\rho'(x)/\rho(x)$ that pushes each individual towards more densely-populated positions. Substituting from the above equation into Eq. (2), the belief update becomes:

$$x_i(t+\Delta t) - x_i(t) \simeq (\alpha_{\Delta t})\frac{\epsilon^2}{3}\Big(\frac{\eta'(x_i(t),t)}{\eta(x_i(t),t)} + \frac{\rho'(x_i(t),t)}{\rho(x_i(t),t)}\Big) + \sigma\Delta W\,. \tag{11}$$

Dividing Equation (11) by Δt and taking the limit of as $\Delta t \to 0$, we get:

$$\frac{dx_i(t)}{dt} \simeq \mu\Big(\frac{\eta'(x_i(t),t)}{\eta(x_i(t),t)} + \frac{\rho'(x_i(t),t)}{\rho(x_i(t),t)}\Big) + \sigma\frac{dW}{dt}\,, \tag{12}$$

where

$$\mu = \frac{\epsilon^2}{3} \lim_{\Delta t \to 0} \frac{\alpha_{\Delta t}}{\Delta t} . \tag{13}$$

Recalling the definition of $\alpha_{\Delta t}$, we call, μ the normalized consumer susceptibility to local influence.

Next, we invoke a well-known result in fluid dynamics that relates the stochastic dynamics of individual particles to the resulting population properties, such as the dynamics of flow and particle density distribution. Specifically, the Fokker-Planck equation of motion (see [27] for a brief introduction) states that if positions, x_i, of particles are governed by a nonlinear stochastic diffusion-drift process of the form:

$$\frac{dx_i}{dt} = \tilde{\mu}(x_i, t) + \sigma \frac{dW}{dt}, \tag{14}$$

then the steady state population density, $\rho(x)$, satisfies the following differential equation:

$$\tilde{\mu}(x,t)\rho(x) - \frac{\sigma^2}{2}\frac{\partial}{\partial x}\rho(x) = 0 \tag{15}$$

(See Appendix B for more detail.) Comparing Equation (12) and Equation (14), we get:

$$\tilde{\mu} = \mu\Big(\frac{\eta'(x_i(t),t)}{\eta(x_i(t),t)} + \frac{\rho'(x_i(t),t)}{\rho(x_i(t),t)}\Big) \tag{16}$$

Substituting with this $\tilde{\mu}$ in Equation (15) and rearranging, the steady state distribution must satisfy:

$$D\frac{\partial}{\partial x}\rho(x) = -\mu\rho(x)\Big(\frac{\partial}{\partial x}V(x) - \kappa\frac{\partial}{\partial x}\rho(x)\Big) \tag{17}$$

which is a generalized diffusion-drift equation [40] containing a nonlinear effect. In the physics interpretation of above equation, $\rho(x)$ is the local density of particles; $D = \sigma^2/2 - \mu$ is the diffusion constant, $V(x) = -\ln \eta_0(x)$ is an external potential applied to the fluid (particles tend to converge to positions with relatively lower potential), and κ can be interpreted as a strength of "attraction" among the particles. It is easy to show that the solution to Equation (17) is:

$$\rho(x) = \Lambda e^{-\mu Y_x / D} \tag{18}$$

where $Y_x = V(x) - \kappa\rho(x)$, and Λ is a constant computed such that $\int_x \rho(x)dx = N$. To see that the above is true, simply obtain the derivative of both sides of Equation (18) with respect to x, and multiply by D to produce Equation (17).

Note how the steady state distribution is related to the potential $V(x)$, where $V(x)$ is shaped by consumer content biases, $\eta_0(x)$. The potential, $V(x)$, produces the "force-field" that gives our model its name.

Substituting for Y_x in Equation (18) with $V(x) - \kappa\rho(x) = -\ln \eta_0(x) - \kappa\rho(x)$, we get the final result.
This completes the proof. ■

B. Properties

Many interesting properties follow directly by inspecting Equation (7). We present them as corollaries of Theorem 1. We mention these corollaries without proof below. For proofs, please refer to Appendix C.

We start by observing that Equation (7) features $\rho(x)$ on both sides. While it is easy enough to solve for $\rho(x)$ numerically, it is possible to remove $\rho(x)$ from the right-hand-side, if desired. The corollary below derives this expression.

Corollary 1: The steady state belief distribution in Theorem 1 can alternatively be expressed by:

$$\rho(x) = \gamma^{-1}\left(\Lambda\eta_0^{\mu/D}(x)\right) .$$

where:

$$\gamma(y) = \left(\frac{D}{\mu\kappa}\right)^2 \Gamma(y; 2, \mu\kappa/D) ,$$

and Γ refers to the regular Gamma distribution in statistics, defined as:

$$\Gamma(y; \alpha, \beta) = \frac{\beta^\alpha}{(\alpha-1)!} y^{\alpha-1} e^{-\beta y} .$$

In other words, the value of $\rho(x)$, at position x in the belief space, is equal to the value of a scaled inverse function, $\gamma^{-1}(y)$, of a Gamma distribution, computed at $y = \Lambda\eta_0^{\mu/D}(x)$, with Gamma distribution parameters $\alpha = 2$ and $\beta = \mu\kappa/D$.

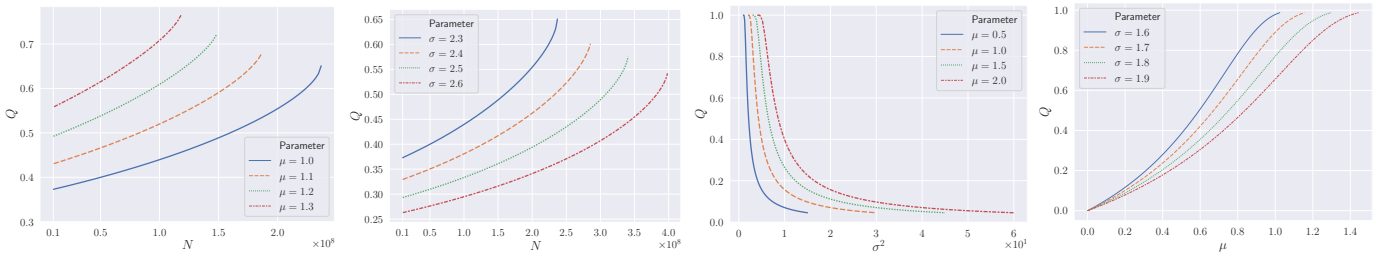
Corollary 2: The equilibrium distribution of population density, $\rho(x)$, is bimodal (i.e., polarized) with peaks aligned with those of the content weighting function, $\eta_0(x)$.

To measure the extent of polarization, it is convenient to define the ratio, Q, as follows:

$$Q = \frac{peak - valley}{peak + valley} = \frac{peak/valley - 1}{peak/valley + 1} \tag{19}$$

where $peak = \rho_{max} = max_x(\rho(x))$ and $valley = \rho_{min} = \rho(0)$. Note that, Q increases when the ratio $peak/valley$ increases. Note also that Q, by definition, ranges between 0 (when $peak = valley$) and 1 (when $valley = 0$). In this paper, we focus on the special case of a bimodal distribution. If there are more than two peaks in the belief space (more opinions), Q can be defined between each two adjacent peaks.

It may be tempting to think that the result in Corollary 2 (that polarization follows the shape of $\eta_0(x)$) trivially follows from the assumption that outlying content has more influence on beliefs. Interestingly, this is not true. Rather, the additional assumptions on (high) information volume and confirmation bias are also necessary for the correctness of Corollary 2. To appreciate this point, note intuitively that if the volume of information was low, and if everyone was able/willing to consume all content, then the resultant force will always move people towards the center, where it cancels out, no matter what η_0 looks like (as long as η_0 is symmetric around the center). Thus, for example, in an age of large monopolies, a content provider might find it more beneficial to strike a neutral tone to maximize distribution among ideologically distinct groups as opposed to limiting sales by appealing to one side only. It is the democratized broadcast (no monopoly), large volume (not being able to consume all content) and confirmation bias (prioritizing one's local neighborhood), embedded in the model, that cause Corollary 2 to be true. The corollary is true regardless of the exact expression of η_0, as long as it is bimodal. This is important because social psychology only gives us qualitative descriptions of shape, not exact equations.



(a) $\sigma = 2.3$, $\kappa = 5.0\text{E}{-9}$ (b) $\mu = 1.0$, $\kappa = 5.0\text{E}{-9}$ (c) $\kappa = 0.0$, $N = 1.0\text{E}8$. (d) $\kappa = 0.0$, $N = 1.0\text{E}8$

Fig. 2: Bifurcation of belief distribution as a function of total volume N, susceptibility μ and random influences σ. As N or μ increases, bifurcation becomes more visible. As σ increases, bifurcation becomes less visible.

Note also that, while our formulation seems to implicitly assume symmetry of $\eta_0(x)$ in x, intuitively, this assumption (while simplifying some proofs) is not strictly necessary. To appreciate the reason, note that confirmation bias causes all interactions among points to be local. Thus, what happens near one peak of $\eta_0(x)$ is largely independent from what happens near another peak. This allows the peaks to be different without invalidating the key results. Also, if the consumer biass were unimodal, at steady state we would observe one peak instead of two. Formal proofs of the above reasoning, however, are beyond the space available in this paper.

Corollary 3: For a sufficiently small κ, the polarization, Q, in the equilibrium distribution of population density, $\rho(x)$, (i) increases with population size N, (ii) decreases with random influence, σ, (iii) increases with susceptibility, μ, and (iv) increases with social influence factor, κ.

Proof: For all corollary proofs, see Appendix C. ■

Interestingly, note that Corollary 3 holds even for $\kappa = 0$. In other words, the super-linear effect of social influence is *not* needed for the above effects to manifest. The effects manifest even at $\kappa = 0$. These effects are driven by confirmation bias and bias for outlying content, balanced against an individual's willingness to seek information in an uncorrelated fashion to their bias. In fact, as pointed out earlier, the diffusion-drift model helps us understand the system by analyzing two combating forces: the diffusion and the drift. It shows that there is a competition between the confirmation bias and other random influences not correlated with bias. If the confirmation bias is weak, the polarization is less pronounced. If cultural factors encourage deliberate efforts to overcome bias, diffusion is higher and polarization is also less pronounced. As pointed out in the introduction, this is very reminiscent of the components of "thinking fast and slow" [2]. (known in psychology as "system 1" and "system 2"). We show that thinking "fast" (following our primitive instincts) increases polarization. Thinking "slow" (e.g., making a conscious effort to process views without correlation with our instincts) decreases it. There are indeed recent examples in history of how a deliberate emphasis on "thinking slow" successfully diminished harmful stereotypes presumably ingrained due to "thinking fast". An example is education campaigns focused on destroying subconscious gender and race biases and stereotypes [41], [42].

C. Numerical Observations

To give a sense of the above trends, in this section, we solve Equation (7) for $\rho(x)$ to compute the steady state belief distribution according to the modeled social dynamics. The figures demonstrate earlier observations, as follows:

- *Effect of volume:* Figures 2a and 2b show how the bifurcation of population density, Q (in the belief space), changes with population size, N, for different values of μ and σ. When N increases, bifurcation increases ($Q \to 1$). The effect is exacerbated when the susceptibility to local influence, μ, increases (higher drift), or when the impact of other diverse influences, σ, decreases (lower diffusion).
- *Effect of diversity:* Figure 2c shows how the diversity of influence affects the population's belief distribution. Parameter σ describes the degree to which belief updates are affected by random (i.e., diverse) factors outside the immediate belief neighborhood of consumers. As expected, increasing the σ (and thus decreasing the relative impact of confirmation bias) has a beneficial effect.
- *Effect of susceptibility:* More susceptible populations have a higher rate of belief change, μ, for the same distribution of neighboring content items. Figure 2d shows that when individuals' susceptibility increases, the degree of bifurcation increases ($Q \to 1$).
- *Effect of social influence:* The more pronounced the nonlinear effect of social influence is (i.e., the higher the parameter κ that describes the super-linear growth of influence with local density), the more pronounced the belief bifurcation, as shown in Figure 3.

The figures confirm that an increase in content volume leads to increased polarization. The effect is further magnified by social influence, κ, and susceptibility, μ.

Because of the motion model (III-C), the belief update always goes towards the center of the mass of the neighbors which is always bounded. The social influence, simply a weight of the mass, only affects the direction within the boundary. Note that, polarization emerges even if κ is zero. This is because bias for outlying content will still drive opinions towards more extreme versions over time.

V. Empirical Evidence

If the model predicts an increase in polarization as a consequence of increased information access and sharing, are these predictions consistent with empirical observations? To offer (at least) anecdotal evidence that addresses this question,

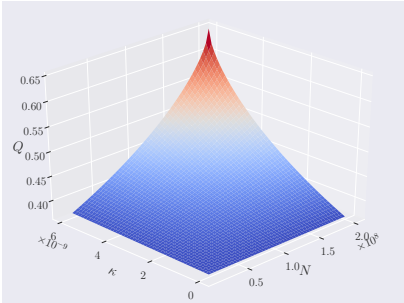


Fig. 3: Joint effect of population size, N, and social influence coefficient, κ. $\sigma = 2.3$, $\mu = 1.0$. When the population size N is fixed, Q increases with κ (i.e., the bifurcation becomes more pronounced as the "winner-takes-all" effect, described in Section (III-D.2) becomes stronger). For a fixed κ, bifurcation becomes more pronounced as the population grows.

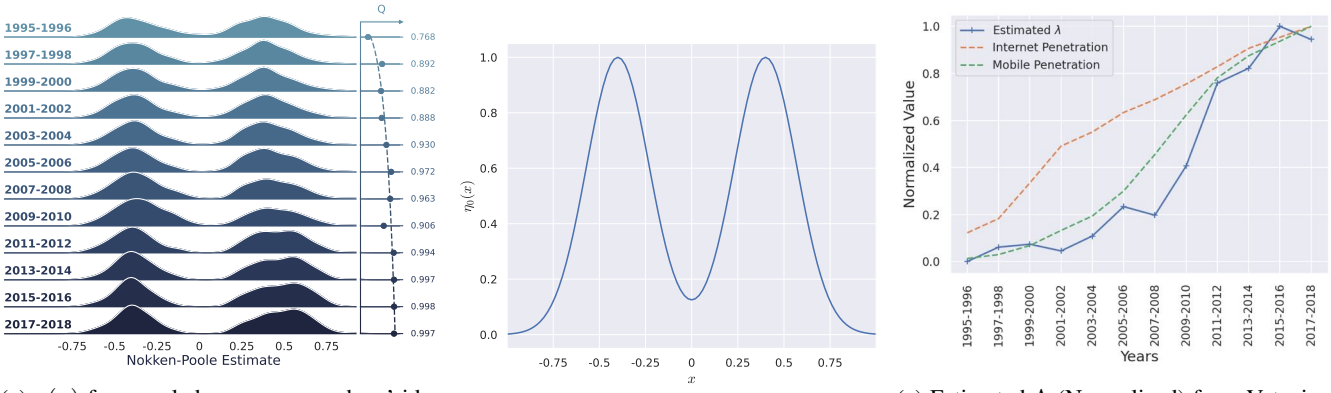
one needs to (i) determine a suitable proxy for the evolution of the number of democratized information broadcast sources, N (or equivalently, Λ, since it is proportional to N) in the age of information, and (ii) find a suitable estimate of the evolution of a population's distribution of beliefs during the same interval. We can then determine whether Equation (7) correctly relates the two quantities. Note that, Equation (7) describes the *steady state* belief distribution as a function of several parameters, including Λ (that is proportional to N). We assume, below, that the rate at which N changes is slow enough that the steady state distribution, $\rho(x)$, predicted by Theorem 1, is reached for a given N, before this N changes significantly. With that in mind, our methodology is as follows:

- *Estimating population belief distribution, $\rho(x)$:* We considered the population of the USA, as an example of a country that remained somewhat geographically distant and thus relatively less impacted by large world disturbances, such as wars (on its territory), famine, mass influx of refugees, or foreign occupation, in the last 25 years. While direct estimates of population beliefs are not as comprehensively documented for the US population, the ideology of members of the US congress is routinely assessed and documented. The Nokken-Poole estimates of Congress member ideologies are publicly available in the Voteview dataset [43]. They include liberalism-conservatism scores ranging from -1 to 1 that map semantically to our variable, x, representing positions in a belief space. Since Congress members who do not adequately reflect their constituents' ideals will likely not get re-elected, we took the distribution of Congress member ideology values as an approximation of the distribution of beliefs of the US population. In general, this is not exactly accurate as was shown in studies on other forces that impact Congress polarization [44]. The polarization in Congress is usually more extreme than that of the whole population. It is a unique data set, however, in that it includes detailed ground truth on member ideology that extends for many decades, making it an interesting case study. We defer a more faithful ideology estimation to future work. With the above caveat, in this paper, we simply scale the ideology distribution of Congress members by the country's population to compute the estimated belief density function, $\rho(x)$, for the US population. Figure 4a shows the results for years 1995 through 2018. Interestingly, note how as time goes by, the peaks become somewhat larger and the valley becomes deeper with a more depleted center. Figure 4a also shows the corresponding value of polarization index Q on the side of each time window (computed from Equation (19), applied after averaging each pair of peaks), demonstrating gradual increase as predicted by Corollary 3.
- *Estimating the number of democratized broadcast sources, N:* While this number cannot be estimated exactly, we expect that it must be correlated with two important technological developments in the last quarter-century; namely, (i) the growth of the Internet as a medium for democratized information exchange, and (ii) the proliferation of mobile phones as devices that enable untethered real-time information access and sharing.

From the empirical data, we can calculate peak/valley estimates for each congress. Further, we used the average population for Λ. By $Peak/Valley = (\eta_{max}^{\mu/D} - \Theta)/(\eta_{min}^{\mu/D} - \Theta)$ and $\Theta = \Lambda(\eta_{min}\eta_{max})^{\mu/D}(\frac{\mu\kappa}{D})$, we get an estimate of κ. Meanwhile, we also have the estimated belief density function, $\rho(x)$. It is clear from the data that $\rho(x)$ is bimodal, so we assumed that $\eta_0(x)$ is a mixture of two Gaussians. Using Theorem 1, we optimized the Gaussian mixture $\eta_0(x)$ to minimize the difference between the left and the right side of the equation across multiple congresses. The result shows that $\eta_0(x)$ has two peaks around $x = \pm 0.4$, $\mu/D = 0.33$, and $\kappa = 7.0\text{E}{-10}$. Figure 4b shows the estimated $\eta_0(x)$. Note that, this function represents human bias for outlying content. Indeed, it increases with distance from the origin up to a point, then decreases when the beliefs become too extreme.

From Equation (7), and using $\eta_0(x)$ from Figure 4b, we then used a least-squares method to find the best Λ for each congress so that the output, $\rho(x)$ of the equation (for the successive years) best matches Figure 4a. We normalized the resulting volume Λ to a 0 to 1 range, and plotted it together with the adoption curve of the Internet (i.e., the number of individuals with Internet access) and the adoption curve of mobile phones (i.e., the number of individuals with mobile phones) within the same period (also normalized to a 0 to 1 range). These curves are compared in Figure 4c.

A significant correlation is observed between Λ (that is proportional to the number of sources, N) and the proliferation of mobile phones. The latter lags behind the penetration of internet access, suggesting that once both technologies were present, their joint adoption correlated with the volume of democratized broadcast sources, which is the driver of polarization in our model (all other "knobs" in the model remaining constant). The slight delay between the mobile phone penetration curve and Λ might refer to a delay reaching the steady state, $\rho(x)$. While no definitive conclusions can be made based on this anecdotal evidence, the quality of the observed correlation is cause for concern. It motivates research community attention to the relation between content



(a) $\rho(x)$ from scaled congress members' ideologies.

(b) Estimated $\eta_0(x)$.

(c) Estimated Λ (Normalized) from Voteview dataset against Internet/Mobile penetration.

Fig. 4: Empirical study showing that mobile phone penetration is a good proxy for "democratization of content sharing" that is consistent with observed growth in congress polarization of the last 25 years.

access/sharing and polarization, as well as to mitigation policies that modulate the competing forces of diffusion and drift.

VI. Discussion

Our model suggests that the biggest factors impacting polarization stem from *consumer behavior and content volume*. Figure 2c suggests that when the impact of diverse random influences, σ, that individuals react to (beyond content that matches their beliefs) increases, polarization dramatically decreases. Cultural trends impacting consumer willingness to engage with diverse content outside their belief neighborhood can therefore account for differences in polarization across different nations, as observed in a recent study [45]. The model also shows that, when susceptibility to local influence, μ, increases, polarization increases. Importantly, when the number or democratized content sources (and thus shared/accessible content volume), N, increases, polarization increases.

The role of customized algorithmic curation (i.e., search and recommendation engines that feed consumers what maximizes engagement) in the emergence of polarization has been the subject of much debate. Our model is agnostic to the attribution of blame for selective exposure and is based only on the assumption of bounded confidence and bias for outlying content. At a high-level, as mentioned earlier in the paper, the work is related to the idea (from psychology) that human behavior is governed by two systems of the mind: *system 1* that operates more quickly, automatically and intuitively, reflecting deeply ingrained subconscious biases, and *system 2* that allocates more deliberate thought and mental attention that is more responsive to cultural effects, teachings, and logical thinking [2]. The drift and diffusion terms in our diffusion-drift model essentially model the struggle between the two. The ingrained biases are "older" (referring to more primitive responses of the brain such as aversion, disgust, and fear) and thus live in System 1. Logical reasoning and learned cultural norms are more teachable and live in System 2. The balance between the two may indeed change over time.

Technological innovations can sometimes be seen as "shocks" that change the norms, much the way, say, the invention of "birth control pills" as an effective means of contraception is sometimes credited with precipitating significant changes in social attitudes towards female sexuality, ultimately breaking previously ingrained gender biases and beliefs. The technology penetration for cell phones in the empirical study may be an example of another shock. In this case, it likely contributed to filter bubbles, greater susceptibility to local influence, and perhaps even rising polarization in congress, amplified by feedback loops due to content curation systems. In the information landscape, while biases (e.g., biases that govern human attention) do evolve, it is not clear if they evolve in the "right direction". Shocking news often attracts much attention initially then becomes normalized, as the amount of stimulus needed to produce the same attentional effect tends to increase over time. A more detailed study of the effects of such shocks and evolving norms in the information space is left for future work.

An interesting question is how to extend the model to account for publishing sources that require, say, a subscription fee but offer more reliable information. Currently there is no notion of reliability of a piece of information in our model. One might even argue that reliability is orthogonal to the problem addressed in this paper. Indeed, in the presence of polarization, both sides often cite reliable information (but add their own interpretations). For example, Russian soldiers may have evacuated some Ukrainian citizens to Russia. Some sources covered it as an unwanted abduction. Others covered it as a humanitarian gesture (to keep those individuals out of harm's way). Reliable (especially incomplete) information can indeed be presented in different ways, and people might choose to believe the version of the story that their ideology agrees with. Therefore, information reliability does not seem to be a solution to the problem addressed in this paper, although it is nevertheless an interesting avenue for future work.

In this paper, we do not make the assumption of rationality about the consumers. The proposed model starts from simple microscopic dynamics, incorporates well-known heuristics and biases in human decision-making [2] and describes population-level phenomena rather than individual-level one. From the

modeling side of opinion dynamics, we have much to expect in future work. For example, to study the actual time-evolving process (reactions to sudden changes), one needs to model the (equivalent of) velocity fields and the viscosity.

Finally, evidence provided in this paper remains anecdotal. It would be interesting to conduct a broader study in countries with different levels of polarization to understand the proportion of the variation in the level of polarization that is predictable from the prevalence of use of democratized broadcast media. It would also be interesting to compute better proxies for democratized media use. The penetration of mobile phones is one factor, but there may be others, such as the popularity of specific social media, the size of the Web, and the level of engagement with online content.

VII. Conclusions

The paper presented a social phenomenon caused by the age of democratized access; namely, growing ideological fragmentation exacerbated by information overload. A diffusion-drift model of this phenomenon was proposed. The model suggests that increasing volume, in the presence of confirmation bias and bias for more outlying content, can contribute to growing polarization. The paper is a call for solutions that may ameliorate this effect.

Acknowledgement

The authors thank Dr. Jonathan Bakdash, ARL, for helpful discussions of a psychological perspective. This work was conducted in part under DARPA award HR001121C0165, and in part under DoD Basic Research Office award HQ00342110002.

References

[1] C. Nall, "The political consequences of spatial policies: How interstate highways facilitated geographic polarization," *The Journal of Politics*, vol. 77, no. 2, pp. 394–406, 2015.

[2] D. Kahneman, *Thinking, fast and slow*. New York: Farrar, Straus and Giroux, 2011.

[3] R. S. Nickerson, "Confirmation bias: A ubiquitous phenomenon in many guises," *Review of general psychology*, vol. 2, no. 2, pp. 175–220, 1998.

[4] J. Lorenz, "Continuous opinion dynamics under bounded confidence: A survey," *International Journal of Modern Physics C*, vol. 18, no. 12, pp. 1819–1838, 2007.

[5] S. T. Fiske, "Attention and weight in person perception: The impact of negative and extreme behavior." *Journal of personality and Social Psychology*, vol. 38, no. 6, p. 889, 1980.

[6] P. J. Shoemaker, "Hardwired for news: Using biological and cultural evolution to explain the surveillance function." *Journal of communication*, 1996.

[7] P. Lamberson and S. Soroka, "A model of attentiveness to outlying news," *Journal of Communication*, vol. 68, no. 5, pp. 942–964, 2018.

[8] L. R. Varshney, "Must surprise trump information?" *IEEE Technology and Society Magazine*, vol. 38, no. 1, pp. 81–87, 2019.

[9] T. Abdelzaher, "On urban event tracking from online media: A social cognition perspective," in *2019 IEEE First International Conference on Cognitive Machine Intelligence (CogMI)*. IEEE, 2019, pp. 160–167.

[10] C. Xu, J. Li, D. Sun, R. Wang, T. Abdelzaher, J. Graham, and B. Szymanski, "The curse of asymmetric polarization dynamics in the age of information overload," in *In Proc. 6th International Workshop on Social Sensing (SocialSens)*, 2021.

[11] C. Xu, J. Li, T. Abdelzaher, H. Ji, B. K. Szymanski, and J. Dellaverson, "The paradox of information access: On modeling social-media-induced polarization," *arXiv preprint arXiv:2004.01106*, 2020.

[12] T. Abdelzaher, H. Ji, J. Li, C. Yang, J. Dellaverson, L. Zhang, C. Xu, and B. K. Szymanski, "The paradox of information access: Growing isolation in the age of sharing," *arXiv preprint arXiv:2004.01967*, 2020.

[13] M. McPherson, L. Smith-Lovin, and J. M. Cook, "Birds of a feather: Homophily in social networks," *Annual review of sociology*, vol. 27, no. 1, pp. 415–444, 2001.

[14] J. G. March, "Exploration and exploitation in organizational learning," *Organization science*, vol. 2, no. 1, pp. 71–87, 1991.

[15] M. H. DeGroot, "Reaching a consensus," *Journal of the American Statistical Association*, vol. 69, no. 345, pp. 118–121, 1974.

[16] G. L. Gilardoni and M. K. Clayton, "On reaching a consensus using degroot's iterative pooling," *The Annals of Statistics*, pp. 391–401, 1993.

[17] N. E. Friedkin, *A structural theory of social influence*. Cambridge University Press, 2006, vol. 13.

[18] ——, "Norm formation in social influence networks," *Social networks*, vol. 23, no. 3, pp. 167–189, 2001.

[19] N. E. Friedkin and E. C. Johnsen, *Social influence network theory: A sociological examination of small group dynamics*. Cambridge University Press, 2011, vol. 33.

[20] F. Ceragioli and P. Frasca, "Continuous and discontinuous opinion dynamics with bounded confidence," *Nonlinear Analysis: Real World Applications*, vol. 13, no. 3, pp. 1239–1251, 2012.

[21] J. Ghaderi and R. Srikant, "Opinion dynamics in social networks: A local interaction game with stubborn agents," in *2013 American Control Conference*. IEEE, 2013, pp. 1982–1987.

[22] J. Lorenz, "Fostering consensus in multidimensional continuous opinion dynamics under bounded confidence," in *Managing complexity: insights, concepts, applications*. Springer, 2008, pp. 321–334.

[23] S. E. Parsegov, A. V. Proskurnikov, R. Tempo, and N. E. Friedkin, "Novel multidimensional models of opinion dynamics in social networks," *IEEE Transactions on Automatic Control*, vol. 62, no. 5, pp. 2270–2285, 2016.

[24] M. Pineda, R. Toral, and E. Hernández-García, "The noisy hegselmann-krause model for opinion dynamics," *The European Physical Journal B*, vol. 86, no. 12, p. 490, 2013.

[25] G. Deffuant, F. Amblard, G. Weisbuch, and T. Faure, "How can extremism prevail? a study based on the relative agreement interaction model," *Journal of artificial societies and social simulation*, vol. 5, no. 4, 2002.

[26] V. Amelkin, F. Bullo, and A. K. Singh, "Polar opinion dynamics in social networks," *IEEE Transactions on Automatic Control*, vol. 62, no. 11, pp. 5650–5665, 2017.

[27] Wikipedia, "Fokker–planck equation." [Online]. Available: https://en.wikipedia.org/wiki/Fokker-Planck_equation

[28] J.-M. Lasry and P.-L. Lions, "Jeux à champ moyen. i–le cas stationnaire," *Comptes Rendus Mathématique*, vol. 343, no. 9, pp. 619–625, 2006.

[29] L. Stella, F. Bagagiolo, D. Bauso, and G. Como, "Opinion dynamics and stubbornness through mean-field games," in *52nd IEEE Conference on Decision and Control*. IEEE, 2013, pp. 2519–2524.

[30] D. Bauso, H. Tembine, and T. Basar, "Opinion dynamics in social networks through mean-field games," *SIAM Journal on Control and Optimization*, vol. 54, no. 6, pp. 3225–3257, 2016.

[31] D. P. Paolo, S. Elena, and T. Marco, "Polarization and coherence in mean field games driven by private and social utility," *arXiv preprint arXiv:2107.06667*, 2021.

[32] A. Baronchelli, M. Felici, E. Caglioti, V. Loreto, and L. Steels, "Sharp transition towards shared vocabularies in multi-agent systems," *J Stat-Mech: TheoryExp*, vol. 2, p. P06014, 2006.

[33] G. Korniss, W. Zhang, C. Lim, and B. Szymanski, "Social consensus through the influence of committed minorities," *Phys. Rev. E*, vol. 84, no. 1, p. 011130, 2011.

[34] J. Xie, J. Emenheiser, M. Kirby, S. Sreenivasan, B. Szymanski, and G. Korniss, "Evolution of opinions on social networks in the presence of competing committed groups," *PLoS One*, vol. 7, no. 3, p. e33215, 2012.

[35] Waagen, A and Verma, G and Chan, K and Swami, A, and D'Souza, R, "Effect of zealotry in high-dimensional opinion dynamics models," *Phys. Rev. E*, vol. 91, no. 2, p. 022811, 2015.

[36] C. Doyle, S. Sreenivasan, B. Szymanski, and G. Korniss, "Social consensus and tipping points with opinion inertia," *Physica A*, vol. 443, pp. 316–323, 2015.

[37] G. Korniss, W. Zhang, C. Lim, and B. Szymanski, "Analytic treatment of tipping points for social consensus in large random networks," *Phys. Rev. E*, vol. 86, no. 6, p. 061134, 2012.

[38] N. E. Friedkin and E. C. Johnsen, "Social influence and opinions," *Journal of Mathematical Sociology*, vol. 15, no. 3-4, pp. 193–206, 1990.

[39] M. Shermer, *The believing brain: From ghosts and gods to politics and conspiracies—How we construct beliefs and reinforce them as truths*. Macmillan, 2011.

[40] L. Landau and E. Lifshitz, "Theoretical physics, vol. 6, fluid mechanics," 1987.
[41] R. D. Godsil, L. R. Tropp, P. A. Goff, and J. MacFarlane, "The effects of gender roles, implicit bias, and stereotype threat on the lives of women and girls," *The Science of Equality*, vol. 2, no. 1, pp. 14–15, 2016.
[42] C. Staats, "Understanding implicit bias: What educators should know." *American Educator*, vol. 39, no. 4, p. 29, 2016.
[43] J. B. Lewis, K. Poole, H. Rosenthal, A. Boche, A. Rudkin, and L. Sonnet, "Voteview: Congressional roll-call votes database," 2021, https://voteview.com/.
[44] J. L. Carson, M. H. Crespin, C. J. Finocchiaro, and D. W. Rohde, "Redistricting and party polarization in the us house of representatives," *American Politics Research*, vol. 35, no. 6, pp. 878–904, 2007.
[45] L. Boxell, M. Gentzkow, and J. M. Shapiro, "Cross-country trends in affective polarization," National Bureau of Economic Research, Tech. Rep., 2020.

APPENDIX A

From Eq. (5), the weighted center of gravity within the neighborhood of ϵ:

$$f(\mathcal{X}^{(i)}(t)) = \frac{\int_{x_i(t)-\epsilon}^{x_i(t)+\epsilon} x\ \rho(x,t)\eta(x,t)\,dx}{\int_{x_i(t)-\epsilon}^{x_i(t)+\epsilon} \rho(x,t)\eta(x,t)\,dx},$$

Substituting from Eq. (8) and Eq. (9) for $\rho(x,t)$ and $\eta(x,t)$, respectively, then integrating, the numerator becomes:

$$\begin{aligned}&\int_{x_i(t)-\epsilon}^{x_i(t)+\epsilon} x\,\rho(x,t)\eta(x,t)\,dx \\ &\simeq 2\epsilon x_i(t)\eta(x_i(t))\rho(x_i(t)) + \frac{2\epsilon^3}{3}\Big[\eta'(x_i(t))\rho(x_i(t)) \\ &+ \eta(x_i(t))\rho'(x_i(t)) + x_i(t)\eta'(x_i(t))\rho'(x_i(t))\Big],\end{aligned} \tag{20}$$

while the denominator becomes:

$$\int_{x_i(t)-\epsilon}^{x_i(t)+\epsilon} \rho(x,t)\eta(x,t)\,dx \tag{21}$$

$$\simeq 2\,\epsilon\,\eta(x_i(t))\rho(x_i(t)) + \frac{2}{3}\epsilon^3\eta'(x_i(t))\rho'(x_i(t))\,. \tag{22}$$

Therefore the weighted center can be estimated by series expansion:

$$f(\mathcal{X}^{(i)}(t)) = x_i(t) + \frac{\epsilon^2}{3}\left[\frac{\eta'(x_i(t))}{\eta(x_i(t))} + \frac{\rho'(x_i(t))}{\rho(x_i(t))}\right] + O[\epsilon^3]\,.$$

APPENDIX B

The Fokker-Planck equation of motion [27] states that if positions, x_i, of particles is governed by a stochastic diffusion-drift process given by the following equation of motion:

$$\frac{dx_i}{dt} = \tilde{\mu}(x_i,t) + \sigma(x_i,t)\frac{dW}{dt}. \tag{23}$$

(where $\tilde{\mu}(x_i,t)$ can be any function) then the probability density, $p(x,t)$, of these particles (in an analogy to our population density $\rho(x,t)$) is given by the following equation:

$$\frac{\partial}{\partial t}p(x,t) = -\frac{\partial}{\partial x}\left[\tilde{\mu}(x,t)p(x,t)\right] + \frac{\partial^2}{\partial x^2}\left[\tilde{D}(x,t)p(x,t)\right] \tag{24}$$

where:

$$\tilde{D}(x,t) = \frac{\sigma^2(x,t)}{2}. \tag{25}$$

Moreover, from the continuity equation, we know that:

$$\frac{\partial}{\partial t}p(x,t) = -\frac{\partial}{\partial x}j(x,t)\,. \tag{26}$$

Combining Eq. (26) with Eq. (24), and integrating with respect to x, we get the flow, $j(x,t)$, as:

$$j(x,t) = \tilde{\mu}(x,t)p(x,t) - \frac{\partial}{\partial x}\left[\tilde{D}(x,t)p(x,t)\right]. \tag{27}$$

Moreover, at steady state:

$$\tilde{\mu}(x,t)p(x,t) - \frac{\partial}{\partial x}\left[\tilde{D}(x,t)p(x,t)\right] = 0 \tag{28}$$

Comparing Equation (12) to Equation (23), it is clear that our model satisfies the diffusion-drift process with:

$$\tilde{\mu}(x_{i,t},t) = \mu(\ln\eta_t(x_{i,t}))' + \mu\rho'(x_{i,t},t)/\rho(x_{i,t},t)\,. \tag{29}$$

Substituting for $\tilde{\mu}(x_{i,t},t)$ in Equation (28) at steady state, we get:

$$\begin{aligned}0 &= \tilde{\mu}(x,t)\rho(x,t) - \frac{\partial}{\partial x}\left[\frac{\sigma^2}{2}\rho(x,t)\right] \\ &= \mu\rho(x,t)\frac{\partial}{\partial x}(\ln\eta_t(x)) + \mu\frac{\partial}{\partial x}\rho(x,t) - \frac{\sigma^2}{2}\frac{\partial}{\partial x}\rho(x,t) \\ &= -(\frac{\sigma^2}{2}-\mu)\frac{\partial}{\partial x}\rho(x,t) + \mu\rho(x,t)\frac{\partial}{\partial x}\left[\ln\eta_0(x) + \lambda\rho(x,t)\right] \\ &= -D\frac{\partial}{\partial x}\rho(x,t) - \mu\rho(x,t)\frac{\partial}{\partial x}\left[V(x) + g\rho(x,t)\right].\end{aligned} \tag{30}$$

Equation (17) thus follows, where $D = \frac{\sigma^2}{2} - \mu$, $V(x) = -\ln\eta_0(x)$, and $g = -\kappa$. ■

APPENDIX C

Proof of Corollary 1: Equation (7) can be rewritten as:

$$\rho(x)\ e^{-\frac{\mu\kappa}{D}\rho(x)} = \Lambda\eta_0^{\mu/D}(x)\,. \tag{31}$$

The left-hand-side, $\rho e^{-\frac{\mu\kappa}{D}\rho}$, has the form of Gamma (Γ) distribution. The standard Gamma distribution reads:

$$\Gamma(y;\alpha,\beta) = \frac{\beta^\alpha}{(\alpha-1)!}y^{\alpha-1}e^{-\beta y} \tag{32}$$

Let us take $\alpha = 2$ and $\beta = \mu\kappa/D$, then define:

$$\gamma(y) = \left(\frac{D}{\mu\kappa}\right)^2\Gamma(y,2,\mu\kappa/D) \tag{33}$$

Substituting in Equation (33) with the definition of Γ from Equation (32), with $\alpha = 2$ and $\beta = \mu\kappa/D$, we get:

$$\gamma(y) = ye^{-\frac{\mu\kappa}{D}y} \tag{34}$$

Comparing with Equation (31), we get:

$$\gamma(\rho(x)) = \Lambda\eta_0^{\mu/D}(x) \tag{35}$$

Thus:

$$\rho(x) = \gamma^{-1}\left(\Lambda\eta_0^{\mu/D}(x)\right)$$

which completes the proof. ■

Proof of Corollary 2: To prove the corollary, we obtain the derivative of (both sides of) Equation (7) from Theorem 1, and look for the peaks (where $\partial\rho/\partial x = 0$). This yields:

$$\frac{\mu}{D}\ \eta_0^{\frac{\mu}{D}-1}\ \frac{\partial \eta_0}{\partial x}\ e^{\frac{\mu\kappa\rho(x)}{D}} = 0 \tag{36}$$

In the non-trivial case (i.e., for a non-zero η_0 and μ), the above equation can only be satisfied when $\partial\eta_0/\partial x = 0$. Thus, the extrema of $\rho(x)$ and η_0 coincide. By obtaining the second derivative around the point where the first derivative is zero, we can further show that the $\partial^2\rho(x)/\partial x^2$ and $\partial^2\eta_0/\partial x^2$ always have the same signs. Thus, the maxima and minima of $\rho(x)$ and η_0 coincide. Since η_0 is bimodal (symmetrically around $x = 0$), so is $\rho(x)$. ■

Proof of Corollary 3: Consider a Taylor series expansion of the exponential term in Equation (7). When κ is small, ignoring higher order terms, the equation becomes:

$$\rho(x) = \frac{\Lambda\eta_0^{\mu/D}}{1 - \Lambda\eta_0^{\mu/D}(\frac{\mu\kappa}{D})} \tag{37}$$

Let η_{max} denote the value of $\eta_0(x)$ when $\rho(x)$ reaches its $peak$, ρ_{max}, and η_{min} denote the value of $\eta_0(x)$ when $\rho(x)$ is at the $valley$, ρ_{min} (i.e., at $x = 0$). Substituting for ρ_{max} and ρ_{min} from Equation (37), after manipulation, we get:

$$\frac{Peak}{Valley} = \frac{\eta_{max}^{\mu/D} - \Theta}{\eta_{min}^{\mu/D} - \Theta} \tag{38}$$

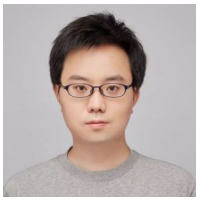
where

$$\Theta = \Lambda(\eta_{min}\eta_{max})^{\mu/D}(\frac{\mu\kappa}{D}) \tag{39}$$

It can be easily seen that the right hand side in Equation (38), and thus the ratio of $Peak/Valley$:

- Increases with Λ (because Θ increases when Λ increases, according to Equation (39)). Note that, Λ is proportional to the number of information sources, N.
- Increases with μ/D. (It can be shown that the denominator decreases faster than the numerator with increased μ/D.) Furthermore, since $D = \sigma^2 - \mu$, this means that $Peak/Valley$ increases with increased μ and decreases with increased σ, as both lead to increased μ/D.
- Increases with κ (because Θ increases when κ increases).

Since Q changes monotonically with $Peak/Valley$, the same trends apply to Q and the corollary follows. ■

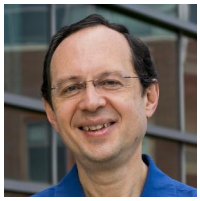


Chao Xu (Ph.D., UCSD, 2021) is now a Postdoc at Kavli Institute for Theoretical Sciences (KITP), Beijing. He studies theoretical condensed matter physics, including unconventional superconductivity, spin model, and strongly-correlated system.

Jinyang Li is a PhD student of Computer Science at University of Illinois at Urbana-Champaign. His research interests include Cyber Physical System, Real-Time System, Internet of Things and Edge AI.

Dachun Sun is currently PhD student of Computer Science at University of Illinois at Urbana-Champaign. He has a background of machine learning and software engineering. His research interest lies in computer vision and computational network analysis.

Jinning Li is a PhD student of Computer Science at University of Illinois at Urbana-Champaign. His research interests include Data Mining and Machine Learning for Social Networks and Cyber-Physical Systems as well as Autonomous Driving, NLP, and Computer Vision.



Tarek Abdelzaher (Ph.D., UMich, 1999) is a Sohaib and Sara Abbasi Professor of CS and Willett Faculty Scholar (UIUC), with over 300 refereed publications in Real-time Computing, Distributed Systems, Sensor Networks, and IoT. He served as Editor-in-Chief of J. Real-Time Systems for 20 years, an AE of IEEE TMC, IEEE TPDS, ACM ToSN, ACM TIoT, and ACM ToIT, among others, and chair of multiple top conferences in his field. He is a fellow of IEEE and ACM.

Jesse Graham (Ph.D., University of Virginia, 2010) is a George S. Eccles Chair in Business Ethics and Associate Professor of Management at Eccles School of Business, University of Utah. He served as Associate Editor of Social Psychological and Personality Science (2015 – 2018), Editorial Board member of Journal of Personality and Social Psychology (2015–2021) and Social Psychological and Personality Science (2011 – 2015).

Michael Macy (Ph.D., Harvard, 1980) is currently Goldwin Smith Professor of Arts and Sciences in Sociology and Director of the Social Dynamics Laboratory at Cornell. His research team studies the interplay between network topology and the dynamics of social interaction. His research has been published in leading journals, including Science, PNAS, American Journal of Sociology, American Sociological Review, and Annual Review of Sociology.

Christian Lebiere (Ph.D., CMU, 1998) is Research Scientist, Human-Computer Interaction Institute, School of Computer Science, Carnegie Mellon University. His main research interest is cognitive architectures and their applications to psychology, artificial intelligence, economics, decision theory, and human-computer interaction.

Boleslaw Szymanski (Ph.D., Polish Academy of Sciences, 1976) is a Claire and Roland Schmitt Distinguished Professor at the Department of Computer Science, Rensselaer Polytechnic Institute. He is the Founding Head of the Center for Network Science and Technology, Rensselaer Polytechnic Institute. He's known for multiple contributions into computer science, including Szymanski's algorithm. He is a IEEE Fellow and a foreign member of the Polish Academy of Sciences.