Learning from Viral Content*

Krishna Dasaratha† Kevin He‡

First version: August 20, 2022
This version: October 17, 2022

Abstract

We study learning on social media with an equilibrium model of users interacting with shared news stories. Rational users arrive sequentially and each observes an original story (i.e., a private signal) and a sample of predecessors' stories in a news feed, then decides which stories to share. The observed sample of stories depends on what predecessors share as well as the sampling algorithm, which represents a design choice of the platform. We focus on how much the algorithm relies on virality (how many times a story has been previously shared) when generating news feeds. Showing users more viral stories can increase information aggregation, but it can also generate steady states where most shared stories are wrong. Such misleading steady states self-perpetuate, as users who observe these wrong stories develop wrong beliefs, and thus rationally continue to share them. We find that these bad steady states appear discontinuously, and even a benevolent platform designer either accepts these misleading steady states or induces fragile learning outcomes in the optimal design.

Keywords: social learning, selective equilibrium sharing, social media, platform design, endogenous virality

*We thank Michel Benaïm, Aislinn Bohren, Tommaso Denti, Mira Frick, Ben Golub, Ryota Iijima, George Mailath, Suraj Malladi, Meg Meyer, Evan Sadler, Philipp Strack, Heidi Thysen, Rakesh Vohra, and numerous seminar participants for valuable comments and discussions. Byunghoon Kim, Stephan Xie, and Tyera Zweygardt provided excellent research assistance. We gratefully acknowledge financial support from NSF Grants SES-2214950 and SES-2215256.

†Boston University. Email: krishnadasaratha@gmail.com

‡University of Pennsylvania. Email: hesichao@gmail.com

1 Introduction

In recent years, viral content on social media platforms has become a major source of news and information for many people. What content users consume often depends on the *news feeds* created by platforms like Twitter and Facebook. Thus, which stories go viral and which disappear is jointly determined by the algorithms generating these feeds and users' actions on the platforms (e.g., sharing, retweeting, or upvoting stories).

How does the design of the news feed affect how users learn on such platforms? Consider a platform deciding how much to push widely shared (or highly upvoted) content into users' news feeds. On the one hand, a news feed that focuses on showing widely shared stories can create a social version of the confirmation bias: incorrect stories spread more widely and determine people's beliefs, even though they only account for a small minority of the information that continues to arrive. The idea is that when stories supporting an incorrect position are shared more, subsequent users tend to see these incorrect stories in their news feeds due to the stories' popularity, and hence form incorrect beliefs. For users who derive utility from sharing content that is accurate or content that agrees with their beliefs, the false stories in the news feeds distort their beliefs and make it rational for the users to share these stories and further increase their popularity. Users have less exposure to the true stories: even if these stories are more numerous, they are shared less than the false stories and therefore shown less by the news-feed algorithm.

But on the other hand, selecting news stories based on their popularity may help aggregate more information. Seeing a positive story in a news feed that focuses on widely shared content gives a user more information than the realization of a single signal. The fact that this story is popular enough to appear in the news feed also tells the user about the past sharing decisions of their predecessors, and thus lets the user draw inferences about the many stories that are not contained in the current news feed. In some circumstances, seeing just a few stories in a news feed that focuses on widely shared content can lead to strong Bayesian beliefs about the state of nature. This can happen even if individual stories are imprecise signals about the state, since rational users can use the selection of these stories to infer the social consensus on the platform.

This work studies the learning implications of the platform's news-feed design and examines the key trade-offs in choosing how much to feature viral content in an equilibrium framework. It relates to contemporary public debates and policy decisions about news-feed algorithms. The impact of social media on society's beliefs has received significant attention in recent years. Misinformation about issues ranging from public health to politics frequently spreads widely on social media. To what extent is this related to platforms that push more "viral" content into users' feeds?

To answer these questions, we develop an equilibrium model of people learning from news feeds and sharing news stories on a social media platform. A large number of users arrive in turn and learn about a binary state. Each user receives a conditionally independent binary signal about the state (which we call a *news story*) and observes a sample of stories from predecessors (which we call a *news feed*). These stories are sampled using a news-feed algorithm that interpolates between choosing a uniform sample of the past stories and choosing each story with probability proportional to its *popularity*, or the number of times that it has been shared. Users are Bayesians and know the news-feed algorithm, so they appropriately account for selection in the stories they see.[1] Users then choose which of these news-feed stories to share. We assume users prefer to share stories that match the true state, given their endogenous beliefs. This relatively straightforward utility specification generates rich learning dynamics that highlight the role of the platform's sampling algorithm.

The platform's design choice in our model is the *virality weight* λ that captures the weight placed on popularity when generating news feeds — higher λ corresponds to a news-feed algorithm that focuses more on showing viral stories. The evolution of content on the platform is described by a stochastic process in $[0, 1]$ we call *viral accuracy*, which measures the relative popularity of the stories that match the true state in each period. We show this fraction almost surely converges to a (random) steady-state value, which depends on the randomness in signal realizations and in news-feed sampling. In equilibrium, there is always an *informative steady state* where most stories in news feeds match the state. But when the virality weight is high enough, there can also be a *misleading steady state* in equilibrium, where most stories in news feeds do not match the state (so viral accuracy is less than $\frac{1}{2}$). At a misleading steady state, users tend to see false stories, and therefore believe in the wrong state and share these false stories. The misleading steady states correspond to the socially-generated confirmation bias described above.

As the platform increases its virality weight, beliefs at informative steady states become more

[1] Many of the main forces we highlight would also appear if we assumed agents fail to account for this selection.

accurate. This result formalizes the intuition mentioned above that a more viral news feed helps aggregate more information. For low enough values of λ, there is a unique steady state and it is informative, so increasing λ monotonically improves learning outcomes. But there is often a *critical virality weight* $\lambda^* < 1$ where a misleading steady state emerges for all $\lambda \geq \lambda^*$. A platform choosing λ therefore faces a trade-off between facilitating more information aggregation and preventing the possibility of a misleading steady state in equilibrium. A key finding is that the misleading steady state emerges *discontinuously*: at the threshold virality level λ^* where the misleading steady state first appears, the probability of learning converging to this bad steady state is strictly positive. As a consequence, the accuracy of beliefs jumps downward at this threshold.

After establishing the existence of this threshold λ^* where misleading steady states appear, we ask how λ^* changes with parameters of the environment. Consuming and interacting with too much social information on the platform relative to the quality of private information from other sources lowers the threshold λ^*, and therefore makes people more susceptible to misleading steady states.

Finally, we consider a benevolent platform designer who chooses the virality weight to maximize users' equilibrium utility from sharing stories on the platform. Because of how the set of equilibrium steady states changes around λ^*, even this benign designer objective must lead to either misleading steady states or fragile social learning, in the sense that an arbitrarily small unanticipated shock to the environmental parameters substantially lowers users' welfare.

At a technical level, our paper applies and extends techniques from a mathematics literature on stochastic approximation and in particular generalized Pólya urns to analyze learning dynamics. Our model is a multidimensional stochastic process that can often converge to multiple steady states. As a result, understanding equilibrium outcomes—or even outcomes under a particular fixed strategy—is complex. To make progress despite this complexity, we show that outcomes under a specific simple strategy (sharing stories that match a majority of one's observations) tell us about the (potentially much less simple) equilibrium outcomes.

1.1 Related Literature

We first discuss how our model relates to a recent literature on learning from shared signals. Several papers have looked at different models of news sharing or signal sharing. As we discuss in detail below, the existing work focuses on the dissemination of a single signal, or on settings where signals are shared once with network neighbors but not subsequently re-shared. Our model differs on these two dimensions. First, we consider a platform where many signals about the same state circulate simultaneously. These signals interact: a user's social information consists of the multiple stories that they see in their news feed, so the probability that they share a given story depends on whether the other stories corroborate it or contradict it. Second, we allow signals to be shared widely through a central platform algorithm that generates news feeds for all users. A signal can become popular due to early agents' sharing decisions and get pushed into a later agent's news feed, and this later agent can re-share the same signal. The combination of these two model features generates the social version of confirmation bias that we outlined earlier.

Bowen, Dmitriev, and Galperti (2022) study a model where signals are selectively shared at most once with network neighbors, but agents are misspecified and partially neglect this selection. This bias leads to mislearning, and it also generates polarization in social networks with echo chambers. By contrast, we focus on rational agents who make endogenous sharing decisions in equilibrium. Bowen et al. (2022) note that:

> "[...]the Internet has also brought an abundance of information, which should lead people to learn quickly and beliefs to converge (not diverge) according to standard economic models."

Our results imply that even if people observe a large amount of information and rationally account for selection, they can converge to a misleading steady state if they mostly observe social information from peers and not private information (as may be the case on real-world social media platforms). Indeed, our comparative statics results in Section 3.4 show that the possibility of a misleading steady state arises precisely when users are exposed to enough social information.

Another group of papers in operations research and economics studies settings with "fake news" where people decide whether to share a story depending on the outcome of a (possibly noisy) fact check (e.g., Papanastasiou (2020), Kranton and McAdams (2022), and Merlino, Pin, and Tabasso (2022)) or depending on their prior beliefs about the story's likelihood (e.g., Bloch, Demange, and Kranton (2018), Acemoglu, Ozdaglar, and Siderius (2022) and Hsu, Ajorlou, and Jadbabaie (2021)). Most of these papers consider the diffusion of a single signal that can be re-shared through a network, while Kranton and McAdams (2022) look at the supply-side decisions of information producers when consumers can share their stories at most once with network neighbors.[2] We focus on a different dimension of platform-design choices. Instead of asking about the network structure that connects users on the platform (e.g., echo chambers) or fact-checking technologies, we consider the platform's choice in terms of showing its users more or less viral content.

Buechel, Klößner, Meng, and Nassar (2022), like our work, also consider an environment where agents can share and re-share copies of a signal. They study a model in which agents' sharing behavior resembles the DeGroot heuristic. In particular, their agents' sharing is independent of beliefs, while we study sharing rules that maximize utility given beliefs. Another difference is that we study a stochastic model where the diffusion of news stories is random even conditional on the realizations of all stories, due to the randomness in news-feed sampling. By contrast, Buechel et al. (2022) analyze a mean-field approximation. The stochastic approximation techniques we use may be useful for understanding a stochastic version of their model.

Finally, there are some similarities between misleading steady states in our model and herding in observational social-learning models (Banerjee (1992), Bikhchandani, Hirshleifer, and Welch (1992), and a large subsequent literature). In both cases, incorrect initial signals can lead to persistent wrong beliefs. A clear difference is that in observational social learning, the obstruction to learning is that agents take actions that only coarsely reflect the private signals informing those actions. In our model, agents observe signals, not actions, and incorrect beliefs instead come from selection in the observed signals. A second distinction we see as important is that misleading steady states persist in our model even though new private information continues to arrive and get shared on the platform. By contrast, the classical herding results rely sharply on the later agents' private signals becoming completely lost once society reaches an information cascade. Perhaps closest to our model within the observational social-learning literature, several papers assume agents observe a random sample of predecessors' actions (Banerjee and Fudenberg (2004), Lévy, Pęski, and Vieille (2022), and Kabos and Meyer (2021)). Our techniques, meanwhile, are based on the same mathematics literature as Arieli, Babichenko, and Mueller-Frank (2022), who model the distribution of actions taken by agents as a generalized Pólya urn.

[2]Merlino, Pin, and Tabasso (2022)'s model features one true and one false message.

2 Model

We consider a finite society with n people learning in sequence about an unknown state of nature $\omega \in \{-1, 1\}$. Everyone starts with the common prior that each state is equally likely. Each agent receives a binary signal $s_i \in \{-1, 1\}$ about the state, interpreted as a *news story*. Call $s_i = -1$ a negative story and $s_i = 1$ a positive story. We assume stories are conditionally independent and symmetric, so that $\mathbb{P}[s_i = -1|\omega = -1] = \mathbb{P}[s_i = 1|\omega = 1] = q$ for some *story precision* $0.5 < q < 1$. We also keep track of the *popularity score* of each story s_i, denoted $\rho(s_i)$. Each story starts with a score of 1 when it is first posted, and the score increases by 1 each time the story is shared by someone else.

We fix a *news feed* size $K \geq 1$. The first K agents receive no information other than their own news stories and mechanically post these stories onto the platform (our analysis will focus on long-run steady states when n is large, and the main insights are robust to other assumptions about what the earliest agents in the sequence observe.) For each agent $i \geq K + 1$, i sees a news feed containing K stories posted by predecessors. The news feed only shows the realizations of the K sampled stories, but not their popularity scores or vintage. Then, agent i *shares* C out of the K stories from their news feed, increasing each shared story's popularity score by 1, for some *capacity* $C \leq K/2$. Agent i gets utility $u > 0$ for each shared story that matches the state ω.[3] Then, i posts their own story s_i onto the platform.

The platform's *viral weight* $\lambda \in [0, 1]$ determines how it samples K stories from $s_1, s_2, ..., s_{i-1}$ to populate i's news feed. For each of the K slots in the news feed, with probability λ, a story is sampled with probabilities proportional to the $i - 1$ stories' current popularity scores. With the complementary probability, a story is sampled uniformly at random from the $i - 1$ stories. We assume for simplicity that all stories are sampled with replacement (as we approach the steady state, the effect of replacement vanishes). All draws are independent.

The platform's sampling rule includes two special cases:

1. **Popularity-based sampling** ($\lambda = 1$): A story that has been shared twice as often as another has twice the probability of being sampled.

[3] Pennycook et al. (2020, 2021) find experimental evidence that people have an intrinsic preference for sharing news from more trustworthy sources, which are more likely to accurately reflect the state.

2. **Uniform sampling** ($\lambda = 0$): Predecessors' sharing decisions do not affect sampling.

More generally, sampling rules with λ between zero and one interpolate between these two cases. The viral weight λ measures how much the news feed shows more popular stories relative to random stories.

The n agents are uniformly randomly placed into the n positions, and do not know their positions. Each agent (correctly) believes that they are in each position $1, \ldots, K$ with equal probabilities if they do not observe a news feed, and in each position $K + 1, \ldots, n$ with equal probabilities if they do observe a news feed. The informational environment is common knowledge.

2.1 Discussion and Interpretation

We view the news stories s_i as original content that users discover from external sources (e.g., opinion pieces from local newspapers, new scientific preprints, etc.) and post on the social media platform (e.g., Twitter, Facebook, or Reddit). The platform presents each user with a news feed of stories that others have posted, and gives users some way of expressing endorsement for the content discovered by others. What we generically refer to as "sharing" in our model corresponds to retweeting on Twitter, reposting content you saw on Facebook on your own timeline, upvoting a story on Reddit, and so forth.

The platform chooses what content to show in the news feed. It could focus on showing more viral content (larger λ) or more "random" content (smaller λ). Displaying "random" content could represent, for example, showing a user the most recent stories that their friends posted without regard for the stories' popularity score. We choose a simple functional form to model this trade-off, but others are also possible. In particular, one could also analyze more extreme sampling rules that place more weight on more viral stories than popularity-based sampling (e.g., the probability of sampling a story is proportional to the square of its popularity).

The virality of the news feed is a design choice that social-media companies pay a lot of attention to in practice. Over the years, different iterations of the Twitter feed gave different levels of emphasis on the trending or most popular tweets on the platform. For Reddit, its ordering algorithm for displaying posts on the front page evolved over a decade. An entry on Reddit's company blog in 2009 shows the platform designers are well aware of the disadvantages of putting too much emphasis on the most highly upvoted content:

> “Once a comment gets a few early upvotes, it’s moved to the top. The higher something is listed, the more likely it is to be read (and voted on), and the more votes the comment gets. It’s a feedback loop” (Munroe, 2009).

We assume an explicit capacity constraint C on how much people can share. The primary reason is tractability, as the analysis is considerably cleaner when the number of stories shared does not depend on the realizations of the sampled stories. A capacity constraint also captures the fact that people tend to only share a small fraction of the content that they consume on social-media platforms.[4] Note that even if the agents are not forced to share exactly C stories out of the K in their news feeds, they would still find it optimal to share up to the capacity constraint. This is because they incur no marginal cost from sharing and no penalty from sharing an incorrect story, and they will always believe that each story has a positive probability of being correct.

We also assume that people do not know their order in the sequence. This is likely more realistic than assuming that everyone knows their precise order. From a technical perspective, it is also the more tractable assumption that lets us focus on analyzing long-run steady states. All agents after K use the same strategy when choosing what to share, as they cannot assign different interpretations to a given news feed based on their positions. This means fixing a strategy, we just need to analyze the stochastic process of the relative popularity scores of the positive and negative stories. Since agents do not know their positions, we do not need to consider how the evolution of this process changes optimal strategies for agents in different positions.

2.2 Strategy, Symmetric BNE in Finite Societies, and Social Equilibrium

An agent who does not see a news feed has no decisions to make. We therefore define a mixed strategy in the game to be $\sigma : \{-1, 1\} \times \{0, ..., K\} \to \Delta(\{0, 1, ..., C\})$, so that $\sigma(s, k)$ gives the distribution over the number of positive stories shared when the agent discovers the story s and sees a news feed with k positive stories out of K. We will regard the space of strategies as a subset of $\mathbb{R}^{C+1}$ with the standard norm. Mixed strategies must satisfy feasibility constraints in terms of

[4] For example, even among the top 10% of tweeters on Twitter, the median number of favorites per month is 70 and the median number of tweets per month is 138, which are presumably much lower than the number of tweets that these users read per month (Wojcik and Hughes, 2019).

the available numbers of positive and negative stories to share, so $\sigma(s, k)$ cannot have values larger than k or smaller than $C + k - K$ in its support for any $0 \leq k \leq K$. Note that we only need to discuss positive stories since the agent must always share C stories in total.

A simple strategy, which will play a central role in our analysis, is to follow the majority of the stories in the news feed (breaking ties in favor of the private signal):

Definition 1. The *majority rule* is the strategy defined by $\sigma^{\text{maj}}(s, k)(C) = 1$ if either $k > K/2$ or $k = K/2$ and $s = 1$, and $\sigma^{\text{maj}}(s, k)(0) = 1$ otherwise.

The majority rule is a pure strategy that either shares C stories with the realization of 1 or C stories with the realization of -1, depending on the news-feed majority. Note the majority rule is feasible because we have assumed $C \leq K/2$.

We apply the solution concept of Bayesian Nash equilibrium (BNE). Note that all possible observations are on-path given any strategy profile. We focus on player-symmetric and state-symmetric BNE: that is, a BNE where each agent uses the same strategy σ, and σ treats positive and negative stories symmetrically.[5] We abbreviate this refinement as "symmetric BNE."

We are mainly interested in analyzing the limits of symmetric BNE when the number of agents on the platform grows large, and studying the accuracy of the resulting news feeds in the long run. Such a "limit" is well defined because for fixed parameters q, K, C, λ, the space of strategies stays constant as the number of agents n grows.

Definition 2. For fixed q, K, C, λ parameters, a mixed strategy σ^* is a *social equilibrium* if there exists a sequence of symmetric BNE $(\sigma^{(j)})_{j=1}^{\infty}$ for finite societies with n_j agents and the same q, K, C, λ parameters, where $n_j \to \infty$ and $\lim_{j \to \infty} \sigma^{(j)} = \sigma^*$.

Symmetric BNE and social equilibria both exist for all parameter values:

Proposition 1. *For any finite n and parameters q, K, C, λ, there exists a symmetric BNE. For any parameters q, K, C, λ, there exists a social equilibrium.*

Using a standard fixed-point argument, we can show that there is a symmetric BNE for all n. Since the space of feasible mixed strategies can be viewed as a compact, convex subset of a finite-dimensional Euclidean space, a social equilibrium must exist.

[5] More precisely, state symmetry means that for every $s \in \{-1, 1\}$ and $0 \leq k \leq K$, we have $\sigma(s, k)(z) = \sigma(-s, K - k)(C - z)$ for each $0 \leq z \leq C$.

3 Steady States and Equilibrium Steady States

We begin this section by describing the structure of steady states under a fixed strategy σ. We then distinguish informative and misleading steady states. The third subsection describes equilibrium strategies and the structure of steady states under equilibrium behavior.

3.1 Definition and Characterization of Steady States

Suppose everyone uses the same strategy σ, which needs not be an equilibrium, and the true state is ω. How will the total popularity score of the correct stories that match the state compare with the popularity score of the incorrect stories in the long run? We define the concept of steady states to study this question.

Given the true state of nature ω, a finite set of t stories $(s_1, ..., s_t)$ on the platform and the popularity scores of these stories $\rho(s_i)$, the *viral accuracy* of the platform is defined to be $\frac{\sum_{i:s_i=\omega} \rho(s_i)}{\sum_{j=1}^{t} \rho(s_j)}$. This metric shows the relative popularity of the stories that match the true state. Imagine a society with infinitely many agents, with all agents $i \geq K+1$ using the strategy σ. This induces a stochastic process $(x(t))_{t=1}^{\infty}$ where $x(t)$ is the viral accuracy of the platform after agent t has acted.

Definition 3. A point x^* such that $x(t) \to x^*$ with positive probability is a *steady state* of strategy σ.

When viral accuracy converges to a steady state x^*, roughly x^* fraction of the total popularity score on the platform is associated with correct stories in all late enough periods. This fraction persists as fresh stories get posted on the platform each period and agents use σ to decide which stories to share from their randomly generated news feeds. The next result tells us that for any state-symmetric strategy, viral accuracy almost surely converges, and the set of steady states X^* is finite.

Proposition 2. *When everyone uses the state-symmetric strategy σ, there exists a finite set of steady states $X^* \subseteq (0, 1)$ such that almost surely $x(t) \to x^*$ for some $x^* \in X^*$.*

The proof uses a convergence result from stochastic approximation (Theorem 2 from Chapter 2 of Borkar, 2009). When X^* contains at least two elements, the limit steady state $x^* \in X^*$ is random and can depend on the early agent's private signal realizations and the random sampling involved in generating news feeds.

The basic idea behind the proof of Proposition 2, as well as much of the subsequent analysis, is that we can describe the state of the system in terms of $x(t)$ and the fraction of stories up to time t that match the state. We can decompose the change $x(t+1) - x(t)$ between periods into a deterministic term (which we will approximate using the inflow accuracy function defined below) and a martingale noise term. Given this decomposition, a martingale convergence theorem shows that $x(t)$ converges almost surely.

In light of Proposition 2, we write $\pi(\cdot \mid \sigma)$ for the distribution over steady states generated by a state-symmetric strategy σ.[6] A substantial challenge in analyzing our model is that we cannot obtain closed-form expressions for $\pi(\cdot \mid \sigma)$ as these probabilities depend on a complicated stochastic process. We focus instead on understanding when the support of $\pi(\cdot \mid \sigma)$, that is the set of steady states, contains certain values of x. We will see this question is already highly non-trivial, and the answers will have interesting implications for understanding equilibrium and design questions.

We begin by describing the support of $\pi(\cdot \mid \sigma)$. Suppose today's viral accuracy is x, and exactly q fraction of the stories on the platform are correct. A new agent must increase the total popularity score on the platform by $C+1$, as they share C existing stories and post a new story. Let $\phi_\sigma(x)$ represent the expected fraction of the incoming $C+1$ popularity score that will be allocated to stories that match the state.

Definition 4. The *inflow accuracy function* is

$$\phi_\sigma(x) := \frac{q + \sum_{k=0}^{K} P_k(x,\lambda) \cdot [q \cdot \mathbb{E}[\sigma(1,k)] + (1-q) \cdot \mathbb{E}[\sigma(-1,k)]]}{1+C}$$

where $P_k(x,\lambda) := \mathbb{P}[\mathrm{Binom}(K, \lambda x + (1-\lambda)q) = k]$ and $\mathrm{Binom}(K,p)$ is the binomial distribution with K trials and success probability p.

To understand the formula in the definition, note that $\lambda x + (1-\lambda)q$ is the *sampling accuracy* of the platform: the probability of each news-feed story being correct, given viral accuracy x and viral weight λ. We can use sampling accuracy to express the probability of getting k positive stories

[6] By symmetry of σ and of the environment, it is clear that the distribution over steady states is the same conditioning on $\omega = 1$ and $\omega = -1$.

out of K in the news feed when $\omega = 1$ for every $0 \leq k \leq K$, then consider how the strategy σ combines the private signal s_i and the number of positive stories in the news feed to make a sharing decision. Finally, $\phi_\sigma(x)$ also takes into account that the story posted by the agent, which starts with a popularity score of 1, has q chance of matching the state. While the definition of $\phi_\sigma(x)$ only describes the expected fraction of the new popularity score assigned to correct stories when $\omega = 1$, the symmetry of the environment and of σ implies that it also describes the same fraction when $\omega = -1$.

We always have $\phi_\sigma(0) > 0$ and $\phi_\sigma(1) < 1$. The idea is that if $x \approx 0$ and almost all of the popularity score are associated with the wrong stories, then the arrival of fresh stories posted by new agents tends to increase x, as a majority of these stories match the state. If on the other hand $x \approx 1$, then these fresh stories will on average lower x, since they have a non-zero probability of mismatching the state. So ϕ_σ must have a fixed point by continuity.

A fixed point of the inflow accuracy function ϕ_σ is a natural candidate for a steady state induced by σ, as it intuitively represents a level of viral accuracy that tends to be exactly maintained on average by the inflow of new popularity score, on a platform with sufficiently many stories so that approximately q fraction of them match the true state. The next result establishes this formally, provided the fixed point is not unstable from both sides.

Theorem 1. *We have $\pi(x^* \mid \sigma) > 0$ if $\phi_\sigma(x^*) = x^*$ and there exists some $\epsilon > 0$ so that either (a) $\phi_\sigma(x) < x$ for all $x \in (x^*, x^* + \epsilon)$, or (b) $\phi_\sigma(x) > x$ for all $x \in (x^* - \epsilon, x^*)$. Conversely, for $x^* \in [0, 1]$, we have $\pi(x^* \mid \sigma) > 0$ only if $\phi_\sigma(x^*) = x^*$.*

The most interesting case is a fixed point of ϕ_σ that is unstable on one side (see Figure 1 for an illustration). A *touchpoint* of ϕ_σ is a fixed point $x^* = \phi_\sigma(x^*)$ where exactly one of condition (a) or condition (b) from Theorem 1 holds (so x^* is only stable from one side). Theorem 1 implies in particular that if ϕ_σ has a touchpoint, then viral accuracy converges there with positive probability. This means the distribution over steady states is discontinuous in the strategies that agents use and discontinuous in parameters of the model such as λ and q. The points of discontinuity are those such that ϕ_σ admits a touchpoint, since the probability of converging to the touchpoint steady state jumps at these points. While inflow accuracy functions that admit touchpoints may seem non-generic, we will see that they play critical roles in classifying social equilibria based on

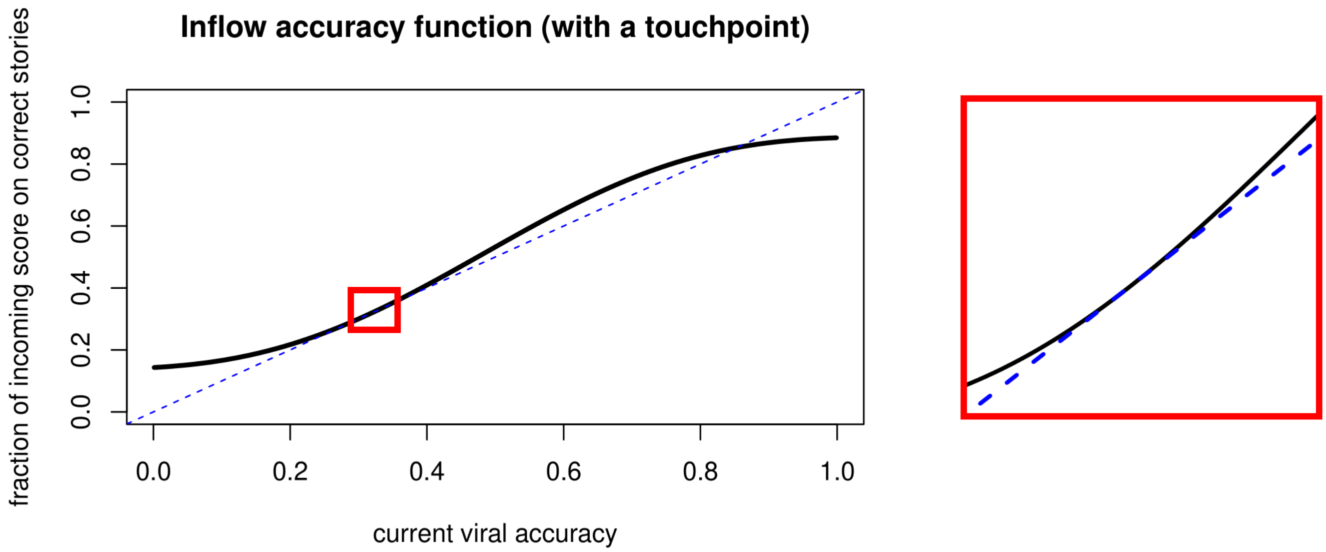

Figure 1: The inflow accuracy function for the majority rule with $K = 7$, $C = 3$, $q = 0.55$, $\lambda \approx 0.76$. Here $\phi_{\sigma^{\mathrm{maj}}}$ has two fixed points: the left fixed point is a touchpoint that is only stable from the left side (the red box shows a zoomed-in view). The right fixed point is stable from both sides. Theorem 1 implies viral accuracy has a positive probability of converging to each of these two fixed points.

the steady states they induce, and that they arise endogenously in some platform optimization problems.

One might expect that the stochastic process of viral accuracy should not converge with positive probability to fixed points of ϕ_σ which are unstable from one side, because random noise in the process $x(t)$ (due to the randomness in news-feed story sampling and private signal realizations) can bring it to the unstable side of the fixed point. But careful analysis shows that there is a positive probability event that $x(t)$ converges to the touchpoint x^* while entirely staying on the stable side, with the noise terms never being large enough to move the process over to the unstable side. The proof extends the techniques of Pemantle (1991), which shows a similar result for generalized Pólya urns.

3.2 Informative and Misleading Steady States

We may classify steady states into two types. One type is an informative steady state, where sampling accuracy is above 1/2 and it is more likely that stories in the news feed are true. The other type is a misleading steady state, where the opposite happens.

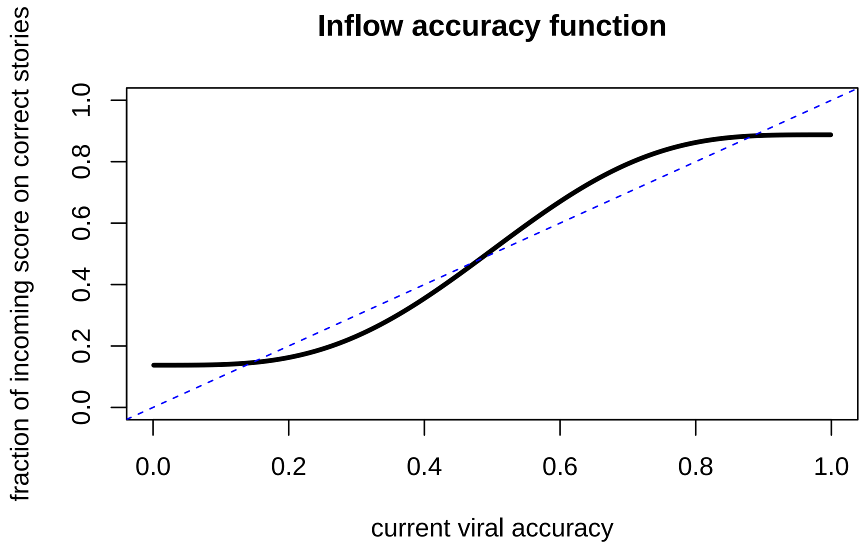

Figure 2: The inflow accuracy function for the majority rule with $K = 7$, $C = 3$, $q = 0.55$, $\lambda = 1$.

Definition 5. A steady state x is *informative* if $\lambda x + (1-\lambda)q \geq 1/2$, and *strictly informative* if this inequality is strict. A steady state x is *misleading* if $\lambda x + (1-\lambda)q \leq 1/2$, and *strictly misleading* if this inequality is strict.

Even reasonable strategies like the majority rule σ^{maj} can generate misleading steady states. As shown in Figure 2, σ^{maj} has two steady states with the parameters $K = 7$, $C = 3$, $q = 0.55$, and $\lambda = 1$ (note the middle fixed point of the inflow accuracy function $\phi_{\sigma^{\text{maj}}}$ is unstable from both sides and thus not a steady state of majority rule). One is informative, but the other is misleading.

In a misleading steady state, the virality of false stories become self-sustaining. The state might be $\omega = 1$ but most people see negative stories in their news feeds, as the platform's virality weight implies the popular false stories tend to get shown to users. This is despite the fact that the total number of negative stories is smaller than the total number of positive stories on the platform. Under the majority rule σ^{maj}, agents will then share the negative stories from their news feeds, which further perpetuates these stories' popularity and makes them more likely to be seen by future agents.

It turns out that $\phi_{\sigma^{\text{maj}}}$, the inflow accuracy function associated with the majority rule, plays an important role in determining the equilibrium steady states of *any* social equilibrium. The steady states of the majority rule σ^{maj} satisfy the following useful property:

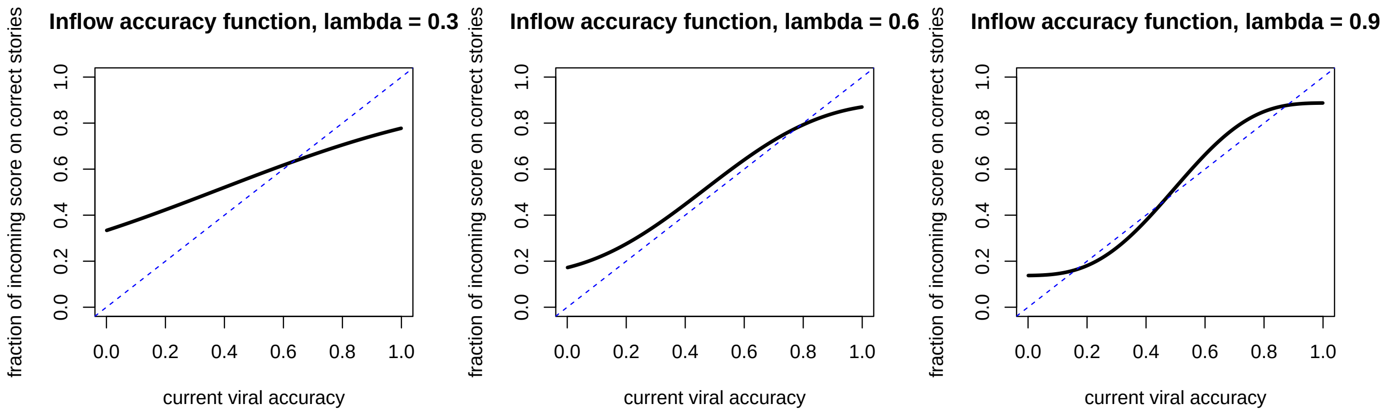

Figure 3: The inflow accuracy function for the majority rule with $q = 0.55$, $K = 7, C = 3$, and $\lambda \in \{0.3, 0.6, 0.9\}$. With $\lambda = 0.3$ and $\lambda = 0.6$, there is a single informative steady state. With $\lambda = 0.9$, a misleading steady state appears.

Lemma 1. *If x is a steady state of σ^{maj}, then it is strictly informative if and only if $x > 1/2$, and strictly misleading if and only if $x < 1/2$. Also, $x = 1/2$ is not a fixed point of $\phi_{\sigma^{maj}}$.*

Steady states are generally classified as informative or misleading depending on the sampling accuracy. The lemma says for the majority rule, we can equivalently classify steady states based on whether viral accuracy is larger than 1/2.

The number of steady states for a fixed strategy depends on λ. The three figures below plot the inflow accuracy function with $q = 0.55$, $K = 7, C = 3$ and the majority rule. The three plots in Figure 3 correspond to three different values of λ: $\lambda = 0.3$, $\lambda = 0.6$, and $\lambda = 0.9$. When $\lambda = 0.3$ and $\lambda = 0.6$, there is only an informative steady state, and this steady state is more accurate when $\lambda = 0.6$. But when $\lambda = 0.9$, there is both an informative steady state and a misleading steady state.

3.3 Equilibrium Steady States

So far, we have discussed steady states associated with arbitrary strategies. We are mainly interested in *equilibrium steady states*, that is the distribution $\pi(\cdot \mid \sigma^*)$ when σ^* is a social equilibrium strategy.

We now define the critical virality weight, which is the smallest λ for which there is a misleading steady state under the majority rule. We will see the set of equilibrium steady states changes sharply around this critical value of λ.

Definition 6. The *critical virality weight* λ^* is

$$\lambda^* := \inf\{\lambda \in [0,1] : \phi_{\sigma^{\mathrm{maj}}}(x^*) = x^* \text{ for some } x^* \in [0,1/2]\},$$

provided this set is non-empty. Otherwise, we let $\lambda^* = 1$.

Depending on the values of the parameters q, K, C, it turns out that either σ^{maj} only has strictly informative steady states for any viral weight (so $\lambda^* = 1$), or there is some smallest $0 < \lambda^* \leq 1$ where a fixed point in $[0,1/2]$ first appears for $\phi_{\sigma^{\mathrm{maj}}}$. For instance, for $q = 0.55, K = 7, C = 3$, Figure 1 shows that $\lambda^* \approx 0.76$. The next theorem fully characterizes when misleading steady states exist across all social equilibria for every level of λ.

Theorem 2. *For $0 < \lambda \leq \lambda^*$, the unique social equilibrium is σ^{maj}. At every $\lambda < \lambda^*$, σ^{maj} only has one equilibrium steady state, and it is strictly higher than q (thus, strictly informative). For $\lambda = \lambda^*$, provided $\lambda^* < 1$, the unique social equilibrium σ^{maj} induces a strictly misleading steady state. For $\lambda > \lambda^*$, every social equilibrium induces at least one strictly misleading steady state.*

This result shows how the platform's virality weight affects the types of equilibrium steady states: there are only informative equilibrium steady states when $\lambda < \lambda^*$, and that there will always be misleading equilibrium steady states when $\lambda \geq \lambda^*$ (as long as $\lambda^* < 1$).[7] It also shows the majority rule is the only possible social equilibrium for non-zero virality weights below the critical virality weight λ^*.[8] For virality weights above λ^*, the majority rule may not be a social equilibrium, and there may be multiple social equilibria. Nevertheless, the result tells us that every social equilibrium has a positive probability of generating a misleading steady state where false stories dominate news feeds.

Theorem 1 and Theorem 2 together imply a discontinuity in equilibrium learning outcomes at λ^*. For λ just below λ^*, we converge almost surely to a steady state where a majority of users believe the true state is more likely. But when $\lambda = \lambda^*$, there is a positive probability of converging to a misleading steady state. This tells us that learning outcomes can be very sensitive to design choices.

[7] We will see in Proposition 5 that λ^* is not equal to one whenever K and C are large enough, so misleading steady states can indeed arise.

[8] When $\lambda = 0$, the only possible equilibrium steady state is q, so every story in the news feed is exactly as informative as one's private signal. There is thus some degree of freedom in tie-breaking when there is one more positive story than negative story in the news feed and one's private signal is negative.

The theorem greatly simplifies checking whether there is a misleading steady state at a social equilibrium under given parameter values. Without the theorem, checking for misleading steady states would require solving for equilibrium strategies, which is a complicated calculation depending on $\pi(\cdot|\sigma)$ and therefore the entire stochastic process. The theorem says we can instead check for misleading steady states under the majority rule, which are simply roots of a polynomial. This reduction relies on two properties, established in the proof: (1) if any strategy sustains a misleading steady state, then the majority rule does too,[9] and (2) when there are no misleading steady states, the majority rule is a best response if the number of agents n in the society is sufficiently large.

We note that when $\lambda > \lambda^*$, the equilibrium strategy may not be simple to characterize. A particular challenge is that it is not clear beliefs need to be monotonic in observations. If there is one informative steady state and one misleading steady state, then viral accuracy is further from 1/2 at the informative steady state than the misleading steady state. So a slight majority of negative stories in a news feed could actually be more indicative of a true state of $\omega = 1$ with the platform trapped in a misleading steady state, rather than a true state of $\omega = -1$ with the platform in an informative steady state. In this case the agent's utility-maximizing action might be to share the positive stories, which make up the minority side of the news feed.

Our next result says that larger a virality weight can lead to a stronger consensus:

Proposition 3. *For $\lambda \in [0, \lambda^*)$, the unique steady state x^* at the unique social equilibrium under λ is strictly increasing in λ.*

In the region of virality weights that do not generate misleading equilibrium steady states, increasing λ allows more information aggregation. This is because a positive story in i's news feed not only tells i about the realization of a single signal, but also lets i draw inferences about the hidden information available to i's predecessors who may have chosen to share that positive story. As λ increases, the sampled news-feed stories are closer to indicating a consensus among many agents.

We state the proposition for $\lambda < \lambda^*$, but a similar argument also shows that increasing λ tends to increase consensus in other regions as well. Formally, consider any interval $(\underline{\lambda}, \overline{\lambda})$ on which the

[9] This makes use of the capacity constraint model of sharing. If the number of stories shared depended on the realization of the sampled signals, it seems plausible there could be a misleading steady state under a strategy that shares fewer stories than the majority rule but not under the majority rule.

set of social equilibria and the set of steady states under each equilibrium are unchanged. Fix a social equilibrium σ^* and a continuous selection $x^*(\lambda)$ of a steady state for each $\lambda \in (\underline{\lambda}, \overline{\lambda})$. Then $|x^*(\lambda) - \frac{1}{2}|$ is strictly increasing in λ on the interval $(\underline{\lambda}, \overline{\lambda})$. Informative steady states become more accurate, while misleading steady states induce a stronger wrong consensus.*

Theorem 2 and Proposition 3 together formalize the trade-off in the virality weight λ described in the introduction. Increasing λ initially increases the steady-state viral accuracy and sampling accuracy. But starting at a critical threshold λ^*, it discontinuously creates the social form of confirmation bias discussed in the introduction, which we have now formalized in terms of a misleading steady state.

3.4 Comparative Statics

Theorem 2 characterizes the critical virality weight λ^* as the smallest value of λ such that the polynomial $\phi_{\sigma^{\text{maj}}}(x)$ has a touchpoint. We now use this characterization to show how the critical virality weight changes with respect to parameters of the environment. A lower critical virality weight means an equilibrium misleading steady state exists for more news-feed designs, so these comparative statics results point to properties of the environment and the social-media platform that tend to create situations where false news stories dominate news feeds.

Proposition 4. *Let $\lambda^*(q, K, C)$ be the critical virality weight for parameters (q, K, C).*

- $\lambda^*(q', K, C) \geq \lambda^*(q, K, C)$ *if* $q' > q$
- $\lambda^*(q, K, C') \geq \lambda^*(q, K, C)$ *if* $C' < C$
- $\lambda^*(q, K-2, C) \geq \lambda^*(q, K, C)$ *for every* K
- $\lambda^*(q, K+1, C) \geq \lambda^*(q, K, C)$ *if* K *is odd*
- $\lambda^*(q, K-1, C) \geq \lambda^*(q, K, C)$ *if* K *is odd*

All of these inequalities are strict whenever $\lambda^*(q, K, C) < 1$.

The overall message is that misleading steady states emerge when individuals consume and interact with a large amount of social information on the platform relative to the amount of private information that they have from other sources. It is easier to generate a misleading equilibrium steady state when individual stories are noisy signals of the state of nature (so there is little private information), when users can share a large number of stories in their news feeds, and when users read many stories on the social-media platform. Intuitively, since each individual's private information is untainted by the popularity of various stories on the social-media platform, a sufficient amount of high-quality private information can counteract the harmful effects of trending false stories in news feeds. On the other hand, if users spend a great deal of time on the platform, browsing many stories in their news feeds and sharing many of the stories they read, then they are more likely to fall prey to a misleading steady state and help perpetuate the virality of inaccurate news.

The comparative statics in K depend on the parity of K. Misleading steady states are less likely for K even because agents with symmetric social observations (i.e., $K/2$ news-feed stories matching each state) follow their private signals, which match the true state more frequently. Aside from this detail, however, increasing K does create misleading steady states for a larger set of parameters.

While Proposition 4 tells us the direction in which λ^* changes as K and C increase, we may still want to know the extent to which these two parameters can affect the critical virality weight. The next result shows the full range of possible critical virality weight values when we fix the story precision and change how many news stories people read and share.

Proposition 5. *For any q, K, C, we have $\lambda^*(q, K, C) > 1 - \frac{1}{2q}$. But for any fixed $1/2 < q < 1$ and any $\overline{\lambda} > 1 - \frac{1}{2q}$, there exist $\underline{K}, \underline{C}$ so that whenever $K \geq \underline{K}$, $C \geq \underline{C}$, we have $\lambda^*(q, K, C) \leq \overline{\lambda}$.*

Proposition 5 says that if users have large enough news feeds and share sufficiently many stories from their feeds, then the critical virality weight will be arbitrarily close to $1 - \frac{1}{2q}$. In particular, no matter how precise the individual stories, every social equilibrium admits a misleading steady state when $\lambda \geq 1/2$ provided K and C are large enough. But $1 - \frac{1}{2q}$ is also a sharp lower bound on the critical virality weight, so news-feed algorithms close enough to uniform sampling remain immune to the possibility of misleading steady states even if we let users access and interact with more and more social information.

4 Optimal News-Feed Design

The characterization of equilibrium steady states from the previous section lets us ask about optimal design of the news feed for different objective functions. We consider a designer who wants to maximize users' equilibrium utility from sharing stories on the platform, for this utility is likely a major component of the value that people derive from the platform and affects their decision to continue using it. But the basic logic of our results also extends to other platform objectives (such as the expected viral accuracy or the degree of consensus on the platform).

Given a fixed finite number n of users, choose a measurable selection $\sigma^*(\lambda)$ of symmetric BNE for each value of λ. Suppose the platform wants to choose λ to maximize the users' expected utility from sharing:

$$W_n(\lambda) = \frac{1}{n-K} \sum_{i=K+1}^{n} \mathbb{E}_{\sigma^*(\lambda)} \left[\sigma^*(\tilde{s}_i, \tilde{k}_i) \cdot u \mid \omega = 1 \right]$$

where the conditional expectation is taken over the realizations of i's story $\tilde{s}_i$ and the number of positive stories $\tilde{k}_i$ in i's news feed, when all agents use the strategy $\sigma^*(\lambda)$ and the state is $\omega = 1$.

In the region $\lambda < \lambda^*$, Theorem 2 says increasing λ will improve viral accuracy and sampling accuracy at the informative steady state without introducing misleading steady states. This improves users' expected sharing utility, because they are more likely to share stories that match the true state. The following proposition says that the optimal λ must be either close to λ^* or above λ^* for the designer's problem in large but finite societies.

Proposition 6. *For any sequence (λ_n) where each λ_n maximizes $W_n(\lambda)$, we have*

$$\liminf_{n\to\infty} \lambda_n \geq \lambda^*.$$

Suppose we are in an environment where $\lambda^* < 1$, so misleading steady states exist for some news-feed designs. The platform endogenously chooses virality weights λ to be either above this critical virality weight or just below it. In the first case, misleading steady states occur under the platform's preferred choice of λ. These steady states appear even though any misleading steady state leads to less agreement, and therefore lower utility for users, than the most informative steady state under the same equilibrium. In the second case, the platform engages in "brinkmanship," choosing λ just below a threshold that corresponds to a discontinuous drop in welfare. This makes the learning outcome fragile, as a small unanticipated shock (e.g., a minuscule decrease in the story precision q) can cause a large shift in people's long-run beliefs.

References

Acemoglu, D., A. Ozdaglar, and J. Siderius (2022): "A model of online misinformation," *Working Paper*.

Arieli, I., Y. Babichenko, and M. Mueller-Frank (2022): "Aggregate implications of probability matching for social learning outcomes," *Working Paper*.

Banerjee, A. and D. Fudenberg (2004): "Word-of-mouth learning," *Games and Economic Behavior*, 46, 1–22.

Banerjee, A. V. (1992): "A simple model of herd behavior," *Quarterly Journal of Economics*, 107, 797–817.

Bikhchandani, S., D. Hirshleifer, and I. Welch (1992): "A theory of fads, fashion, custom, and cultural change as informational cascades," *Journal of Political Economy*, 100, 992–1026.

Bloch, F., G. Demange, and R. Kranton (2018): "Rumors and social networks," *International Economic Review*, 59, 421–448.

Borkar, V. S. (2009): *Stochastic Approximation: A Dynamical Systems Viewpoint*, vol. 48, Springer.

Bowen, R., D. Dmitriev, and S. Galperti (2022): "Learning from shared news: When abundant information leads to belief polarization," *Working Paper*.

Buechel, B., S. Klössner, F. Meng, and A. Nassar (2022): "Misinformation due to asymmetric information sharing," *Working Paper*.

Hill, B. M., D. Lane, and W. Sudderth (1980): "A strong law for some generalized urn processes," *The Annals of Probability*, 214–226.

Hsu, C.-C., A. Ajorlou, and A. Jadbabaie (2021): "Persuasion, news sharing, and cascades on social networks," *Working Paper*.

Kabos, E. and M. Meyer (2021): "A welfare analysis of a steady-state model of observational learning," *Working Paper*.

Kranton, R. and D. McAdams (2022): “Social connectedness and the market for information,” *Working Paper*.

Lévy, R., M. Pęski, and N. Vieille (2022): “Stationary social learning in a changing environment,” *Working Paper*.

Merlino, L. P., P. Pin, and N. Tabasso (2022): “Debunking rumors in networks,” *American Economic Journal: Microeconomics (Forthcoming)*.

Munroe, R. (2009): “Reddit’s new comment sorting system,” *The Reddit Blog*.

Papanastasiou, Y. (2020): “Fake news propagation and detection: A sequential model,” *Management Science*, 66, 1826–1846.

Pemantle, R. (1991): “When are touchpoints limits for generalized Pólya urns?” *Proceedings of the American Mathematical Society*, 113, 235–243.

Pennycook, G., Z. Epstein, M. Mosleh, A. A. Arechar, D. Eckles, and D. G. Rand (2021): “Shifting attention to accuracy can reduce misinformation online,” *Nature*, 592, 590–595.

Pennycook, G., J. McPhetres, Y. Zhang, J. G. Lu, and D. G. Rand (2020): “Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention,” *Psychological Science*, 31, 770–780.

Wojcik, S. and A. Hughes (2019): “Sizing up Twitter users,” *Pew Research Center*.

A Proofs

A.1 Proof of Proposition 1

Proof. Fixing n gives a symmetric finite game. So by Kakutani’s fixed point theorem, there exists a symmetric BNE.

Now fix q, K, C, and λ. For each n, there exists a symmetric BNE $\sigma^{(n)}$. Because the space of strategies σ is compact, we can choose a convergent subsequence. The limit of this subsequence is a social equilibrium. □

A.2 Proof of Proposition 2

Proof. The proof applies a convergence result from stochastic approximation from Chapter 2 of Borkar (2009). Suppose agents use strategy σ. Without loss of generality, we can condition on $\omega = 1$.

Let

$$Y = \{\mathbf{y} = (x, z) \in [0,1]^2\}.$$

Recall $\rho_t(s_i)$ is the popularity of signal s_i after agent t acts. For each t, define $\mathbf{y}(t) \in Y$ by

$$x(t) = \frac{\sum_{s_i=1} \rho_t(s_i)}{(C+1)t} \text{ and } z(t) = \frac{\sum_{s_i=1} 1}{t}.$$

The first entry of $\mathbf{y}(t) = (x(t), z(t))$ measures the fraction of shares which are shares of signals with realization 1 up to time t. The second entry measures the fraction of private signals which have realization 1 up to time t.

We can write

$$\mathbf{y}(t+1) = \mathbf{y}(t) + \frac{1}{t+1}\left(\xi(t+1) - \mathbf{y}(t)\right),$$

where $\xi(t+1)$ is the random variable with first entry equal to the fraction of shared signals which have realization 1 in period $t+1$ and second entry is a binary indicator for whether $s_{t+1} = 1$.

Following the notation of Borkar (2009), we write

$$h(\mathbf{y}(t)) = \mathbb{E}\left[\xi(t+1) \,|\, \mathbf{y}(t)\right] - \mathbf{y}(t) \text{ and } M(t+1) = \xi(t+1) - \mathbb{E}\left[\xi(t+1) \,|\, \mathbf{y}(t)\right].$$

We can then decompose the change in the stochastic process $\mathbf{y}(t)$ into $h(\mathbf{y}(t))$, which depends deterministically on $\mathbf{y}(t)$, and $M(t+1)$, which is a martingale:

$$\mathbf{y}(t+1) = \mathbf{y}(t) + \frac{1}{t+1}\left(h(\mathbf{y}(t)) + M(t+1)\right).$$

We would like to apply Theorem 2 of Chapter 2 of Borkar (2009), which requires the following assumptions:

(A1) h is Lipschitz continuous.

(A2) $\sum_t \frac{1}{t+1} = \infty$ while $\sum_t \frac{1}{(t+1)^2} < \infty$.

(A3) $\mathbb{E}\left[M(t+1) \mid \mathbf{y}(t)\right] = 0$ and $\{M(t)\}$ are square-integrable with

$$\mathbb{E}\left[\|M(t+1)\|^2 \mid \mathbf{y}(t)\right] \leq \kappa(1 + \|\mathbf{y}(t)\|^2)$$

a.s. for all n and some $\kappa > 0$.

(A4) $\|\mathbf{y}(t)\|$ remains bounded a.s.

Properties **(A2)** and **(A4)** are immediate. For **(A3)**, the martingale property holds by the construction of $M(t)$ and the remaining properties hold because $M(t)$ is bounded (independent of t).

Property **(A1)** remains. Since $-\mathbf{y}(t)$ is Lipschitz continuous in $\mathbf{y}(t)$, we must check that $\mathbb{E}\left[\xi(t+1) \mid \mathbf{y}(t)\right]$ is Lipschitz continuous in $\mathbf{y}(t)$.

Write $\sigma_1(s,k)$ for the expected number of "1" signals that strategy σ shares, when the agent's private signal is s and k signals in the sample are "1". Write $P_k(x,z,\lambda)$ for $\mathbb{P}[\text{Binom}(K, \lambda x + (1-\lambda)z) = k]$, where $\text{Binom}(k,p)$ is the binomial distribution with k trials and success probability p.

Then for $t \geq K$, the conditional expectation of the random variable $\xi_1(t+1)$ is equal to

$$\frac{1}{C+1}\left(q + \sum_{0 \leq k \leq K} P_k(x(t), z(t), \lambda)(q\sigma_1(1,k) + (1-q)\sigma_1(-1,k))\right).$$

This is a polynomial of degree at most K in $x(t)$ and $z(t)$, and therefore is Lipschitz continuous on Y. The conditional expectation of $\xi_2(t+1)$ is constant, and therefore Lipschitz continuous in Y as well.

We can define a continuous-time differential equation by letting

$$\dot{\mathbf{r}}(t) = h(\mathbf{r}(t)), t \geq 0. \tag{1}$$

An invariant set A of (1) is a set such that $\mathbf{r}(0) \in A$ implies $\mathbf{r}(t) \in A$ for all $t \geq 0$. An invariant set is internally chain transitive if for any $\mathbf{r}, \mathbf{r}' \in A$, $\epsilon > 0$ and $T > 0$, there exists $\mathbf{r}^0 = \mathbf{r}, \mathbf{r}^1, \ldots, \mathbf{r}^n = \mathbf{r}' \in A$ such that the trajectory of $\mathbf{r}(t)$ starting from $\mathbf{r}(0) = \mathbf{r}^i$ meets with an ϵ-neighborhood of $\mathbf{r}^{i+1}$ at some time $t \geq T$.

By Theorem 2 of Chapter 2 of Borkar (2009), the stochastic process $\mathbf{y}(t)$ converges to an internally chain transitive invariant set of equation (1). Because $r_2(t) \to q$, any internally chain transitive invariant set must be contained in $[0,1] \times \{q\}$. We claim that at any $\mathbf{r}$ contained in an internally chain transitive invariant set A, we must have

$$\frac{dr_1(t)}{dt} = 0$$

when $\mathbf{r}(t) = \mathbf{r}$. Suppose an internally chain transitive invariant set A of (1) contains a point $\mathbf{r}$ at which

$$\frac{dr_1(t)}{dt} > 0.$$

Letting $\mathbf{r}(0) = \mathbf{r}$, we can choose some $t' > 0$ such that $r_1(t') > r_1(0)$ and

$$\frac{dr_1(t)}{dt} > 0$$

at $t = t'$. Now let $\mathbf{r}' = \mathbf{r}(t')$. We have $\mathbf{r}' \in A$ by invariance.

If we consider the trajectory $\mathbf{r}(t)$ beginning with $\mathbf{r}(0) = \mathbf{r}'$, we cannot have $r_1(t)$ fall below $r_1(0)$ since $\dot{r}_1(0) > 0$ and the sign of $\dot{r}_1(t)$ depends only on $r_1(t)$. For $\epsilon > 0$ sufficiently small this implies that the trajectory $\mathbf{r}(t)$ beginning with $\mathbf{r}(0) = \mathbf{r}'$ never enters an ϵ-neighborhood of $\mathbf{r}$. This contradicts the assumption that A is internally chain transitive.

If A contains a point $\mathbf{r}$ at which

$$\frac{dr_1(t)}{dt} < 0,$$

we obtain a contradiction similarly. This shows the claim that

$$\frac{dr_1(t)}{dt} = 0$$

at all $\mathbf{r}$ contained in an internally chain transitive invariant set.

Values of $r_1(t)$ for which $\frac{dr_1(t)}{dt} = 0$ correspond to the roots of a non-linear polynomial, and therefore there are at most finitely many such values. Calling the set of such values $X^*(1)$, we can conclude that $x(t)$ converges almost surely to $x^* \in X^*(1)$. $\square$

A.3 Proof of Theorem 1

We say a fixed point x^* of $\phi_\sigma(x)$ is a *touchpoint* if there exists $\epsilon > 0$ such that $\phi_\sigma(x) < x$ for all $x \neq x^*$ in $(x^* - \epsilon, x^* + \epsilon)$ or $\phi_\sigma(x) > x$ for all $x \neq x^*$ in $(x^* - \epsilon, x^* + \epsilon)$.

Case (i): x^* is a touchpoint.

The proof extends the arguments from Theorem 1 of Pemantle (1991).[10] Suppose that $\phi_\sigma(x) > x$ for all $x \neq x^*$ in $(x^* - \epsilon, x^* + \epsilon)$. The other case is the same.

Fix $v \in (0, \frac{1}{2})$ and $v_1 \in (v, \frac{1}{2})$. Choose $\gamma > 1$ such that $\gamma v_1 < \frac{1}{2}$. Define $g(r) = re^{(1-r)/(2v_1\gamma)}$. Then $g(1) = 1$ and $g'(1) = 1 - 1/(2v_1\gamma) < 0$, so we can choose $r_0 \in (0, 1)$ with $g(r_0) > 1$. Also define

$$T(n) = e^{n(1-r_0)/(\gamma v_1)}.$$

Then

$$g(r_0)^n = r_0^n T(n)^{1/2} > 1.$$

Choose N such that $\gamma r_0^N < \epsilon$. Since $T(1)^{1/2} r_0 = g(r_0) > 1$, we can find $\alpha > 0$ such that $T(1)^{1/2-\alpha} > r_0$ and therefore $T(n)^{1/2-\alpha} r_0^{-n} \to \infty$. Let $\tau_N = \inf\{j > T(N) : x(j-1) < x^* - r_0^N < x(j)\}$ if there is any such j and $\tau_N = -\infty$ otherwise. For each $n \geq N$, define

$$\tau_{n+1} = \inf\{j \geq \tau_n : x(j) > x^* - r_0^{n+1}\}.$$

So τ_n is the first time the stochastic process crosses $x^* - r_0^n$.

We will show the probability that $\tau_n > T(n)$ for all $n \geq N$ is positive. Since $x(t) \to x^*$ from below whenever $\tau_n > T(n)$ for all $n \geq N$, this will complete the case.

Let $z(t)$ be the fraction of private signals up to period t with realization $s_i = 1$. We first bound the probability that $z(t)$ is far from q. Define a function

$$\phi_{\sigma,z}(x) := \frac{q + \sum_{k=0}^{K} \mathbb{P}[\text{Binom}(K, \lambda x + (1-\lambda)z) = k] \cdot [q \cdot \sigma(1, k) + (1-q) \cdot \sigma(-1, k)]}{1 + C}$$

to be the inflow accuracy when a fraction z of past private signals have value 1.

We begin by defining an event $\mathscr{C}$ under which the number of private signals with positive

[10] Our model does not fit Pemantle (1991)'s definition of a generalized Pólya urn because (1) the relevant stochastic process is two-dimensional and (2) signals are shared in correlated groups of $C + 1$ signals rather than one at a time.

realization is close to q for t sufficiently large. Let $\mathscr{C}_1$ be the event that for all $n \geq N$ and for all $t \geq T(n)$,

$$\phi_{\sigma,z(t)}(x) - x \geq -1/T(n)^{1/2-\alpha}$$

on $(x^* - \epsilon, x^* + \epsilon)$. Because $\phi_{\sigma,z}(x) - x$ is polynomial (in z and x) and is non-negative on this interval when $z = q$, this holds for $|z(t) - q| < B/T(n)^{1/2-\alpha}$ for some $B > 0$.

Suppose event $\mathscr{C}_1$ holds and $\tau_n > T(n)$. Then we have

$$\begin{aligned}
\sum_{t=\tau_n}^{j} h_1(\mathbf{y}(t))/(t+1) &= \sum_{t=\tau_n}^{j} (\phi_{\sigma,z(t)}(x(t)) - x(t))/(t+1) \\
&\geq -\sum_{m=n}^{\infty} \frac{1}{T(m)^{1/2-\alpha}} \sum_{T(m) \leq t < T(m+1)} \frac{1}{t+1} \text{ by the definition of } \mathscr{C}_1 \\
&\geq -\sum_{m=n}^{\infty} \frac{\log(\lceil T(m+1) \rceil) - \log(\lceil T(m) \rceil)}{T(m)^{1/2-\alpha}} \\
&\geq -\sum_{m=n}^{\infty} \left(\frac{1-r_0}{\gamma v_1} + 1 \right) \cdot e^{-m(1/2-\alpha)(1-r_0)/(\gamma v_1)} \\
&= -\left(\frac{1-r_0}{\gamma v_1} + 1 \right) \cdot \frac{e^{-n(1/2-\alpha)(1-r_0)/(\gamma v_1)}}{1 - e^{-(1/2-\alpha)(1-r_0)/(\gamma v_1)}}. \qquad (2)
\end{aligned}$$

We define $\mu = \left(\frac{1-r_0}{\gamma v_1} + 1 \right) \cdot \frac{1}{1-e^{-(1/2-\alpha)(1-r_0)/(\gamma v_1)}}$, so that the right-hand side is $-\mu T(n)^{-(1/2-\alpha)}$.

Let $\mathscr{C}_2$ be the event that for all $n \geq N$ and for all $t \geq T(n)$,

$$\phi_{\sigma,z(t)}(x) - x \leq v\gamma r_0^n \qquad (3)$$

for all $x \in [x^* - \gamma r_0^n, x^*]$. Because $\phi_{\sigma,z}(x) - x$ is polynomial (in z and x) and

$$\frac{d(\phi_{\sigma,q}(x) - x)}{dx}(x^*) = 0,$$

we can choose B' such that for all $n \geq N$ we have

$$\phi_{\sigma,z}(x) - x \leq v\gamma r_0^n$$

for $x \in [x^* - \gamma r_0^n, x^*]$ whenever $|z - q| < B' r_0^n$ (since we can bound the entries of the Hessian of $\phi_{\sigma,z}(x) - x$ above by a constant on the rectangle $[x^* - \gamma r_0^N, x^*] \times [q - r_0^N, q + r_0^N]$). Because

$r_0^n > T(n)^{1/2-\alpha}$, this holds for $|z(t) - q| < B'/T(n)^{1/2-\alpha}$ for some $B' > 0$.

Define the event $\mathscr{C} = \mathscr{C}_1 \cap \mathscr{C}_2$ to be the intersection of these two events. The event $\mathscr{C}$ holds when $|z(t) - q| < \min(B, B')/T(n)^{1/2-\alpha}$ for all $n \geq N$ and all $t \geq T(n)$. By the Chernoff bound and the inequalities $t \geq T(n)$ and $q > 1-q$, the probability of $|z(t)-q| > \min(B, B')/T(n)^{1/2-\alpha}$ is at most $2e^{-\min(B,B')^2 t^{2\alpha}/(2q^2)}$. So the probability that the event $\mathscr{C}$ does not hold for some $n \geq N$ and all $t \geq T(n)$ is at most

$$2 \sum_{n=N}^{\infty} \sum_{t=T(n)}^{\infty} 2e^{-\min(B,B')^2 t^{2\alpha}/(2q^2)}.$$

For N sufficiently large, this sum is approximately

$$\sum_{n=N}^{\infty} \frac{1}{\alpha} \left(\frac{\min(B, B')^2}{2q^2} \right)^{-\frac{1}{2\alpha}} \Gamma \left(\frac{1}{2\alpha}, T(n)^{2\alpha} \min(B, B')^2/(2q^2) \right)$$

where $\Gamma(s, x)$ is the incomplete Gamma function. Since $\Gamma(s,x)/(x^{s-1}e^{-x}) \to 1$ as $x \to \infty$, this sum converges to zero as $N \to \infty$. Increasing N if necessary, we can conclude that the event $\mathscr{C}$ has positive probability. For the remainder of case (i), we condition on this event $\mathscr{C}$.

Now let $\mathscr{B}$ be the event $\{\inf_{j>\tau_n} x(j) \geq x^* - \gamma r_0^n\}$. We will bound the probability of this event conditional on $\tau_n > T(n)$. Let $Z_{m,n} = \sum_{t=m}^{n-1} M(t+1)$ be the sum of the martingale parts of the stochastic process. Because the differences $M(t)$ are martingales with $|M(t)| \leq (C+1)/(t+1)$, we have

$$\mathbb{E}[Z_{m,\infty}^2] = \sum_{t=m}^{\infty} \mathbb{E}[M(t)^2] \leq \sum_{t=m}^{\infty} \left(\frac{C+1}{t+1} \right)^2 \leq \frac{(C+1)^2}{m}. \tag{4}$$

We have:

$$\begin{aligned}
\mathbb{P}\left[\mathscr{B}^c \,|\, \tau_n > T(n)\right] &= \mathbb{P}\left[\inf_{j>\tau_n} x(j) < x^* - \gamma r_0^n \,\middle|\, \tau_n > T(n) \right] \\
&\leq \mathbb{P}\left[\inf_{j>\tau_n} Z_{\tau_n, j} < -(\gamma-1) r_0^n + \mu T(n)^{-(1/2-\alpha)} \,\middle|\, \tau_n > T(n) \right] \text{ by equation (2)} \\
&= \mathbb{P}\left[Z_{\tau_n,\infty} < -(\gamma-1) r_0^n + \mu T(n)^{-(1/2-\alpha)} \,\middle|\, \tau_n > T(n) \right] \\
&\leq \mathbb{E}\left[Z_{\tau_n,\infty}^2 \,\middle|\, \tau_n > T(n) \right] / ((\gamma-1) r_0^n - \mu T(n)^{-(1/2-\alpha)})^2 \text{ by Chebyshev's inequality} \\
&\leq (C+1)^2 e^{-n(1-r_0)/(v_1\gamma)} ((\gamma-1) r_0^n - \mu T(n)^{-(1/2-\alpha)})^{-2}
\end{aligned}$$

by inequality (4) and the definition of $T(n)$.

Recall that $T(n)^{1/2-\alpha} r_0^{-n} \to \infty$, so for n sufficiently large

$$(\gamma - 1)r_0^n - \mu T(n)^{-(1/2-\alpha)} \geq \frac{\gamma - 1}{2} r_0^n.$$

We conclude that

$$\mathbb{P}\left[\mathscr{B}^c \,|\, \tau_n > T(n)\right] \leq (C+1)^2 \left(\frac{\gamma - 1}{2}\right)^{-2} g(r_0)^{-2n}.$$

This bounds the conditional probability of the event $\mathscr{B}$ not holding.

When the event $\mathscr{B}$ does hold and $\tau_n > T(n)$,

$$\begin{aligned}
\sum_{\substack{T(n)<t<T(n+1)\\ x(t)<x^*}} h_1(\mathbf{y}(t))/(t+1) &= \sum_{\substack{T(n)<t<T(n+1)\\ x(t)<x^*}} (\phi_{\sigma,z(t)}(x(t)) - x(t))/(t+1) \\
&\leq \sum_{\substack{T(n)<t<T(n+1)\\ x(t)<x^*}} v\gamma r_0^n/(t+1) \text{ by equation (3)} \\
&\leq (\log\lceil T(n+1)\rceil - \log\lceil T(n)\rceil)(v\gamma r_0^n) \\
&\quad \text{by the partial sums of the harmonic series} \\
&\leq (v\gamma r_0^n)((1-r_0)/(\gamma v_1) + 1/T(n)) \\
&= (v/v_1)(r_0^n - r_0^{n+1}) + v\gamma r_0^n/T(n).
\end{aligned}$$

Now suppose $\mathscr{B}$ holds and $\tau_n > T(n)$ but $\tau_{n+1} \leq T(n+1)$. Then

$$\begin{aligned}
Z_{\tau_n,\tau_{n+1}} &= x(\tau_{n+1}) - x(\tau_n) - \sum_{t=\tau_n}^{\tau_{n+1}-1} h_1(\mathbf{y}(t))/(t+1) \\
&\geq x(\tau_{n+1}) - x(\tau_n) - \sum_{\substack{T(n)<t<T(n+1)\\ x(t)<x^*}} h_1(\mathbf{y}(t))/(t+1) \\
&\geq r_0^n - r_0^{n+1} - \xi_n - (v/v_1)(r_0^n - r_0^{n+1}) - v\gamma r_0^n/T(n) \text{ by the inequality above and definition of } \tau_n \\
&= r_0^n(1-r_0)(1-v/v_1) - \xi_n - v\gamma r_0^n/T(n),
\end{aligned}$$

where ξ_n is an error term since $x(\tau_n)$ may be larger than $x^* - r_0^n$ and $\widetilde{\xi}_n = \xi_n + v\gamma r_0^n/T(n)$. Since the error term ξ_n is at most $1/T(n)$ and therefore is lower order than r_0^n, we have

$$\frac{r_0^n(1-r_0)(1-v/v_1)-\widetilde{\xi}_n}{r_0^n(1-r_0)(1-v/v_1)} \to 1 \tag{5}$$

as $n \to \infty$.

Combining our bounds, we have:

$$\begin{aligned}
\mathbb{P}[\tau_{n+1} \le T(n+1) \,|\, \tau_n > T(n)] \le{}& \mathbb{P}\left[\mathscr{B}^c \,|\, \tau_n > T(n)\right] + \\
& \mathbb{P}\left[\mathscr{B}, \sup_{\tau_{n+1}} Z_{\tau_n,\tau_{n+1}} \ge r_0^n(1-r_0)(1-v/v_1)-\widetilde{\xi}_n \,\middle|\, \tau_n > T(n)\right] \\
& \le (C+1)^2\left(\frac{\gamma-1}{2}\right)^{-2} g(r_0)^{-2n} + \frac{\mathbb{E}[Z^2_{\tau_n,\infty} \,|\, \tau_n > T(n)]}{(r_0^n(1-r_0)(1-v/v_1)-\widetilde{\xi}_n)^2} \\
& \text{by Chebyshev's inequality} \\
& \le (C+1)^2\left(\frac{\gamma-1}{2}\right)^{-2} g(r_0)^{-2n} + \frac{(C+1)^2 T(n)^{-1}}{(r_0^n(1-r_0)(1-v/v_1)-\widetilde{\xi}_n)^2} \\
& \text{by inequality (4)} \\
& \le (C+1)^2\left(\frac{\gamma-1}{2}\right)^{-2} g(r_0)^{-2n} + \\
& (C+1)^2((1-r_0)(1-v/v_1))^{-2} g(r_0)^{-2n} \cdot \frac{r_0^n(1-r_0)(1-v/v_1)-\widetilde{\xi}_n}{r_0^n(1-r_0)(1-v/v_1)}.
\end{aligned}$$

We claim that the sum of these probabilities converges. The sum of the first terms converges because $g(r_0) > 1$. For the second term, recall that the fraction $\frac{r_0^n(1-r_0)(1-v/v_1)-\widetilde{\xi}_n}{r_0^n(1-r_0)(1-v/v_1)}$ converges to 1. So the sum of the second terms also converges because $g(r_0) > 1$.

We have

$$\mathbb{P}[\tau_n > T(n) \text{ for all } n \ge N] = \mathbb{P}[\tau_N > T(N)] \prod_{n=M}^{\infty} (1 - \mathbb{P}[\tau_{n+1} \le T(n+1) \,|\, \tau_n > T(n)]).$$

On the right-hand side, each factor in the product is positive and

$$\sum_{n=M}^{\infty} \mathbb{P}[\tau_{n+1} \le T(n+1) \,|\, \tau_n > T(n)]$$

is finite. By a standard result on infinite products, this implies the product is positive. So the probability that $\tau_n > T(n)$ for all $n \geq N$ is positive, which implies that the probability $\pi(x^*|\sigma)$ of converging to x^* is positive.

Case (ii): There exists $\epsilon > 0$ such that $\phi_\sigma(x) > x$ for all $x \in (x^* - \epsilon, x^*)$ and $\phi_\sigma(x) < x$ for all $x \in (x^*, x^* + \epsilon)$.

Our argument is based on the related result for generalized Pólya urns from Hill, Lane, and Sudderth (1980). We begin with a lemma, which says that suitably changing a stochastic process away from a neighborhood of a fixed point does not affect whether we converge to that fixed point with positive probability:

Lemma 2. *Suppose*

$$\widetilde{\mathbf{y}}(t+1) = \widetilde{\mathbf{y}}(t) + \frac{1}{t+1}\left(\widetilde{\xi}(t+1) - \widetilde{\mathbf{y}}(t)\right),$$

where the conditionally i.i.d. random variables $\widetilde{\xi}(t+1)$ have the same conditional distribution as $\xi(t+1)$ in a neighborhood U of (x^*, q), have the same support as $\xi(t+1)$ for all $(x, z) \in (0,1)^2$, and have expectations $\mathbb{E}[\widetilde{\xi}(t+1)]$ that are Lipschitz continuous in (x, z). Then $x(t)$ converges to x^* with positive probability if and only if $\widetilde{x}(t) = \widetilde{\mathbf{y}}_1(t)$ does.

Proof. The stochastic process $x(t)$ converges to x^* with positive probability if and only if there exists some T and some $(x(T), z(T))$ reached with positive probability under $\mathbf{y}(t)$ such that starting with initial condition $(x(T), z(T))$, with positive probability $x(t) \to x^*$ and $(x(t), z(t)) \in U$ for $t \geq T$.

Because the random variables $\widetilde{\xi}(t)$ have the same support as $\xi(t)$ whenever x and z are interior, the point $(x(T), z(T))$ is reached with positive probability under $\widetilde{\mathbf{y}}(t)$ if and only if it is reached with positive probability under $\mathbf{y}(t)$. Because $\widetilde{\xi}(t)$ and $\xi(t)$ agree on U, starting with initial condition $(x(T), z(T))$, with positive probability $\widetilde{x}(t) \to x^*$ and $(\widetilde{x}(t), \widetilde{z}(t)) \in U$ for $t \geq T$ if and only if the same holds for $(x(t), z(t))$. These conditions hold for some $(x(T), z(T))$ if and only if $\widetilde{x}(t)$ converges to x^* with positive probability. $\square$

Now choose $\widetilde{\xi}(t)$ satisfying the conditions of the lemma, agreeing with $\xi(t)$ in the second coordinate, and such that the unique fixed point of the corresponding function $\widetilde{\phi}_\sigma(x)$ is x^*. To do so, choose an open neighborhood U of (x^*, q) such that x^* is the unique fixed point of $\phi_\sigma(x)$ with $(x, q) \in \overline{U}$. Let $\widetilde{\xi}(t) = \xi(t)$ on the closure $\overline{U}$ of U. For each z, let $\widetilde{\xi}(t)$ be constant in x outside of the neighborhood U.

Then $\widetilde{\xi}(t)$ and $\xi(t)$ have the same support because both have full support for all interior x and z. Lipschitz continuity follows from Lipschitz continuity of the expectations of $\xi(t)$ in x and z, which we checked in the proof of Proposition 2.

Since x^* is the unique fixed point of $\phi_\sigma(x)$, by the same argument as in Proposition 2, we have $\widetilde{x}(t) \to x^*$ almost surely. Note that this step uses Lipschitz continuity of $\mathbb{E}[\widetilde{\xi}(t+1)]$. So by Lemma 2, $x(t) \to x^*$ with positive probability.

A.4 Proof of Lemma 1

Proof. If $x > 1/2$, then sampling accuracy is $\lambda x + (1-\lambda)q > 1/2$ since $q > 1/2$ also. If $x < 1/2$ and $\phi_{\sigma^{\text{maj}}}(x) = x$, then the sampling accuracy must be strictly less than $1/2$. Otherwise, if sampling accuracy is $1/2$ or higher, then the majority rule implies $\sum_{k=0}^{K} P_k(x,\lambda) \cdot [q \cdot \sigma^{\text{maj}}(1,k) + (1-q) \cdot \sigma^{\text{maj}}(-1,k)] \geq C/2$ and so $\phi_{\sigma^{\text{maj}}}(x) > 1/2$. Finally, if $x = 1/2$ were a steady state, then its sampling accuracy would be at least $1/2$, so again by the same reason we would have $\phi_{\sigma^{\text{maj}}}(x) > 1/2$, a contradiction. □

A.5 Proof of Theorem 2

A.5.1 Preliminary Lemmas

We first state and prove three preliminary lemmas.

Lemma 3. *Suppose σ is state symmetric, $\mathbb{E}[\sigma(1,k)] \geq \mathbb{E}[\sigma(-1,k)]$ for every $0 \leq k \leq K$, and that $\sigma(1, K/2)(C) = 1$, $\sigma(-1, K/2)(0) = 1$ if K is even. If sampling accuracy at x is weakly smaller than 1/2, then $\phi_{\sigma^{maj}}(x) \leq \phi_\sigma(x)$. If sampling accuracy at x is strictly smaller than $1/2$ and $\sigma \neq \sigma^{maj}$, then $\phi_{\sigma^{maj}}(x) < \phi_\sigma(x)$.*

Proof. We have

$$\phi_{\sigma^{\text{maj}}}(x) = \frac{q + P_{K/2}(x,\lambda) \cdot q \cdot C + \sum_{j>K/2} P_j(x,\lambda) \cdot C}{1+C}$$

We first note that in a news feed with $K/2$ out of K positive stories, by assumption on σ it must share C positive stories that match the private signal, so both strategies contribute $q \cdot C$ correct shares in expectation.

For each $j > K/2$, by symmetry we have $\mathbb{E}[\sigma(1,j)] = C - \mathbb{E}[\sigma(-1,K-j)]$ and $\mathbb{E}[\sigma(-1,j)] = C - \mathbb{E}[\sigma(1,K-j)]$.

We have

$$
\begin{aligned}
&P_j(x,\lambda)\cdot(q\cdot\mathbb{E}[\sigma(1,j)] + (1-q)\cdot\mathbb{E}[\sigma(0,j)]) + P_{K-j}(x,\lambda)\cdot(q\cdot\mathbb{E}[\sigma(1,K-j)] + (1-q)\cdot\mathbb{E}[\sigma(-1,K-j)]) \\
=&P_j(x,\lambda)\cdot(q\cdot(C-\mathbb{E}[\sigma(-1,K-j)]) + (1-q)\cdot(C-\mathbb{E}[\sigma(1,K-j)])) \\
&+ P_{K-j}(x,\lambda)\cdot(q\cdot\mathbb{E}[\sigma(1,K-j)] + (1-q)\cdot\mathbb{E}[\sigma(-1,K-j)]) \\
=&P_j(x,\lambda)\cdot C - P_j(x,\lambda)\cdot(q\mathbb{E}[\sigma(-1,K-j)] + (1-q)\mathbb{E}[\sigma(1,K-j)]) \\
&+ P_{K-j}(x,\lambda)\cdot(q\cdot\mathbb{E}[\sigma(1,K-j)] + (1-q)\cdot\mathbb{E}[\sigma(-1,K-j)]) \\
\geq&P_j(x,\lambda)\cdot C - P_j(x,\lambda)\cdot(q\mathbb{E}[\sigma(-1,K-j)] + (1-q)\mathbb{E}[\sigma(1,K-j)]) \\
&+ P_{K-j}(x,\lambda)\cdot((1-q)\cdot\mathbb{E}[\sigma(1,K-j)] + q\cdot\mathbb{E}[\sigma(-1,K-j)])
\end{aligned}
$$

using the fact that $q > 1/2$ and $\mathbb{E}[\sigma(1,K-j)] \geq \mathbb{E}[\sigma(-1,K-j)]$ by the hypothesis on σ.

Suppose x is weakly misleading, so $\lambda x + (1-\lambda)q \leq 1/2$. Then $P_{K-j}(x,\lambda) \geq P_j(x,\lambda)$ since $j > K/2$. This shows when $\omega = 1$, the expected number of positive stories shared by σ with a j majority in the news feed is weakly larger than that shared by σ^{maj}. So $\phi_{\sigma^{\text{maj}}}(x) \leq \phi_\sigma(x)$.

Now suppose x strictly misleading, so $\lambda x + (1-\lambda)q < 1/2$. Then $P_j(x,\lambda) < P_{K-j}(x,\lambda)$ and the final term is strictly larger than $P_j(x,\lambda)\cdot C$ except when $\mathbb{E}[\sigma(-1,K-j)] = \mathbb{E}[\sigma(1,K-j)] = 0$. This shows we get $\phi_{\sigma^{\text{maj}}}(x) < \phi_\sigma(x)$ except when $\sigma(-1,k) = \sigma(1,k)$ is the degenerate distribution on 0 for any $k < K/2$, which by symmetry only happens when $\sigma = \sigma^{\text{maj}}$. □

Lemma 4. *Suppose σ is state symmetric and $\sigma(1,k)(C) = 1$ for every $k \geq K/2$. Then, ϕ_σ does not have any fixed point x with $\lambda x + (1-\lambda)q > 1/2$ and $x \leq q$.*

Proof. Suppose by way of contradiction that such a fixed point x exists. Let $y = \lambda x + (1-\lambda)q$ be the sampling accuracy, and note $x \leq y \leq q$, with $y > 1/2$. Expected number of positive stories shared by each new agent when $\omega = 1$ is:

$$q\cdot\sum_{k=0}^{K} P_k(x,\lambda)\cdot\mathbb{E}[\sigma(1,k)] + (1-q)\cdot\sum_{k=0}^{K} P_k(x,\lambda)\cdot\mathbb{E}[\sigma(-1,k)].$$

We know $\mathbb{E}[\sigma(1,k)] = C \geq \mathbb{E}[\sigma(-1,k)]$ for each $k \geq K/2$, and conversely $\mathbb{E}[\sigma(-1,k)] = 0 \leq$

$\mathbb{E}[\sigma(1,k)]$ for each $k < K/2$. This means $\sum_{k=0}^{K} P_k(x,\lambda) \cdot \mathbb{E}[\sigma(1,k)] \geq \sum_{k=0}^{K} P_k(x,\lambda) \cdot \mathbb{E}[\sigma(-1,k)]$. Since $q \geq y > 1/2$, this expected number of shared positive stories is at least

$$y \cdot \sum_{k=0}^{K} P_k(x,\lambda) \cdot \mathbb{E}[\sigma(1,k)] + (1-y) \cdot \sum_{k=0}^{K} P_k(x,\lambda) \cdot \mathbb{E}[\sigma(-1,k)].$$

For each $0 \leq k < K/2$,

$$\begin{aligned}(1-y) \cdot P_{K-k}(x,\lambda) &= (1-y) \cdot \binom{K}{k} y^{K-k}(1-y)^k \\ &\geq y \cdot \binom{K}{k} y^k (1-y)^{K-k},\end{aligned}$$

so

$$(1-y) \cdot P_{K-k}(x,\lambda) \cdot \mathbb{E}[\sigma(-1,K-k)] \geq y \cdot P_k(x,\lambda) \cdot \mathbb{E}[\sigma(0,K-k)]$$

and the inequality is strict for $k=0$ because $y > 1/2$. So we have

$$\begin{aligned}
&y \cdot \sum_{k=0}^{K} P_k(x,\lambda) \cdot \mathbb{E}[\sigma(1,k)] + (1-y) \cdot \sum_{k=0}^{K} P_k(x,\lambda) \cdot \mathbb{E}[\sigma(-1,k)] \\
\geq &y \cdot \sum_{k \geq K/2}^{K} P_k(x,\lambda) \cdot C + y \cdot \sum_{k<K/2}^{K} P_k(x,\lambda) \cdot \mathbb{E}[\sigma(1,k)] + (1-y) \cdot \sum_{k>K/2}^{K} P_k(x,\lambda) \cdot \mathbb{E}[\sigma(-1,k)] \\
> &y \cdot \sum_{k \geq K/2}^{K} P_k(x,\lambda) \cdot C + y \cdot \sum_{k<K/2}^{K} P_k(x,\lambda) \cdot \mathbb{E}[\sigma(1,k)] + y \cdot \sum_{k<K/2}^{K} P_k(x,\lambda) \cdot \mathbb{E}[\sigma(-1,K-k)] \\
= &y \cdot \sum_{k \geq K/2}^{K} P_k(x,\lambda) \cdot C + y \cdot \sum_{k<K/2}^{K} P_k(x,\lambda) \cdot C
\end{aligned}$$

where the last step uses the symmetry condition $\mathbb{E}[\sigma(1,k) + \sigma(-1,K-k)] = C$ for $k < K/2$. The last expression is yC, thus $\phi_\sigma(x) > \frac{q+yC}{1+C} \geq \frac{y+yC}{1+C} = y \geq x$. This contradicts x being a fixed point. Hence there are no fixed points of ϕ_σ that satisfy both conditions in the statement of the lemma. $\square$

Lemma 5. *For each $\epsilon', \epsilon'' > 0$, $p \in (0,1)$, strategy σ^* and $0 \leq \overline{\lambda} \leq 1$ with $\phi_{\sigma^*}^{\overline{\lambda}}(x) - x \geq 2\epsilon'$ for every x with $\overline{\lambda}x + (1-\overline{\lambda})q \leq p + 2\epsilon'$ (where $\phi_{\sigma^*}^{\overline{\lambda}}$ is the inflow accuracy function with virality weight $\overline{\lambda}$), there is some N and some $\delta > 0$ so that for every σ with $\|\sigma - \sigma^*\|_2 < \delta$ and λ with $|\lambda - \overline{\lambda}| < \delta$, we have $\mathbb{P}_{\sigma,\lambda}[\lambda x(t) + (1-\lambda)q \geq p + \epsilon'/2] > 1 - \epsilon''$ for every $t \geq N$.*

Proof. Because $\phi_\sigma^\lambda(x)$ is polynomial in λ, σ, and x, there exists $\delta > 0$ such that $\phi_\sigma^\lambda(x) - x \geq \epsilon'$ for every x with $\lambda x + (1-\lambda)q \leq p + \epsilon'$ when $\|\sigma^* - \sigma\|_2 < \delta$ and $|\overline{\lambda} - \lambda| < \delta$. Shrinking δ if necessary, we can also assume that λ is bounded away from zero when $|\overline{\lambda} - \lambda| < \delta$.

For the remainder of the proof, fix σ and λ in these neighborhoods. We will observe at the end of the proof that the bounds we will prove are uniform in the choice of σ and λ.

Let $p' > p + \epsilon'$ be the largest number in $(0,1)$ such that

$$\phi_\sigma(x) - x \geq \epsilon'/2 \tag{6}$$

for all x satisfying $\lambda x + (1-\lambda)q \leq p'$. Let $N_1 < N_2$ be positive integers with $N_2 \geq bN_1$ for some integer $b > 1$. We will first show that for N_1 and N_2 large enough, the probability that $\lambda x(t) + (1-\lambda)q < p'$ for all $t \in [N_1, N_2]$ is small. We will then show that if $\lambda x(t_1) + (1-\lambda)q > p'$ for some $N_1 \leq t_1 < N_2$, then the probability that $\lambda x(N_2) + (1-\lambda)q < p + \epsilon'/2$ is small.

By the Chernoff bound applied to $z(t)$ and compactness of the set of strategies σ under consideration, we can choose a constant $B > 0$ independent of σ such that

$$\max_{x\in[0,1]} |\phi_{\sigma,z(t)}(x) - \phi_\sigma(x)| < \epsilon'/4 \tag{7}$$

with probability at least $1 - 2e^{-Bt}$ for t sufficiently large.

Recall that we can decompose $\mathbf{y}(t)$ as

$$\mathbf{y}(t+1) = \mathbf{y}(t) + h(\mathbf{y}(t)) + M(t+1),$$

where $h(\mathbf{y}(t))$ is deterministic and $M(t+1)$ is a martingale. We have $|M(t)| < 2(C+1)/t$ for all t. So by Theorem 14 from Appendix C of Borkar (2009), for any $\alpha > 0$ and any t_1 and t_2,

$$\mathbb{P}\left(\sup_{t_1<t<t_2}\left|\sum_{i=t_1}^{t} M(t)\right| > \alpha\right) \leq 2e^{-\frac{\alpha^2}{\sum_{i=t_1}^{t_2} 4(C+1)^2/i^2}}. \tag{8}$$

Consider the event E that $\lambda x(t) + (1-\lambda)q < p'$ for all $N_1 \leq t \leq N_2$. Suppose inequality (7) holds for all $N_1 \le t < N_2$. Then we have

$$\begin{aligned}
x(N_2) - x(N_1) &= \sum_{t=N_1}^{N_2-1} \frac{\phi_{\sigma,z(t)}(x(t)) - x(t)}{t+1} + \sum_{t=N_1}^{N_2-1} M(t+1) \\
&= \sum_{t=N_1}^{N_2-1} \frac{\phi_{\sigma,z(t)}(x(t)) - \phi_\sigma(x(t))}{t+1} + \sum_{t=N_1}^{N_2-1} \frac{\phi_\sigma(x(t)) - x(t)}{t+1} + \sum_{t=N_1}^{N_2-1} M(t+1) \\
&\ge \sum_{t=N_1}^{N_2-1} \epsilon'/4 \cdot \frac{1}{t+1} + \sum_{t=N_1}^{N_2-1} M(t+1) \text{ by inequalities (6) and (7)} \\
&\ge (\epsilon'/4)(\log(N_2) - \log(N_1)) + \sum_{t=N_1}^{N_2-1} M(t+1).
\end{aligned}$$

When event E holds, the right-hand side must be at most p'/λ. Taking b and therefore N_2/N_1 sufficiently large, we can assume that

$$(\epsilon'/4)(\log(N_2) - \log(N_1)) > 2p'/\lambda.$$

By equation (8), the absolute value of the sum of martingales is greater than p'/λ with probability at most

$$2e^{-\frac{(p'/\lambda)^2}{\sum_{i=N_1}^{N_2} 4(C+1)^2/(i+1)^2}} \le 2e^{-\frac{(p'/\lambda)^2 N_1 N_2}{8(C+1)^2(N_2-N_1)}} < 2e^{-\frac{(p'/\lambda)^2 N_1}{8(C+1)^2}}.$$

Along with the Chernoff bound, this gives an upper bound on the probability of event E.

If event E does not hold, there exists some $N_1 \le t \le N_2$ such that $\lambda x(t) + (1-\lambda)q \ge p'$. Choose t_1 so that $t_1 - 1$ is the largest such t.

Suppose $\lambda x(N_2) + (1-\lambda)q \le p + \epsilon'/2$. For N_1 sufficiently large, this implies $t_1 \le N_2$. So we must have

$$x(N_2) - x(t_1) \le ((p + \epsilon'/2) - p')/\lambda < -\epsilon'/(2\lambda).$$

On the other hand, when inequality (7) holds for all $N_1 \leq t < N_2$,

$$\begin{aligned}
x(N_2) - x(t_1) &= \sum_{t=t_1}^{N_2-1} \frac{\phi_{\sigma,z(t)}(x(t)) - x(t)}{t+1} + \sum_{t=t_1}^{N_2-1} M(t+1) \\
&= \sum_{t=t_1}^{N_2-1} \frac{\phi_{\sigma,z(t)}(x(t)) - \phi_\sigma(x(t))}{t+1} + \sum_{t=t_1}^{N_2-1} \frac{\phi_\sigma(x(t)) - x(t)}{t+1} + \sum_{t=t_1}^{N_2-1} M(t+1) \\
&\geq \sum_{t=t_1}^{N_2-1} \epsilon'/4 \cdot \frac{1}{t+1} + \sum_{t=t_1}^{N_2-1} M(t+1) \text{ by inequalities (6) and (7)} \\
&\geq (\epsilon'/4)(\log(N_2) - \log(t_1)) + \sum_{t=t_1}^{N_2-1} M(t+1).
\end{aligned}$$

Applying equation (8) with $\alpha = \epsilon'/(2\lambda)$, the absolute value of the sum of martingales is greater than $\epsilon'/(2\lambda)$ with probability at most

$$2e^{-\frac{(\epsilon')^2/(2\lambda)^2}{\sum_{i=t_1}^{N_2} 4(C+1)^2/(i+1)^2}} \leq 2e^{-\frac{(\epsilon'/\lambda)^2 N_2 t_1}{32(C+1)^2(N_2-t_1)}} \leq 2e^{-\frac{(\epsilon'/\lambda)^2 t_1}{32(C+1)^2}}.$$

When this does not hold and the Chernoff bounds apply, $x(N_2) - x(t_1)$ is greater than $-\epsilon'/(2\lambda)$ and therefore $\lambda x(N_2) + (1-\lambda)q > p + \epsilon'/2$ if N_1 is sufficiently large. This gives an upper bound on the probability that $\lambda x(N_2) + (1-\lambda)q \leq p + \epsilon'/2$.

We conclude that

$$\mathbb{P}_\sigma[\lambda x(N_2) + (1-\lambda)q < p + \epsilon'/2] \leq 2e^{-\frac{(p'/\lambda)^2 N_1}{8(C+1)^2}} + \sum_{t=N_1+1}^{N_2-1} 2e^{-\frac{(\epsilon'/\lambda)^2 t}{32(C+1)^2}} + 2\sum_{t=N_1}^{N_2-1} e^{-Bt}$$

for N_1 sufficiently large. Because the second and third terms are geometric series, we can choose N_1 sufficiently large so that this probability is less than ϵ'' for all $N_2 \geq bN_1$. Because λ is bounded away from zero, we can make this choice uniformly in λ and σ (subject to the constraints $|\lambda - \overline{\lambda}| < \delta$ and $\|\sigma - \sigma^*\|_2 < \delta$). So for N_1 sufficiently large, we have

$$\mathbb{P}_\sigma[\lambda x(t) + (1-\lambda)q \geq p + \epsilon'/2] > 1 - \epsilon''$$

for $t \geq N = bN_1$. $\square$

A.5.2 Proof of Theorem 2

Proof. **Part 1**: Fix $0 < \lambda \leq \lambda^*$ and suppose σ^* is a social equilibrium.

Step 1: Either $\sigma^* = \sigma^{\text{maj}}$, or all fixed points of ϕ_{σ^*} are strictly informative.

We verify that σ^* satisfies the hypotheses of Lemma 3. Note σ^* is the limit of a sequence of symmetric BNEs $(\sigma^{(i)})$, where every $\sigma^{(i)}$ is state symmetric. Also, in the i-th finite society under the equilibrium $\sigma^{(i)}$, belief about $\{\omega = 1\}$ must be weakly higher after observing k positive stories and $s = 1$ than k positive stories and $s = -1$ for every $0 \leq k \leq K$. So by optimality of $\sigma^{(i)}$, we have $\mathbb{E}[\sigma^{(i)}(1, k)] \geq \mathbb{E}[\sigma^{(i)}(-1, k)]$ for every i and every $0 \leq k \leq K$. The limit σ^* must also satisfy state symmetry and $\mathbb{E}[\sigma^*(1, k)] \geq \mathbb{E}[\sigma^*(-1, k)]$ for every $0 \leq k \leq K$. Also, when K is even, by the state symmetry of the equilibrium $\sigma^{(i)}$ we know that a news feed with $K/2$ positive stories in society i generates an equilibrium posterior belief that both states are equally likely. Thus, optimality of $\sigma^{(i)}$ implies $\sigma^{(i)}(1, K/2)(C) = 1$ and $\sigma^{(i)}(-1, K/2)(0) = 1$. The limit σ^* must then also satisfy $\sigma^*(1, K/2)(C) = 1$, $\sigma^*(-1, K/2)(0) = 1$.

If ϕ_{σ^*} has a strictly misleading fixed point and $\sigma^* \neq \sigma^{\text{maj}}$, that is some $x \in [0, 1]$ with $\lambda x + (1 - \lambda)q < 1/2$ and such that $\phi_{\sigma^*}(x) = x$, then by Lemma 3 we get $\phi_{\sigma^{\text{maj}}}(x) < x$. But we also have $\phi_{\sigma^{\text{maj}}}(0) > 0$, which means $\phi_{\sigma^{\text{maj}}}$ has a strictly misleading fixed point in $(0, x)$ by the intermediate-value theorem, and further $\phi_{\sigma^{\text{maj}}}$ will continue to have a nearby fixed point for nearby values of λ. Since $x \leq 1/2$, this implies for some $\lambda' < \lambda^*$, $\phi_{\sigma^{\text{maj}}}$ has a fixed point in $[0, 1/2]$, which contradicts the definition of λ^*.

If there is some x with $\lambda x + (1 - \lambda)q = 1/2$ and such that $\phi_{\sigma^*}(x) = x$, then by Lemma 3 we get $\phi_{\sigma^{\text{maj}}}(x) \leq x$. But since the sampling accuracy at x is exactly $1/2$, every sample is as likely as its mirror image, so the majority rule is expected to share at least $C/2$ correct stories out of C, hence $\phi_{\sigma^{\text{maj}}}(x) > 1/2$ after accounting for the arrival of new stories that tend to match the true state. This is a contradiction. Thus, every fixed point of ϕ_{σ^*} must be strictly informative unless $\sigma^* = \sigma^{\text{maj}}$.

Step 2: If $\lambda < \lambda^*$, ϕ_{σ^*} only has fixed points in $(q, 1]$. If $\lambda = \lambda^*$, either $\sigma^* = \sigma^{\text{maj}}$ or ϕ_{σ^*} only has fixed points in $(q, 1]$.

We first show all fixed points of ϕ_{σ^*} are strictly informative, except when $\lambda = \lambda^*$ and $\sigma^* = \sigma^{\text{maj}}$. If $\lambda < \lambda^*$ and $\sigma^* = \sigma^{\text{maj}}$, by definition of λ^* we know that all fixed points of ϕ_{σ^*} are strictly informative. And if $\sigma^* \neq \sigma^{\text{maj}}$, then by **Step 1**, all fixed points of ϕ_{σ^*} are strictly informative. If $\lambda = \lambda^*$ and $\sigma^* \neq \sigma^{\text{maj}}$, again **Step 1** implies all fixed points of ϕ_{σ^*} are strictly informative.

We verify that, except when $\lambda = \lambda^*$ and $\sigma^* = \sigma^{\text{maj}}$, σ^* is such that $\sigma^*(1,k)(C) = 1$ for every $k \geq K/2$, and thus satisfies the hypotheses of Lemma 4. Since all fixed points of ϕ_{σ^*} are strictly informative, there exists some $\epsilon' > 0$ so that $\phi_{\sigma^*}(x) - x > 2\epsilon'$ for every x where $\lambda x + (1-\lambda)q \leq 0.5 + 2\epsilon'$. Find some $\epsilon'' > 0$ so that

$$\frac{(0.5+\epsilon'/4)^{\lfloor K/2 \rfloor +1} \cdot (0.5-\epsilon'/4)^{K-\lfloor K/2 \rfloor -1} \cdot (1-3\epsilon'')}{(0.5+\epsilon'/4)^{K-\lfloor K/2 \rfloor -1} \cdot (0.5-\epsilon'/4)^{\lfloor K/2 \rfloor +1} \cdot (1-3\epsilon'') + 3\epsilon''} > 1. \tag{9}$$

Apply Lemma 5 to these ϵ', ϵ'' and $p = 1/2$ to find N and δ. Also, by the law of large numbers, we may find N' so that $\mathbb{P}[|\text{Binom}(t,q)/t - q| > \epsilon'/4] < \epsilon''$ whenever $t \geq N'$, where $\text{Binom}(t,q)$ refers to a binomial random variable with t trials and q success probability. Since $\sigma^{(i)} \to \sigma^*$, there exists I so that $\|\sigma^{(i)} - \sigma^*\|_2 < \delta$ whenever $i \geq I$. For all $i \geq I$ large enough, we must have $\max(N,N')/(n_i - K) < \epsilon''$. But in the equilibrium $\sigma^{(i)}$ in society i, we know that agents' equilibrium belief about sampling accuracy for any position in $\{\max(N,N'), \max(N,N')+1, ..., n_i\}$ assigns mass at least $1-2\epsilon''$ to the region $[0.5+\epsilon'/4, 1]$. This is because if we have both $\lambda x(t) + (1-\lambda)q \geq 0.5 + \epsilon'/2$ and $|z(t) - q| < \epsilon'/4$, then we also have $\lambda x(t) + (1-\lambda)z(t) \geq 0.5 + \epsilon'/4$ — the former event has probability at least $1-\epsilon''$ when $t \geq N$ and latter event has probability at least $1-\epsilon''$ when $t \geq N'$. Hence, equilibrium belief about sampling accuracy for a uniformly random position in $\{K+1, K+2, ..., \max(N,N'), \max(N,N')+1, ..., n_i\}$ assigns mass at least $1-3\epsilon''$ to the same region. Thus, the expression on the LHS of Equation (9) gives a lower bound on the the posterior likelihood ratio of $\omega = 1$ and $\omega = -1$ after seeing a sample containing k positive stories in equilibrium $\sigma^{(i)}$, for any $k > K/2$. Hence, by optimality, $\sigma^{(i)}(1,k)(C) = 1$ for every $k > K/2$. Also, for any belief about sampling accuracy, a sample with $k = K/2$ is uninformative, so if $K/2$ is an integer then $\sigma^{(i)}(1,K/2)(C) = 1$ by optimality. Thus we see for all large enough j, $\sigma^{(i)}(1,k)(C) = 1$ for every $k \geq K/2$, hence the same must hold for the limit σ^*.

Combining the conclusion of Lemma 4 (which rules out steady states at or lower than q with a sampling accuracy strictly higher than $1/2$) with the argument at the beginning of **Step 2** (which rules out steady states with sampling accuracy $1/2$ or lower), we have completed this step.

Step 3: $\sigma^* = \sigma^{\text{maj}}$.

By **Step 2**, we just need to establish this when ϕ_{σ^*} only has fixed points in $(q, 1]$. By state symmetry it suffices to show that $\sigma^*(-1, k)(C) = 1$ for every $k > K/2$. Since ϕ_{σ^*} only has fixed points in $(q, 1]$, there exists some $\epsilon' > 0$ so that $\phi_{\sigma^*}(x) - x \geq 2\epsilon'$ for every x where $\lambda x + (1-\lambda)q \leq q + 2\epsilon'$. Find some $\epsilon'' > 0$ so that

$$\frac{(q+\epsilon'/4)^{\lfloor K/2 \rfloor + 1} \cdot (1-q-\epsilon'/4)^{K - \lfloor K/2 \rfloor - 1} \cdot (1-3\epsilon'')}{(q+\epsilon'/4)^{K-\lfloor K/2 \rfloor - 1} \cdot (1-q-\epsilon'/4)^{\lfloor K/2 \rfloor + 1} \cdot (1-3\epsilon'') + 3\epsilon''} > \frac{q}{1-q}. \tag{10}$$

Apply Lemma 5 to these ϵ', ϵ'' and $p = q$ to find N and δ. Also, by the law of large numbers, we may find N' so that $\mathbb{P}[|\text{Binom}(t, q)/t - q| > \epsilon'/4] < \epsilon''$ whenever $t \geq N'$, where $\text{Binom}(t, q)$ refers to a binomial random variable with t trials and q success probability. Since $\sigma^{(i)} \to \sigma^*$, there exists I so that $\|\sigma^{(i)} - \sigma^*\|_2 < \delta$ whenever $i \geq I$. For all $i \geq I$ large enough, by the same arguments as in **Step 2**, the expression on the LHS of Equation (10) gives a lower bound on the the posterior likelihood ratio of $\omega = 1$ and $\omega = -1$ after seeing a sample containing k positive stories in equilibrium $\sigma^{(i)}$ in society i, for any $k > K/2$. So even if the private signal is $s = -1$, the $\omega = 1$ state is still strictly more likely, and $\sigma^{(i)}(-1, k)(C) = 1$ by optimality. So we also have in the limit $\sigma^*(-1, k)(C) = 1$ for every $k > K/2$.

Part 2: For $\lambda < \lambda^*$, σ^{maj} has a unique steady state.

For K odd, we have

$$\phi_{\sigma^{\text{maj}}}(x) = \frac{q + C\sum_{k > K/2} P_k(x, \lambda)}{1 + C}.$$

By straightforward algebra, we can show that

$$\phi'_{\sigma^{\text{maj}}}(x) = \frac{C\lambda K}{1+C}\mathbb{P}\left[\text{Binom}\left(K-1, \lambda x + (1-\lambda)q\right) = \frac{K-1}{2}\right].$$

If $\lambda = 0$, then $\phi'_{\sigma^{\text{maj}}}(x)$ is constant in x, so $\phi_{\sigma^{\text{maj}}}(x)$ cannot intersect $y = x$ multiple times. If $\lambda > 0$, then $\phi'_{\sigma^{\text{maj}}}(x)$ is strictly decreasing in $x \in (\frac{1}{2}, 1]$, so $\phi_{\sigma^{\text{maj}}}(x)$ is strictly concave for $x \in (\frac{1}{2}, 1]$. Since $\phi_{\sigma^{\text{maj}}}(0) > 0$, at the first fixed point x^* with $\phi_{\sigma^{\text{maj}}}(x^*) = x^*$, we must have $\phi'_{\sigma^{\text{maj}}}(x^*) \leq 1$. We have $x^* > 1/2$ by **Part 1**, so strict concavity of $\phi_{\sigma^{\text{maj}}}(x)$ in $(\frac{1}{2}, 1]$ ensures that there are no fixed points larger than x^*. There is a unique informative steady state.

For K even, we have

$$\phi_{\sigma^{\text{maj}}}(x) = \frac{q + qC\sum_{k\geq K/2} P(k, \lambda x + (1-\lambda)q) + (1-q)C\sum_{k\geq (K/2)+1} P(k, \lambda x + (1-\lambda)q)}{1+C}.$$

So

$$\phi'_{\sigma^{\text{maj}}}(x) = \frac{C\lambda K}{1+C} \cdot \left(q\mathbb{P}\left[\text{Binom}\left(K-1, y\right) = \frac{K}{2} - 1\right] + (1-q)\mathbb{P}\left[\text{Binom}\left(K-1, y\right) = \frac{K}{2}\right]\right),$$

where $y = \lambda x + (1-\lambda)q$. The term in parentheses is proportional to

$$g(y) = (q(1-y) + (1-q)y)(y(1-y))^{K/2-1}.$$

For $\lambda > 0$, the derivative $\frac{\partial g}{\partial x}$ has the same sign as the derivative

$$\frac{\partial g}{\partial y} = (y(1-y))^{K/2-2}((K/2-1)(2y-1)(q(2y-1)-y) - (2q-1)(1-y)y),$$

which is strictly negative for $y \in (\frac{1}{2}, 1)$.

Since $y \in (\frac{1}{2}, 1)$ whenever $x \in (\frac{1}{2}, 1)$, we conclude that $\phi_{\sigma^{\text{maj}}}(x)$ is strictly concave for $x \in (\frac{1}{2}, 1)$ if $\lambda > 0$. By the same arguments as before, there is a unique informative steady state.

Part 3: Now, suppose $\lambda > \lambda^*$ and suppose σ^* is a social equilibrium.

Step 1: ϕ_{σ^*} must have a weakly misleading fixed point.

If not, then there exists some $\epsilon > 0$ so that $\phi_{\sigma^*}(x) - x > \epsilon$ for every x where $\lambda x + (1-\lambda)q \leq 0.5 + \epsilon$. By repeating the arguments in Part 1, Steps 2 and 3, we conclude $\sigma^* = \sigma^{\text{maj}}$.

But we show σ^{maj} has a strictly misleading fixed point for every $\lambda > \lambda^*$. By the definition of λ^*, we can choose some λ' with $\lambda^* \leq \lambda' < \lambda$ such that there exists a strictly misleading fixed point x' under σ^{maj} at λ' (we get "strictly" because by Lemma 1, $1/2$ is not a fixed point of σ^{maj} and all fixed points in $[0, 1/2)$ are strictly misleading). We rewrite the inflow accuracy function $\phi_\sigma(x)$ as $\phi_\sigma(x, \lambda)$ to make explicit its dependence on λ.

Observe $\phi_\sigma(x, \lambda)$ only depends on x and λ through the value of $\lambda x + (1-\lambda)q$. We can define x by

$$\lambda x + (1-\lambda)q = \lambda' x' + (1-\lambda')q.$$

Since $\lambda' < \lambda$ and $x' < q$, this equality implies that $x > x'$. For x' to be a strictly misleading fixed point under the majority rule we must have $\lambda' x' + (1-\lambda')q < \frac{1}{2}$, and therefore $x < \frac{1}{2}$ as well.

So

$$\phi_{\sigma^{\text{maj}}}(x, \lambda) = \phi_{\sigma^{\text{maj}}}(x', \lambda') = x',$$

where the second inequality holds because x' is a fixed point under σ^{maj} and λ'. So we conclude that $\phi_{\sigma^{\text{maj}}}(x, \lambda) < x$. Since $\phi_{\sigma^{\text{maj}}}(-1, \lambda) > 0$, by the intermediate value theorem there is some fixed point of $\phi_{\sigma^{\text{maj}}}$ between 0 and x. Since $x < \frac{1}{2}$, this is a strictly misleading fixed point, contradiction.

Note that since $\phi_{\sigma^*}(0) > 0$, the first weakly misleading fixed point of ϕ_{σ^*} is stable at least from the left, so it is also a weakly misleading steady state.

Step 2: ϕ_{σ^*} cannot have a fixed point with a sampling accuracy of exactly 1/2.

Each $\sigma^{(i)}$, by optimality, has the property that $\mathbb{E}[\sigma^{(i)}(1,k)] \geq \mathbb{E}[\sigma^{(i)}(-1,k)]$ for every $0 \leq k \leq K$. So we must have $\mathbb{E}[\sigma^*(1,k)] \geq \mathbb{E}[\sigma^*(-1,k)]$ for each $0 \leq k \leq K$. Suppose $\lambda x + (1-\lambda)q = 1/2$ and $\phi_{\sigma^*}(x) = x$. For each $0 \leq k < K/2$, we get

$$\begin{aligned}
&P_k(x,\lambda)\cdot[q\cdot\mathbb{E}[\sigma^*(1,k] + (1-q)\cdot\mathbb{E}[\sigma^*(-1,k)]] + P_{K-k}(x,\lambda)\cdot[q\cdot\mathbb{E}[\sigma^*(1,K-k)] + (1-q)\cdot\mathbb{E}[\sigma^*(-1,K-k)]] \\
=&P_k(x,\lambda)\cdot[q\cdot\mathbb{E}[\sigma^*(1,k)] + (1-q)\cdot\mathbb{E}[\sigma^*(-1,k)] + q\cdot\mathbb{E}[\sigma^*(1,K-k)] + (1-q)\cdot\mathbb{E}[\sigma^*(-1,K-k)]] \\
&\text{since } P_k(x,\lambda) = P_{K-k}(x,\lambda) \\
=&P_k(x,\lambda)\cdot[q\cdot\mathbb{E}[\sigma^*(1,k)] + (1-q)\cdot\mathbb{E}[\sigma^*(-1,k)] + q\cdot(C-\mathbb{E}[\sigma^*(-1,k)]) + (1-q)\cdot(C-\mathbb{E}[\sigma^*(1,k)])] \\
=&P_k(x,\lambda)\cdot[C + (2q-1)\cdot\mathbb{E}[\sigma^*(1,k)] - (2q-1)\mathbb{E}[\sigma^*(-1,k)]] \\
\geq&P_k(x,\lambda)\cdot C \text{ because } \mathbb{E}[\sigma^*(1,k)] \geq \mathbb{E}[\sigma^*(-1,k)] \text{ and } 2q-1>0 \\
\geq&P_k(x,\lambda)\cdot[C/2] + P_{K-k}(x,\lambda)\cdot[C/2]
\end{aligned}$$

Also, if $K/2$ is an integer, we have $\mathbb{E}[\sigma^*(1,K/2)] + \mathbb{E}[\sigma^*(-1,K/2)] = C$ and $\mathbb{E}[\sigma^*(1,K/2)] \geq \mathbb{E}[\sigma^*(-1,K/2)]$, so $q > 1-q$ implies $q\cdot\mathbb{E}[\sigma^*(1,K/2)] + (1-q)\cdot\mathbb{E}[\sigma^*(-1,K/2)] \geq C/2$. Thus, we conclude $\sum_{k=0}^K P_k(x,\lambda)\cdot[q\cdot\mathbb{E}[\sigma^*(1,k)] + (1-q)\cdot\mathbb{E}[\sigma^*(-1,k)]] \geq C/2$. So

$$\phi_{\sigma^*}(x) := \frac{q + \sum_{k=0}^K P_k(x,\lambda)\cdot[q\cdot\mathbb{E}[\sigma^*(1,k)] + (1-q)\cdot\mathbb{E}[\sigma^*(-1,k)]]}{1+C} \geq \frac{q + C/2}{1+C} > 1/2$$

since $q > 1/2$. But this means $\lambda\phi_{\sigma^*}(x) + (1-\lambda)q > 1/2$, contradiction. □

A.6 Proof of Proposition 3

Proof. Let $\lambda < \lambda' < \lambda^*$ and suppose that x^* is a steady state under λ. We want to show that there exists a steady state $(x')^* > x^*$ under λ'.

As in the proof of Part 3 of Theorem 2, let $\phi_\sigma(x, \lambda)$ be the inflow accuracy function with its dependence on λ. By monotonicity, $\phi_{\sigma^{\text{maj}}}(x, \lambda)$ is strictly increasing in λ when $x > q$. By Theorem 2, we have $x^* > q$ and therefore

$$x^* = \phi_{\sigma^{\text{maj}}}(x^*, \lambda) < \phi_{\sigma^{\text{maj}}}(x^*, \lambda').$$

Since $\phi_{\sigma^{\text{maj}}}(1, \lambda') < 1$, by the intermediate value theorem there exists $(x')^* \in (x^*, 1)$ such that

$$\phi_{\sigma^{\text{maj}}}((x')^*, \lambda') = (x')^*.$$

This is a steady state under λ' that is greater than x^*.

□

A.7 Proof of Proposition 4

Proof. Fix some K, and let $\lambda^*(q, C)$ be the critical virality value for (K, C, q). Write $\phi_{\sigma^{\text{maj}}}(x; \lambda, q, C)$ for the inflow accuracy function, emphasizing its dependence on the parameters.

First, note that if $\lambda^*(q, C) < 1$, then at every $\lambda \geq \lambda^*(q, C)$, $\phi_{\sigma^{\text{maj}}}(x; \lambda, q, C)$ has a strictly misleading steady state.

Part 1: $\lambda^*(q, C)$ increases when q increases. We have $q \mapsto \phi_{\sigma^{\text{maj}}}(x; \lambda, q, C)$ is always strictly increasing. For K odd, we may write $\phi_{\sigma^{\text{maj}}}(x; \lambda, q, C) = \frac{q + \sum_{k=(K+1)/2}^{K} P_k(x,\lambda) \cdot C}{1+C}$. For any $p' > p$, we can think of the experiment of tossing K coins each with a probability p' of landing heads as the experiment of first tossing K coins each with a probability p of landing heads, and then independently flipping each tail to a head with some probability h so that $p + (1-p)h = p'$. This shows that $\sum_{k=(K+1)/2}^{K} P_k(x, \lambda)$ is strictly increasing in $\lambda x + (1-\lambda)q$ (since the possibility of changing some tails to heads can only increase the total number of heads in the experiment), so it is weakly increasing in q. Also, the numerator of $\phi_{\sigma^{\text{maj}}}(x; \lambda, q, C)$ contains the term q, so this shows the entire numerator is strictly increasing in q. For K even, we may write $\phi_{\sigma^{\text{maj}}}(x; \lambda, q, C) =$

$\frac{q+[qP_{K/2}(x,\lambda)+\sum_{k=(K/2)+1}^{K}P_k(x,\lambda)]\cdot C}{1+C}$. For any fixed $\overline{q} < 1$, $\overline{q}P_{K/2}(x,\lambda) + \sum_{k=(K/2)+1}^{K} P_k(x,\lambda)$ is strictly increasing in $\lambda x + (1-\lambda)q$. This is because the possibility of changing some heads to tails keeps outcomes where $k \geq (K/2)+1$ in this same class, while outcomes with $k = K/2$ have a positive probability of changing to the class $k \geq (K/2)+1$, thus contribution 1 instead of $\overline{q} < 1$ to the sum. Hence $qP_{K/2}(x,\lambda) + \sum_{k=(K/2)+1}^{K} P_k(x,\lambda)$ is weakly increasing in q, and the numerator of $\phi_{\sigma^{\text{maj}}}(x;\lambda,q,C)$ is strictly increasing in q.

Suppose $\lambda^*(q,C) = 1$ and $\phi_{\sigma^{\text{maj}}}(x;\lambda,q,C)$ has no root in $x \in [0,1/2]$ for any $\lambda \in [0,1]$. Since $\phi_{\sigma^{\text{maj}}}(0;\lambda,q,C) > 0$, by continuity this means for every $\lambda \in [0,1]$, $\phi_{\sigma^{\text{maj}}}(x;\lambda,q,C) > x$ for each $x \in [0,1/2]$. For any $q' > q$, we have $\phi_{\sigma^{\text{maj}}}(x;\lambda,q',C) > \phi_{\sigma^{\text{maj}}}(x;\lambda,q,C) > x$ for every $x \in [0,1/2]$ and $\lambda \in [0,1]$. So again, $\lambda^*(q',C) = 1$.

Now suppose $\lambda^*(q,C) < 1$. We know also that $\lambda^*(q,C) > 0$ since $\phi_{\sigma^{\text{maj}}}(x;\lambda,q,C)$ has no fixed point in $x \in [0,1/2]$ when λ is near enough 0. If we have $\phi_{\sigma^{\text{maj}}}(x';\lambda^*(q,C),q,C) < x'$ for any $x' \in [0,1/2]$, then by continuity there is some $0 < \lambda < \lambda^*(q,C)$ that still has $\phi_{\sigma^{\text{maj}}}(x';\lambda,q,C) < x'$, which means $\phi_{\sigma^{\text{maj}}}(x;\lambda,q,C)$ has a root in $x \in [0,1/2)$ by the intermediate-value theorem. This contradicts the definition of $\lambda^*(q,C)$. So we must instead have $\phi_{\sigma^{\text{maj}}}(x;\lambda^*(q,C),q,C) \geq x$ for every $x \in [0,1/2]$. This means for every $q' > q$, $\phi_{\sigma^{\text{maj}}}(x;\lambda^*(q,C),q',C) > x$ for every $x \in [0,1/2]$, that is $\phi_{\sigma^{\text{maj}}}(x;\lambda^*(q,C),q',C)$ has no fixed point in $[0,1/2]$. This means that $\phi_{\sigma^{\text{maj}}}(x;\lambda^*(q,C),q',C)$ does not have a strictly misleading steady state. So either $\lambda^*(q',C) = 1$, or $\lambda^*(q',C) < 1$ but $\lambda^*(q,C) < \lambda^*(q',C)$.

Part 2: $\lambda^*(q,C)$ increases when C decreases. If $C' < C$, then $\phi_{\sigma^{\text{maj}}}(x;\lambda,q,C') > \phi_{\sigma^{\text{maj}}}(x;\lambda,q,C)$ at every x where $\lambda x + (1-\lambda)q \leq 1/2$. To see this, first suppose K is odd. Then at such x, $\sum_{k=(K+1)/2}^{K} P_k(x,\lambda) \leq 1/2$, and we have

$$\frac{d}{dC}\frac{q + \sum_{k=(K+1)/2}^{K} P_k(x,\lambda)\cdot C}{1+C} = \frac{\sum_{k=(K+1)/2}^{K} P_k(x,\lambda)\cdot(1+C) - (q + \sum_{k=(K+1)/2}^{K} P_k(x,\lambda)\cdot C)}{(1+C)^2}$$
$$= \frac{\sum_{k=(K+1)/2}^{K} P_k(x,\lambda) - q}{(1+C)^2} < 0$$

using the fact that $q > 1/2$.

If instead K is even, then we have $P_k(x,\lambda) \leq P_{K-k}(x,\lambda)$ for every $k \geq (K/2)+1$. This means $\sum_{k=(K/2)+1}^{K} P_k(x,\lambda) \leq \frac{1}{2}\cdot[1 - P_{K/2}(x,\lambda)]$, so then $qP_{K/2}(x,\lambda) + \sum_{k=(K/2)+1}^{K} P_k(x,\lambda) < q$ since

$q > 1/2$. We have:

$$\begin{aligned}
&\frac{d}{dC}\left(\frac{q+[qP_{K/2}(x,\lambda)+\sum_{k=(K/2)+1}^{K}P_k(x,\lambda)]\cdot C}{1+C}\right)\\
=&\frac{[qP_{K/2}(x,\lambda)+\sum_{k=(K/2)+1}^{K}P_k(x,\lambda)]\cdot(1+C)-(q+[qP_{K/2}(x,\lambda)+\sum_{k=(K/2)+1}^{K}P_k(x,\lambda)]\cdot C)}{(1+C)^2}\\
=&\frac{qP_{K/2}(x,\lambda)+\sum_{k=(K/2)+1}^{K}P_k(x,\lambda)-q}{(1+C)^2}<0
\end{aligned}$$

Suppose $\lambda^*(q,C)=1$ and $\phi_{\sigma^{\mathrm{maj}}}(x;\lambda,q,C)$ has no root in $x\in[0,1/2]$ for any $\lambda\in[0,1]$. Since $\phi_{\sigma^{\mathrm{maj}}}(0;\lambda,q,C)>0$, by continuity this means for every $\lambda\in[0,1]$, $\phi_{\sigma^{\mathrm{maj}}}(x;\lambda,q,C)>x$ for each x with $\lambda x+(1-\lambda)q\le 1/2$. For any $C'<C$, we have $\phi_{\sigma^{\mathrm{maj}}}(x;\lambda,q,C')>\phi_{\sigma^{\mathrm{maj}}}(x;\lambda,q,C)>x$ for every x with $\lambda x+(1-\lambda)q\le 1/2$ and $\lambda\in[0,1]$. That is, $\phi_{\sigma^{\mathrm{maj}}}(x;\lambda,q,C')$ does not have a strictly misleading fixed point for any $\lambda\in[0,1]$, which means $\lambda^*(q,C')=1$.

Now suppose $\lambda^*(q,C)<1$. Again, we have $\lambda^*(q,C)>0$ and $\phi_{\sigma^{\mathrm{maj}}}(x;\lambda^*(q,C),q,C)\ge x$ for every $x\in[0,1/2]$ by similar arguments as before. This means for every $C'<C$, $\phi_{\sigma^{\mathrm{maj}}}(x;\lambda^*(q,C),q,C')>x$ for every x with $\lambda x+(1-\lambda)q\le 1/2$, that is $\phi_{\sigma^{\mathrm{maj}}}(x;\lambda^*(q,C),q,C')$ has no strictly misleading steady state. So either $\lambda^*(q,C')=1$, or $\lambda^*(q,C')<1$ but $\lambda^*(q,C)<\lambda^*(q,C')$.

Part 3: Comparative statics in K. Now, fix q and C. For simplicity, denote $\phi_{\sigma^{\mathrm{maj}}}(x;\lambda,q,K,C)$ by $\phi(x;\lambda,K)$, and let $p:=\lambda x+(1-\lambda)q$. Then, $P_k^{(K)}(x,\lambda)=\binom{K}{k}p^k(1-p)^{K-k}$. We can rewrite:

$$P_k^{(K+1)}(x,\lambda)=p\cdot P_{k-1}^{(K)}(x,\lambda)+(1-p)P_k^{(K)}(x,\lambda). \tag{11}$$

By the same arguments as in **Part 1** and **Part 2**, it suffices to show that for $0<\lambda<1$ and for every x such that $\lambda x+(1-\lambda)q<1/2$:

- If K is odd, then $\phi(x;\lambda,K+1)>\phi(x;\lambda,K)$
- If $K+1$ is even, then $\phi(x;\lambda,K+1)>\phi(x;\lambda,K+2)$
- For any K, $\phi(x;\lambda,K)>\phi(x;\lambda,K+2)$

Case 1: K is odd (K to $K+1$).

Note that

$$\phi(x;\lambda,K)=\frac{q+\sum_{k=\frac{K+1}{2}}^{K}P_k^{(K)}(x,\lambda)\cdot C}{1+C}$$

and

$$\phi(x;\lambda,K+1)=\frac{q+q\cdot P^{(K+1)}_{\frac{K+1}{2}}(x,\lambda)\cdot C+\sum_{k=\frac{K+3}{2}}^{K+1}P^{(K+1)}_k(x,\lambda)\cdot C}{1+C}$$

Applying Equation (11), we have

$$\begin{aligned}\sum_{k=\frac{K+3}{2}}^{K+1}P^{(K+1)}_k(x,\lambda)&=p\sum_{k=\frac{K+3}{2}}^{K+1}P^{(K)}_{k-1}(x,\lambda)+(1-p)\sum_{k=\frac{K+3}{2}}^{K+1}P^{(K)}_k(x,\lambda)\\&=p\sum_{k=\frac{K+1}{2}}^{K}P^{(K)}_{k}(x,\lambda)+(1-p)\sum_{k=\frac{K+3}{2}}^{K}P^{(K)}_k(x,\lambda)\\&=p\cdot P^{(K)}_{\frac{K+1}{2}}(x,\lambda)+\sum_{k=\frac{K+3}{2}}^{K}P^{(K)}_k(x,\lambda).\end{aligned}$$

Then,

$$\begin{aligned}&\phi(x;\lambda,K)-\phi(x;\lambda,K+1)\\&=\frac{-q\cdot P^{(K+1)}_{\frac{K+1}{2}}(x,\lambda)\cdot C+\left[\sum_{k=\frac{K+1}{2}}^{K}P^{(K)}_k(x,\lambda)-\sum_{k=\frac{K+3}{2}}^{K+1}P^{(K+1)}_k(x,\lambda)\right]\cdot C}{1+C}\\&=\frac{-q\cdot P^{(K+1)}_{\frac{K+1}{2}}(x,\lambda)\cdot C+\left[P^{(K)}_{\frac{K+1}{2}}(x,\lambda)-p\cdot P^{(K)}_{\frac{K+1}{2}}(x,\lambda)\right]\cdot C}{1+C}\\&=\frac{-q\cdot P^{(K+1)}_{\frac{K+1}{2}}(x,\lambda)\cdot C+(1-p)\cdot P^{(K)}_{\frac{K+1}{2}}(x,\lambda)\cdot C}{1+C}\\&=\frac{\left(-q+\frac{1}{2}\right)\cdot P^{(K+1)}_{\frac{K+1}{2}}(x,\lambda)\cdot C}{1+C}<0,\ \left(\text{using } (1-p)\cdot P^{(K)}_{\frac{K+1}{2}}(x,\lambda)=\frac{1}{2}P^{(K+1)}_{\frac{K+1}{2}}(x,\lambda)\right)\end{aligned}$$

since $q>\frac{1}{2}$.

Case 2: $K+1$ is even ($K+1$ to $K+2$).

Note that

$$\phi(x;\lambda,K+1)=\frac{q+q\cdot P^{(K+1)}_{\frac{K+1}{2}}(x,\lambda)\cdot C+\sum_{k=\frac{K+3}{2}}^{K+1}P^{(K+1)}_k(x,\lambda)\cdot C}{1+C}$$

and

$$\phi(x;\lambda,K+2)=\frac{q+\sum_{k=\frac{K+3}{2}}^{K+2}P^{(K+2)}_k(x,\lambda)\cdot C}{1+C}$$

As in the first case, we have

$$\sum_{k=\frac{K+3}{2}}^{K+2} P_k^{(K+2)}(x,\lambda) = p \cdot P_{\frac{K+1}{2}}^{(K+1)}(x,\lambda) + \sum_{k=\frac{K+3}{2}}^{K+1} P_k^{(K+1)}(x,\lambda).$$

Therefore,

$$\begin{aligned} &\phi(x;\lambda,K+1) - \phi(x;\lambda,K+2) \\ &= \frac{(q-p)\cdot P_{\frac{K+1}{2}}^{(K+1)}(x,\lambda)\cdot C}{1+C} > 0, \end{aligned}$$

since $q > \lambda x + (1-\lambda)q = p$ for all $x \in [0,1/2)$ and $\lambda > 0$.

Case 3: K **to** $K+2$ **for odd** K.

Combining results from **Case 1** and **Case 2**,

$$\begin{aligned} \phi(x;\lambda,K) - \phi(x;\lambda,K+2) &= \phi(x;\lambda,K) - \phi(x;\lambda,K+1) + \phi(x;\lambda,K+1) - \phi(x;\lambda,K+2) \\ &= \frac{\left(-q+\frac{1}{2}\right)\cdot P_{\frac{K+1}{2}}^{(K+1)}(x,\lambda)\cdot C + (q-p)\cdot P_{\frac{K+1}{2}}^{(K+1)}(x,\lambda)\cdot C}{1+C} \\ &= \frac{\left(-p+\frac{1}{2}\right)\cdot P_{\frac{K+1}{2}}^{(K+1)}(x,\lambda)\cdot C}{1+C} > 0, \end{aligned}$$

if $p = \lambda x + (1-\lambda)q < \frac{1}{2}$.

Case 4: $K+1$ **to** $K+3$ **for even** $K+1$.

Combining results from **Case 2** and **Case 1**,

$$\begin{aligned} \phi(x;\lambda,K+1) - \phi(x;\lambda,K+3) &= \phi(x;\lambda,K+1) - \phi(x;\lambda,K+2) + \phi(x;\lambda,K+2) - \phi(x;\lambda,K+3) \\ &= \frac{(q-p)\cdot P_{\frac{K+1}{2}}^{(K+1)}(x,\lambda)\cdot C}{1+C} + \frac{\left(-q+\frac{1}{2}\right)\cdot P_{\frac{K+3}{2}}^{(K+3)}(x,\lambda)\cdot C}{1+C}. \end{aligned}$$

We have $P_{\frac{K+1}{2}}^{(K+1)}(x,\lambda) \geq P_{\frac{K+3}{2}}^{(K+3)}(x,\lambda)$, so this expression is weakly larger than

$$\frac{\left(-q+\frac{1}{2}\right)\cdot P_{\frac{K+3}{2}}^{(K+3)}(x,\lambda)\cdot C}{1+C} + \frac{(q-p)\cdot P_{\frac{K+3}{2}}^{(K+3)}(x,\lambda)\cdot C}{1+C} = \frac{(\frac{1}{2}-p)\cdot P_{\frac{K+3}{2}}^{(K+3)}(x,\lambda)\cdot C}{1+C} > 0$$

when $p = \lambda x + (1-\lambda)q < 1/2$. $\square$

A.8 Proof of Proposition 5

Proof. To show that $\lambda^*(q, K, C) > 1 - \frac{1}{2q}$ for any q, K, C, we prove that $\phi_{\sigma^{\text{maj}}}(x)$ does not have fixed points in $[0, 1/2]$ when $\lambda \le 1 - \frac{1}{2q}$. We have $\lambda x + (1-\lambda)q \ge (1-\lambda)q = \frac{1}{2}$. This means that for K odd, $\sum_{k=(K+1)/2}^{K} P_k(x, \lambda) \ge 1/2$. For K even, we have $q \cdot P_{K/2}(x, \lambda) + \sum_{k=(K/2)+1}^{K} P_k(x, \lambda) > 1/2$. So in both cases, the numerator of $\phi_{\sigma^{\text{maj}}}(x)$ is at least $q + C/2$, which means $\phi_{\sigma^{\text{maj}}}(x) \ge \frac{q+C/2}{1+C} > 1/2$ since $q > 1/2$. This shows $\phi_{\sigma^{\text{maj}}}(x) > x$ for every $x \in [0, 1/2]$.

Next, fix any $1/2 < q < 1$ and any $\overline{\lambda} > 1 - \frac{1}{2q}$. Let $0 < x' < 1/2$ be any number such that $\overline{\lambda}x' + (1-\overline{\lambda})q < 1/2$ (such x' exists by the bound on $\overline{\lambda}$). We find integers $\underline{K}$ and $\underline{C}$ so that whenever $K \ge \underline{K}$, $C \ge \underline{C}$, we get $\phi_{\sigma^{\text{maj}}}(x'; q, K, C, \overline{\lambda}) < x'$. Since $x' < 1/2$ and since $\phi_{\sigma^{\text{maj}}}(0; q, K, C, \overline{\lambda}) > 0$, we know that $\phi_{\sigma^{\text{maj}}}(\cdot; q, K, C, \overline{\lambda})$ has a fixed point in $(0, 1/2)$ by the intermediate-value theorem. This implies $\lambda^*(q, C, K) \le \overline{\lambda}$.

To construct $\underline{K}$ and $\underline{C}$, let $\epsilon = x'/2$. By the law of large numbers, we can find $\underline{K}$ so that whenever $K \ge \underline{K}$, the probability that a binomial distribution with K trials and success probability $\overline{\lambda}x' + (1-\overline{\lambda})q < 1/2$ has strictly less than $K/2$ successes is larger than $1 - \epsilon$. Thus, whenever $K \ge \underline{K}$, $\phi_{\sigma^{\text{maj}}}(x'; q, K, C, \overline{\lambda}) \le \frac{q+\epsilon C}{1+C}$. Now, increasing $\underline{K}$ further if necessary, we can find $\underline{C}$ large enough so that for all $C \ge \underline{C}$, we have $\frac{q+\epsilon C}{1+C} < 2\epsilon$. Whenever $K \ge \underline{K}$ and $C \ge \underline{C}$, we have $\phi_{\sigma^{\text{maj}}}(x'; q, K, C, \overline{\lambda}) < 2\epsilon = x'$ as desired. $\square$

A.9 Proof of Proposition 6

Proof. Suppose that $\liminf_n \lambda_n < \lambda^*$. Then there exists $\hat{\lambda} < \lambda^*$ such that $\lambda_n < \hat{\lambda}$ infinitely often. By passing to a subsequence if necessary, we can find some $\overline{\lambda} < \lambda^*$ and a sequence $(n_j) \in \mathbb{N}$ so that $\lim_{j\to\infty} \lambda_{n_j} = \overline{\lambda}$ and $\lim_{j\to\infty} \sigma_{(n_j)} = \sigma^*$ for some strategy σ^*. Fix $0 < \delta < \lambda^* - \overline{\lambda}$. We will show that $W_n(\overline{\lambda} + \delta) > W_n(\lambda_n)$ for some large n, which contradicts the optimality of λ_n.

By Theorem 2, for $\lambda < \lambda^*$ the unique social equilibrium is σ^{maj} and there is a unique steady state $x^*(\lambda) > \frac{1}{2}$ under σ^{maj}. We now show a lemma, which says that when $\lambda = \overline{\lambda} + \delta$, the value of $x(t)$ under the equilibrium strategy is likely to be at least $x^*(\overline{\lambda} + \delta/2)$ except perhaps in the first finitely many periods.

Lemma 6. *Fix $\epsilon > 0$. For each n, let $\sigma_n^*(\overline{\lambda}+\delta)$ be the equilibrium strategy chosen by the measurable selection. There exists N and T such that with n agents, when $\lambda = \overline{\lambda}+\delta$ and the strategy is $\sigma_n^*(\overline{\lambda}+\delta)$,*

$$\mathbb{P}_{\sigma_n^*(\overline{\lambda}+\delta)}[x(t) < x^*(\overline{\lambda} + \delta/2)] < \epsilon$$

for all $t \geq T$ and $n \geq N$.

Proof. By Theorem 2, the unique social equilibrium is σ^{maj}. So we must have $\sigma_n^*(\overline{\lambda} + \delta) \to \sigma^{\text{maj}}$. By Proposition 3, we have $x^*(\overline{\lambda} + \delta) > x^*(\overline{\lambda} + 3\delta/4)$. So $\phi_{\sigma^{\text{maj}}}^{\bar{\lambda}+\delta}(x) - x > 0$ on $[0, x^*(\overline{\lambda} + 3\delta/4)]$. We can choose small enough $\epsilon' > 0$ and some appropriate p so that:

- $p + \epsilon'/2 = (\overline{\lambda} + \delta)x^*(\overline{\lambda} + \delta/2) + (1 - \overline{\lambda} - \delta)q$;
- $p + 2\epsilon' < (\overline{\lambda} + \delta)x^*(\overline{\lambda} + 3\delta/4) + (1 - \overline{\lambda} - \delta)q$; and
- $\phi_{\sigma^{\text{maj}}}^{\overline{\lambda}+\delta}(x) - x \geq 2\epsilon'$ for every x with $x \in [0, x^*(\overline{\lambda} + 3\delta/4)]$ (by continuity of $\phi_{\sigma^{\text{maj}}}^{\overline{\lambda}+\delta}(x)$).

Apply Lemma 5 with these values of ϵ', p and with $\epsilon'' = \epsilon$, making use of the convergence $\sigma_n^*(\bar{\lambda}+\delta) \to \sigma^{\text{maj}}$, there is some T, N so that whenever $t \geq T$ and $n \geq N$, we get

$$\mathbb{P}_{\sigma_n^*(\overline{\lambda}+\delta)}[x(t) < x^*(\overline{\lambda} + \delta/2)] < \epsilon$$

as desired. □

Applying the lemma we have just proven, we can choose T and N such that

$$\mathbb{P}_{\sigma_n^*(\overline{\lambda}+\delta)}[x(t) < x^*(\overline{\lambda} + \delta/2)] < \epsilon$$

whenever $t \geq T$ and $n \geq N$. Increasing T if necessary, by the Chernoff bound, we can assume that $|z(t) - q| < \epsilon$ for all $t \geq T$ with probability at least $1 - \epsilon$ whenever $n \geq N$. Finally, since $\sigma_n^*(\overline{\lambda}+\delta) \to \sigma^{\text{maj}}$, we may find some N' so that whenever $n \geq N'$, $|\sigma_n^*(\overline{\lambda} + \delta)(s,k)(z) - \sigma^{\text{maj}}(s,k)(z)| < \epsilon$ for every s, k. For $n \geq \max(N, N')$, we can therefore bound welfare when $\lambda = \overline{\lambda} + \delta$ from below as:

$$\begin{aligned} W_n(\overline{\lambda} + \delta) \geq (1 - 2\epsilon) \cdot \frac{n - T}{n - K} \cdot (&\mathbb{P}[\text{Binom}(K, (\overline{\lambda} + \delta)x^*(\overline{\lambda} + \delta/2) + (1 - (\overline{\lambda} + \delta))(q - \epsilon)) > K/2] \\ &+ \mathbb{P}[\text{Binom}(K, (\overline{\lambda} + \delta)x^*(\overline{\lambda} + \delta/2) + (1 - (\overline{\lambda} + \delta))(q - \epsilon)) = K/2]q) \cdot (1 - \epsilon)Cu. \quad (12) \end{aligned}$$

Lemma 7. *Fix $\epsilon > 0$. For each n, let $\sigma_n^*(\lambda_n)$ be the equilibrium strategy chosen by the measurable selection. There exists a subsequence $(n_j) \to \infty$ and integers N and T such that with n_j agents, when $\lambda = \lambda_{n_j}$ and the strategy is $\sigma_{n_j}^*(\lambda_{n_j})$,*

$$\mathbb{P}_{\sigma^*_{n_j}(\lambda_{n_j})}[x(t) > x^*(\overline{\lambda} + \delta/4)] < \epsilon$$

for all $t \geq T$ and $n_j \geq N$.

Proof. Take the sequence $(n_j) \in \mathbb{N}$ so that $\lim_{j\to\infty} \lambda_{n_j} = \overline{\lambda}$ and $\lim_{j\to\infty} \sigma_{(n_j)} = \sigma^*$ for some strategy σ^*. We show that σ^* is the majority rule. This argument follows the same steps as **Part 1** in the proof of Theorem 2.

Step 1: All fixed points of $\phi^{\overline{\lambda}}_{\sigma^*}$ are strictly informative.

This is because σ^* satisfies the hypotheses of Lemma 3 for the same reason as in the proof of Theorem 2, so the same arguments about $\phi^{\overline{\lambda}}_{\sigma^*}$ go through.

Step 2: $\phi^{\overline{\lambda}}_{\sigma^*}$ only has fixed points in $(q, 1]$.

We now verify that σ^* is such that $\sigma^*(1,k)(C) = 1$ for every $k \geq K/2$, and thus satisfies the hypotheses of Lemma 4. By **Step 1**, we know that there is some $\epsilon' > 0$ so that $\phi^{\overline{\lambda}}_{\sigma^*}(x) - x \geq 2\epsilon'$ for every x with $\overline{\lambda}x + (1-\overline{\lambda})q \leq 0.5 + 2\epsilon'$. Find some $\epsilon'' > 0$ so that

$$\frac{(0.5+\epsilon'/4)^{\lfloor K/2\rfloor+1}\cdot(0.5-\epsilon'/4)^{K-\lfloor K/2\rfloor-1}\cdot(1-3\epsilon'')}{(0.5+\epsilon'/4)^{K-\lfloor K/2\rfloor-1}\cdot(0.5-\epsilon'/4)^{\lfloor K/2\rfloor+1}\cdot(1-3\epsilon'')+3\epsilon''} > 1. \tag{13}$$

Apply Lemma 5 to these ϵ', ϵ'' and $p = 1/2$ to find N and δ. Also, by the law of large numbers, we may find N' so that $\mathbb{P}[|\text{Binom}(t,q)/t - q| > \epsilon'/4] < \epsilon''$ whenever $t \geq N'$, where $\text{Binom}(t,q)$ refers to a binomial random variable with t trials and q success probability. Since $\lambda_{n_j} \to \overline{\lambda}$ and $\sigma_{(n_j)} \to \sigma^*$, there exists J so that $\|\sigma_{(n_j)} - \sigma^*\|_2 < \delta$ and $|\lambda_{n_j} - \overline{\lambda}| < \delta$ whenever $j \geq J$. For all $j \geq J$ large enough, we must have $\max(N,N')/(n_j - K) < \epsilon''$. But in the equilibrium $\sigma_{(n_j)}$ with virality weight λ_{n_j}, we know that agents' equilibrium belief about sampling accuracy for any position in $\{\max(N,N'), \max(N,N')+1, ..., n_j\}$ assigns mass at least $1-2\epsilon''$ to the region $[0.5+\epsilon'/4, 1]$. This is because if we have both $\lambda_{n_j}x(t) + (1-\lambda_{n_j})q \geq 0.5+\epsilon'/2$ and $|z(t) - q| < \epsilon'/4$, then we also have $\lambda_{n_j}x(t) + (1-\lambda_{n_j})z(t) \geq 0.5+\epsilon'/4$ — the former event has probability at least $1-\epsilon''$ when $t \geq N$ and latter event has probability at least $1-\epsilon''$ when $t \geq N'$. Hence, equilibrium belief about sampling accuracy for a uniformly random position in $\{K+1, K+2, ..., \max(N, N'), \max(N, N')+1, ..., n_j\}$ assigns mass at least $1-3\epsilon''$ to the same region. Thus, the expression on the LHS of Equation (13) gives a lower bound on the the posterior likelihood ratio of $\omega = 1$ and $\omega = 0$ after seeing a sample containing k positive stories in equilibrium $\sigma_{(n_j)}$ with virality weight λ_{n_j}, for any $k > K/2$. Hence, by optimality, $\sigma_{(n_j)}(1, k)(C) = 1$ for every $k > K/2$. Also, for any belief about sampling accuracy, a sample with $k = K/2$ is uninformative, so if $K/2$ is an integer then $\sigma_{(n_j)}(1, K/2)(C) = 1$ by optimality. Thus we see for all large enough j, $\sigma_{(n_j)}(1, k)(C) = 1$ for every $k \geq K/2$, hence the same must hold for the limit σ^*.

Combining the conclusion of Lemma 4 with the previous step (which says every fixed point x of $\phi^{\overline{\lambda}}_{\sigma^*}$ satisfies $\overline{\lambda}x + (1-\overline{\lambda})q > 1/2$), we have shown that $\phi^{\overline{\lambda}}_{\sigma^*}$ only has fixed points in $(q, 1]$.

Step 3: $\sigma^* = \sigma^{\text{maj}}$.

To finally show that $\sigma^* = \sigma^{\text{maj}}$, by state symmetry it suffices to show that $\sigma^*(-1, k)(C) = 1$ for every $k > K/2$. Since $\phi^{\overline{\lambda}}_{\sigma^*}$ only has fixed points in $(q, 1]$, there exists some $\epsilon' > 0$ so that $\phi^{\overline{\lambda}}_{\sigma^*}(x) - x \geq 2\epsilon'$ for every x where $\overline{\lambda}x + (1-\overline{\lambda})q \leq q + 2\epsilon'$. Find some $\epsilon'' > 0$ so that

$$\frac{(q+\epsilon'/4)^{\lfloor K/2 \rfloor + 1} \cdot (1-q-\epsilon'/4)^{K - \lfloor K/2 \rfloor - 1} \cdot (1-3\epsilon'')}{(q+\epsilon'/4)^{K - \lfloor K/2 \rfloor - 1} \cdot (1-q-\epsilon'/4)^{\lfloor K/2 \rfloor + 1} \cdot (1-3\epsilon'') + 3\epsilon''} > \frac{q}{1-q}. \tag{14}$$

Apply Lemma 5 to these ϵ', ϵ'' and $p = q$ to find N and δ. Also, by the law of large numbers, we may find N' so that $\mathbb{P}[|\text{Binom}(t, q)/t - q| > \epsilon'/4] < \epsilon''$ whenever $t \geq N'$, where $\text{Binom}(t, q)$ refers to a binomial random variable with t trials and q success probability. Since $\lambda_{n_j} \to \overline{\lambda}$ and $\sigma_{(n_j)} \to \sigma^*$, there exists J so that $\|\sigma_{(n_j)} - \sigma^*\|_2 < \delta$ and $|\lambda_{n_j} - \overline{\lambda}| < \delta$ whenever $j \geq J$. For all $j \geq J$ large enough, by the same arguments as in **Step 2**, the expression on the LHS of Equation (14) gives a lower bound on the the posterior likelihood ratio of $\omega = 1$ and $\omega = -1$ after seeing a sample containing k positive stories in equilibrium $\sigma_{(n_j)}$ with virality weight λ_{n_j}, for any $k > K/2$. So even if the private signal is $s = -1$, the $\omega = 1$ state is still strictly more likely, and $\sigma_{(n_j)}(-1, k)(C) = 1$ by optimality. So we also have in the limit $\sigma^*(-1, k)(C) = 1$ for every $k > K/2$.

The remainder of the proof proceeds symmetrically to the proof of Lemma 6 applied to the subsequence (n_j), and we omit the details. $\square$

Applying the lemma we have just proven, we can choose a subsequence (n_j) and integers T and

N such that

$$\mathbb{P}_{\sigma^*_{n_j}(\lambda_{n_j})}[x(t) > x^*(\overline{\lambda} + \delta/4)] < \epsilon$$

whenever $t \geq T$ and $n_j \geq N$. Increasing T if necessary, by the Chernoff bound, we can assume that $|z(t) - q| < \epsilon$ for all $t \geq T$ with probability at least $1 - \epsilon$ whenever $n_j \geq N$. Again using the convergence of the equilibrium strategies to the majority rule,we may find some N' so that whenever $n_j \geq N'$, $|\sigma^*_{n_j}(\overline{\lambda} + \delta)(s,k)(z) - \sigma^{\mathrm{maj}}(s,k)(z)| < \epsilon$ for every s, k. For $n_j \geq \max(N, N')$, we can therefore bound welfare when $\lambda = \lambda_{n_j}$ from above as:

$$\begin{aligned} W_{n_j}(\lambda_{n_j}) \leq (1-2\epsilon) \cdot \frac{n_j - T}{n_j - K} \cdot (&\mathbb{P}[\mathrm{Binom}(K, \lambda_{n_j} x^*(\overline{\lambda} + \delta/4) + (1 - \lambda_{n_j})(q + \epsilon)) > K/2] + \\ &\mathbb{P}[\mathrm{Binom}(K, \lambda_{n_j} x^*(\overline{\lambda} + \delta/4) + (1 - \lambda_{n_j})(q + \epsilon)) = K/2]q) \cdot Cu + \\ &(\frac{T - K}{n_j - K} + 2\epsilon \cdot \frac{n_j - T}{n_j - K}) \cdot Cu + \epsilon C^2 u. \end{aligned}$$

Recall that inequality (12) stated

$$\begin{aligned} W_n(\overline{\lambda} + \delta) \geq (1-2\epsilon) \cdot \frac{n - T}{n - K} \cdot (&\mathbb{P}[\mathrm{Binom}(K, (\overline{\lambda} + \delta) x^*(\overline{\lambda} + \delta/2) + (1 - (\overline{\lambda} + \delta))(q - \epsilon)) > K/2] + \\ &\mathbb{P}[\mathrm{Binom}(K, (\overline{\lambda} + \delta) x^*(\overline{\lambda} + \delta/2) + (1 - (\overline{\lambda} + \delta))(q - \epsilon)) = K/2]q) \cdot (1 - \epsilon) Cu. \end{aligned}$$

Since $\lambda_{n_j} < \overline{\lambda} + \delta$ for j sufficiently large and $x^*(\overline{\lambda} + \delta/4) < x^*(\overline{\lambda} + \delta/2)$, we can choose ϵ sufficiently small and j sufficiently large such that these inequalities imply

$$W_{n_j}(\lambda_{n_j}) < W_{n_j}(\overline{\lambda} + \delta).$$

But this contradicts the optimality of λ_{n_j}, so we must have $\liminf_n \lambda_n \geq \lambda^*$. $\square$