Fake News Propagation and Detection: A Sequential Model

Yiangos Papanastasiou
Haas School of Business · University of California, Berkeley
yiangos@haas.berkeley.edu

In the wake of the 2016 US presidential election, social media platforms are facing increasing pressure to combat the propagation of "fake news" (i.e., articles whose content is fabricated). Motivated by recent attempts in this direction, we consider the problem faced by a social media platform that is observing the sharing actions of a sequence of rational agents and is dynamically choosing whether to conduct an inspection (i.e., a "fact-check") of an article whose validity is ex ante unknown. We first characterize the agents' inspection and sharing actions and establish that in the absence of any platform intervention, the agents' news-sharing process is prone to the proliferation of fabricated content, even when the agents are intent on sharing only truthful news. We then study the platform's inspection problem. We find that because the optimal policy is adapted to crowdsource inspection from the agents, it exhibits features that may appear a priori nonobvious; most notably, we show that the optimal inspection policy is nonmonotone in the ex ante probability that the article being shared is fake. We also investigate the effectiveness of the platform's policy in mitigating the detrimental impact of fake news on the agents' learning environment. We demonstrate that in environments characterized by a low (high) prevalence of fake news, the platform's policy is more effective when the rewards it collects from content sharing are low relative to the penalties it incurs from the sharing of fake news (when the rewards it collects from content sharing are high in absolute terms).

Key words: fake news, social learning, crowdsourcing, platform operations

1. Introduction

In the run-up to the 2016 US presidential election, articles containing fabricated information (now widely referred to as "fake news") were actively shared millions of times between US voters on social media platforms such as Facebook, YouTube, and Twitter. Alarmingly, one post-election analysis found that in the final three months of the presidential campaign, the 20 top-performing fake news stories, which included

"Pope Francis Shocks World, Endorses Donald Trump for President" and
"ISIS Leader Calls for American Muslim Voters to Support Hillary Clinton,"

received more Facebook engagements than the 20 top-performing truthful news stories (Buzzfeed 2016; see also Vosoughi et al. 2018). While the debate over the impact of such since-debunked news stories is still ongoing, the possibility that fake news may have played a role in determining the US presidency has raised serious concerns from individuals and organizations throughout society.

In the aftermath of the election, much of the blame for the fake news epidemic has fallen on the platforms that facilitate the sharing of this content. In response to increasing pressure to combat the propagation of fake news, social media platforms have recently adopted various measures, both preventive (i.e., with the goal of curbing the production of fake news) and corrective (i.e., with the goal of detecting and removing fake news already in circulation). This paper is motivated by attempts falling in the latter category, the most prominent example of which is Facebook's approach of forming partnerships with third-party news organizations (e.g., ABC News, the Associated Press) and performing fact-checks of articles whenever this is deemed necessary (Financial Times 2017, Tech Times 2017). We aim to develop insights regarding both the optimal implementation of such measures from the platform's perspective, but also their impact on the society's learning environment.

To do so, we develop a model of news propagation, where a sequence of rational agents receive a news article of unknown validity, decide first whether to conduct a costly inspection of its contents (i.e., a fact-check), and then whether to share it with the next agent. With this model as a basis, we consider the problem faced by a social media platform that is observing the sharing actions of its users and is dynamically choosing whether to intervene by conducting its own costly inspection and disclosing its findings. In choosing whether to conduct an inspection, we assume that the platform collects a reward whenever its users share content on the platform (e.g., from ad revenues), but incurs a penalty whenever the content being shared is fake (e.g., from loss of goodwill).

We first study the platform's optimal inspection policy. We start by establishing that in the absence of any platform intervention, the more an article is initially shared by the agents, the more likely it is to be shared even further. As a result, articles are seen to gain momentum up until a critical number of shares, after which they "go viral" (i.e., they are shared by the agents in perpetuity). Building on this result, we demonstrate that the platform's optimal inspection policy reduces to the solution of a finite-horizon optimal stoping problem, whereby the platform either conducts an inspection before the article reaches its critical number of shares, or it conducts no inspection at all. Because the platform's optimal policy is adapted to crowdsource inspection from the agents, we find that it exhibits features that may appear a priori nonobvious. Most notably, we show that the policy is nonmonotone in the ex ante probability that the article being shared is fake: when this probability is very low, the platform opts not to conduct an inspection; when it is moderately low, the platform conducts an inspection from the onset of the sharing process; when it is moderate, the platform conducts a delayed inspection; and when it is high, the platform either conducts an inspection from the onset or no inspection at all. We characterize the aforementioned cases analytically, and extract various insights on how the platform's optimal policy depends on its sharing rewards and penalties, the article's contents, and the agents' characteristics.

Having established the platform's optimal inspection policy in the presence of fake news, we then consider the implications of this policy for the agents' learning environment. We point out that the platform's objective of maximizing sharing rewards is not necessarily aligned with that of ensuring efficiency in the social learning process, and investigate this misalignment in detail.

Using our model of the agents' news-sharing process, we first quantify the expected impact of a news article on the agents' beliefs as a function of the article's validity. In doing so, we identify two sources of inefficiency, which we refer to as the "direct" and "indirect" effects of fake news: on one hand, the circulation of fake news directly impacts the agents' beliefs, through fake news articles that evade detection long enough to go viral; on the other, the presence of fake news also impacts the agents' beliefs indirectly, by causing truthful articles to be discontinued by agents who are suspicious of information that contradicts their prior beliefs, and/or by reducing the capacity of truthful articles to influence the agents' beliefs. We then analyze whether and to what extent the platform's inspection policy restores efficiency in the agents' learning environment. We find that the effectiveness of the platform's policy in this respect exhibits an asymmetric structure: while the direct effect of fake news is mitigated as long as the platform's policy prescribes an inspection before an article goes viral, the indirect effect is mitigated only if this inspection occurs from the onset of the sharing process. Furthermore, we demonstrate that when the proportion of fake news in circulation is low, the platform's policy is more effective when its rewards from content sharing are low relative to the penalties it incurs from the sharing of fake news; by contrast, when the proportion of fake news in circulation is high, the platform's policy is more effective when its rewards from content sharing are high in absolute terms.

2. Related Literature

At the core of our paper is a model of social learning, where the objective is to discern whether the contents of a news article are truthful or fabricated. Our model utilizes a sequential decision paradigm which at a high level is much in the spirit of the seminal papers by Banerjee (1992) and Bikhchandani et al. (1992): after observing the actions of her predecessors, each agent takes an action whose payoff is related to some unobservable state of the world (see also Acemoglu et al. (2011) and Drakopoulos et al. (2013), where each agent observes only a subset of her predecessors). Our paper shares with these studies the observation that sequential decision making often leads to inefficient outcomes, where the learning process is seen to converge to erroneous conclusions.

Indeed, several studies that incorporate the sequential paradigm in a wide range of contexts arrive at similar conclusions; for instance, Veeraraghavan and Debo (2009) find that customers may choose a server of inferior quality as a result of observing the queue-joining behavior of their predecessors, while Scharfstein and Stein (1990) show that managers may choose to ignore their own information in favor of mimicking the investments of other managers. In the context of fake news, our focus is on how this phenomenon affects the virality of fake news and their impact on the agents' beliefs, and on how it interacts with the content inspection policy of an online social media platform.

The implications of social learning for various operational decisions have received significant attention in the recent OM literature. Among such work, Crapis et al. (2017), Papanastasiou and Savva (2016), Shin and Zeevi (2017) and Yu et al. (2015) focus on the impact of consumer reviews on optimal pricing policies; Papanastasiou (2018) and Papanastasiou et al. (2016) identify implications of social learning for inventory decisions; Feldman et al. (2018) show that the presence of social learning may lead to a decrease in the quality of new experience goods. In a related line of work, other papers study the effects of social interactions more broadly defined. For instance, Candogan et al. (2012) study optimal pricing policies when consumers embedded in a social network experience local consumption externalities, while Momot et al. (2016) investigate the value of social network information when selling to conspicuous consumers. Manshadi and Misra (2016) study new product diffusion in networks with limited social interactions. Allon and Zhang (2017) consider how customers' social network affects a service firm's optimal service-level differentiation strategy. In a two-product newsvendor setting, Hu et al. (2015) consider the effects of social influence on optimal inventory decisions. This paper contributes to the above area of research by studying the implications of social learning for a relatively new type of operational function of social media platforms, namely, that of conducting inspections to verify the validity of the content they host. In terms of operational context, our work also contributes to the emerging research that focuses on the operations of online service platforms. Chakraborty and Swinney (2017) examine how an entrepreneur operating in a crowdfunding platform can signal the quality of her product to potential contributors by choosing an investment level and funding target; Marinesi et al. (2017) investigate the operational advantages of threshold-discounting offers; Gao and Su (2016) consider whether a retailer benefits by providing consumers the option to buy online and pick up in store; Papanastasiou et al. (2017) study optimal information disclosure with the goal of promoting exploration in an online review platform.

The misinformation aspect of our work relates to a line of research that considers sender-receiver games and demonstrates that in the presence of reputational concerns senders may strategically misreport the information in their possession (e.g., Durbin and Iyer 2009, Gentzkow and Shapiro 2006, Morris 2001, Ottaviani and Sørensen 2006). Our paper differs from this work in that truthful and fake messages (in the form of news articles) are assumed to be generated by an exogenous process, and the focus is instead on the propagation and detection of fake messages once these have started to circulate in a social media platform. Finally, this paper is one of several recent studies, belonging to a range of disciplines and utilizing a range of methodologies, that focus on various aspects of the fake news phenomenon. Among such work, Vosoughi et al. (2018) conduct a statistical analysis of Twitter stories and demonstrate that fake news tends to diffuse farther and faster than truthful news; Allcott and Gentzkow (2017) use survey data to study the impact of fake news on the 2016 US election and find that fake news are unlikely to have played a major role in the election outcome; Candogan and Drakopoulos (2017) develop a theoretical model to study how a platform with private knowledge of its content accuracy can balance user engagement against misinformation by designing appropriate signaling mechanisms; Jun et al. (2017) report experimental results where subjects were found to be less likely to fact-check claims when the claims were evaluated in the presence of others; Qiu et al. (2017) use simulation to highlight the role of behavioral limitations (information overload and limited attention) in the prevalence of low-quality information in a social network.

3. A Model of Fake News

We now present our model of fake news, followed by an example that illustrates its main features. We assume that there is a binary state of the world $\theta \in \{T, C\}$, whose realization is of interest to a society of agents. While the state realization itself is unobservable at all times, the agents may receive information pertaining to the state realization in the form of a news article.

The contents of an article are fully described by the observable pair (m, a), where $m \in \{\text{“}T\text{”}, \text{“}C\text{”}\}$ is the state realization that the information contained in the article supports, and $a \in (0.5, 1)$ is the persuasiveness of the information *when this is taken at face value*. More specifically, we assume that an article (m, a) is a binary signal that *claims* to have been generated according to the signal-generating process

$$P(m = \text{“}T\text{”} \mid T) = P(m = \text{“}C\text{”} \mid C) = a.$$

If the article is “truthful” (i.e., if its information is factual), the claimed signal-generating process is accurate, and the article is therefore informative with respect to θ.[1] By contrast, if the article is “fake” (i.e., if its information is fabricated), the claimed signal-generating process is inaccurate,

[1] Note that in the case of a truthful article, the signal-generating process is informative but noisy; that is, the generated signal is not always consistent with θ.

and the article in reality is completely uninformative with respective to θ. To capture this, we assume that the signal-generating process of a fake article is

$$P(m = \text{“}T\text{”} \mid T) = P(m = \text{“}T\text{”} \mid C) = \rho_T,$$

that is, independent of the state (see also Table 1).[2] For simplicity in exposition we set $\rho_T = 0.5$, although the analysis extends readily to any $\rho_T \in [0,1]$.

Truthful		
$P(m \mid \theta)$	$\theta = T$	$\theta = C$
$m = \text{“}T\text{”}$	a	$1-a$
$m = \text{“}C\text{”}$	$1-a$	a

Fake		
$P(m \mid \theta)$	$\theta = T$	$\theta = C$
$m = \text{“}T\text{”}$	ρ_T	ρ_T
$m = \text{“}C\text{”}$	$1-\rho_T$	$1-\rho_T$

Table 1 The signal-generating process of a truthful article (left) and a fake article (right).

3.1. Illustrative Example

To illustrate the main features of the described model, we present the following example.

Example 1. Suppose that because of an upcoming senate vote, the public is interested in whether the burning of fossil fuels is linked to climate change. Suppose further that an article is circulating on social media, whose contents can be summarized as follows:

“UC Berkeley study finds no connection between climate change and the burning of fossil fuels.”

In this example, the unobservable state of the world can be described as $\theta \in \{Y, N\}$, where Y represents “yes” (i.e., there is a link between the burning of fossil fuels and climate change) and N represents “no” (i.e., there is no link). Furthermore, the article may be described as $(m, a) = (\text{“}N\text{”}, 0.95)$, since the article's content supports $\theta = N$, and the information on which this support is based (taken at face value) is quite persuasive, in the sense that one would expect the outcome of a UC Berkeley study on this topic to be highly informative regarding the true state of the world (i.e., $a = P(m = \text{“}N\text{”} \mid \theta = N) = 0.95$).

To see how our model distinguishes between truthful and fake news, consider the following two extreme scenarios. Suppose first that it is commonly known that all articles circulating on social media are truthful. In this case, an agent receiving the article and whose prior belief is $b = P(\theta = N)$ will update her belief via Bayes' rule to b', where

$$b' = \frac{ab}{ab + (1-a)(1-b)} > b,$$

[2] Note that our model corresponds to cases where the information conveyed in a news article is either truthful or outright false. Another interesting case of misinformation is that of articles which have their basis in real facts, but construe these in a particular way; such articles do not fall into the category of fake news as considered in this paper.

so that her belief is updated towards $\theta = N$, consistent with the article's truthful report.

Now suppose instead that it is commonly known that all articles circulating on social media are fake. In this case, the receiving agent updates her belief via Bayes' rule to

$$b' = \frac{\rho_N b}{\rho_N b + \rho_N (1-b)} = b,$$

so that the fake article has no impact whatsoever on her belief about θ.

4. The Agents' News-Sharing Process

4.1. Process Description

In this section, we develop a sequential model of news propagation among a population of rational agents. We assume that there is an infinite number of agents, who take actions in sequence over a discrete-time horizon indexed by $t \in \{1, 2, ...\}$. At the beginning of time, each agent is endowed with a prior opinion over the unobservable state of the world $\theta \in \{T, C\}$, and a news article is generated which claims to be informative with respect to the true underlying state.

An agent's prior opinion is captured through a prior belief that the underlying state is T. Agents are heterogeneous in their opinions; in particular, we assume that each agent's opinion is an *iid* draw from a distribution with *cdf* $F(\cdot)$, where $F(\cdot)$ is continuous and strictly increasing on the positive unit interval, with $F(0) = 0$ and $F(1) = 1$.[3] We denote agent i's prior opinion by $b_{i0} := P_{i0}(\theta = T)$.

The article's contents are fully characterized by the observable pair (m, a), where $m \in \{\text{“}T\text{”}, \text{“}C\text{”}\}$ and $a \in (0.5, 1)$, and the article can be truthful or fake, as described in §3. We henceforth denote the article's validity by $v \in \{t, f\}$, where t (f) represents “truthful” (“fake”). We refer throughout to m and a as the article's “message” and “persuasiveness” respectively (the latter term reflects the fact that, conditional on being true, an article's influence on an agent's ex post opinion over θ increases with a). We assume that the ex ante probability that the news article is fake is $q_0 \in (0, 0.5)$, and that this probability is common knowledge; for instance, q_0 may be the proportion of fake news circulating online.[4] Without loss of generality, in our analysis we restrict attention to the case where the generated article carries the message $m = \text{“}T\text{”}$ (i.e., the article advocates in favor of $\theta = T$).

In the first period, the article is exogenously shared with a randomly chosen agent. The agent makes two decisions. First, whether to “inspect” (i.e., fact-check) the article to determine its validity. We assume that inspecting the article incurs a cost $K > 0$ for the agent, and that a decision

[3] For other models where agents are heterogeneous in their prior beliefs see, for example, Banerjee and Somanathan (2001) and Che and Kartik (2009); see also the discussion in Morris (1995).

[4] Note that our model assumes that the probability q_0 is independent of an article's persuasiveness a, although it is straightforward to extend the model to allow for dependance between the two.

to inspect perfectly reveals to the agent whether the article is fake.[5] Second, irrespective of whether the agent chose to inspect the article or not, the agent chooses whether to share the article or not. For an article of validity v, the agent's utility gain from a sharing action $\mathrm{s} \in \{0,1\}$ is given by $u^v(\mathrm{s})$, where $\mathrm{s}=1$ ($\mathrm{s}=0$) denotes sharing (not sharing); we assume that $u^t(1) > u^t(0)$ (i.e., for a truthful article, the dominant action is to share) and $u^f(1) < u^f(0)$ (i.e., for a fake article, the dominant action is to not share).[6] In addition, for ease of exposition in our analysis we set $u^t(1) = u^f(0) = 1$, $u^f(1) = u^t(0) = 0$, and, accordingly, we impose $K < 0.5$ to avoid trivial cases where agents never perform an inspection.

If a receiving agent chooses not to share the article, the article "dies" (i.e., sharing of the article is discontinued indefinitely), while if she chooses to share, the article is received by another agent in the next period, who repeats the decision process described above. Thus, in our analysis we assume that each agent is connected to, and can therefore share the article with, exactly one downstream agent; however, we note that our main results extend qualitatively to the case where each agent shares the article with more than one other agents.[7] Consistent with the motivating context of news sharing on social media, we assume that each agent receiving the article knows the number of times the article has been previously shared, but not how many times the article has undergone inspection by the preceding agents, or the private opinions of the preceding agents.[8]

4.2. Analysis

We now analyze the dynamics of the agents' news-sharing process, with a focus on properties of the process that are both of independent interest, but also crucial in solving the platform's inspection problem considered in §5. The main building block in the analysis of this section is the decision process of an individual agent that receives the article, which is where we begin our discussion.

Upon receiving the article, each agent must first choose whether to inspect it or not, and then whether to share it or not. Since inspection is assumed to yield a perfect outcome, if the agent

[5] Assuming that each agent's inspection cost is random and/or that the inspection outcome is imperfect has no qualitative bearing on our analysis.

[6] Note that this utility function assumes that agents are impartial and interested in disseminating (blocking) truthful (fake) articles; for instance, this may be motivated by altruistic or socially-responsible behaviour, or by agents' personal reputational concerns. Alternatively, we may assume that agents are partisan, so that they are interested in propagating articles that support their own beliefs, even if these articles are fake; in §7, we illustrate that such an assumption results in qualitatively similar agent behavior.

[7] Analysis of this case is available by the author.

[8] For example, the number of shares an article has received on Facebook is observable, as is the number of times an article has been retweeted on Twitter. Moreover, while one may argue that an agent knows the private opinion of the agent who shared the article with her, she is less likely to know the opinions of the agents further upstream in the sharing process.

chooses to inspect the item she is then guaranteed to take the "correct" sharing action (i.e., she will share the article if and only if she finds it to be truthful). By contrast, if she chooses not to inspect the item, the agent makes a sharing decision on the basis of her uncertain belief over the article's validity. Therefore, for an agent who receives the article in period t, we may restrict attention to the action set $\alpha_{it} \in [s, n, c]$, with action s representing *sharing* without inspecting, action n representing *not sharing* without inspecting, and action c representing inspecting (i.e., fact-*checking*) the article and subsequently deciding (correctly) whether to share it or not.

The action α_{it} which maximizes the agent's expected utility is determined by the agent's belief over the article's validity, conditional on receiving the article in period t. Thus, in order to characterize the period-t agent's actions, it is necessary to first characterize this belief. We start with the following crucial observation.

Lemma 1. *Suppose that an agent receives the article in period t. The agent's posterior belief that the article is fake, q_{it}, is strictly decreasing in her prior opinion over θ, b_{i0}.*

All proofs are provided in Appendix A. Recall that before receiving the article, all agents share the same prior belief q_0 that any given article is fake. However, upon receiving the article and reading its content, the validity beliefs of two agents with different prior opinions on the topic θ rationally diverge. In particular, the article advocates that $\theta = T$ (recall we have restricted attention to articles of type $m =$ "T"), while the prior opinion of an agent captures her prior belief that $\theta = T$; thus, Lemma 1 makes the important point that an agent whose prior opinion is more strongly aligned with the article's content will *rationally* perceive the article as less likely to be fake.

The diversity in the agents' assessments over the article's validity gives rise to diversity in their inspect-and-share actions. Using the monotonicity result of Lemma 1, we may characterize the period-t agent's actions as a function of her prior opinion b_{i0}. In particular,

Proposition 1. *In any period t, there exist thresholds $z_t^l \leq z_t^h$, where $z_t^l, z_t^h \in [0,1]$, such that the period-t agent's behavior is described as follows:*

(i) If $b_{i0} \leq z_t^l$, then $\alpha_{it} = n$; that is, the agent (a) does not inspect, and (b) does not share the article.

(ii) If $z_t^h < b_{i0} \leq z_t^h$, then $\alpha_{it} = c$; that is, the agent (a) inspects, and (b) shares the article only if it is found to be truthful.

(iii) If $b_{i0} > z_t^h$, then $\alpha_{it} = s$; that is, the agent (a) does not inspect, and (b) shares the article.

Figure 1 provides an example that illustrates the expected utility of the first-period agent from each action $\alpha_{it} \in [s, n, c]$, as a function of her prior opinion b_{i0} (along with the thresholds z_1^l and z_1^h). Since the agent's belief that the article is fake is strictly decreasing in her prior opinion (Lemma

1), the expected utility from sharing the article without first inspecting it (i.e., action $\alpha_{it} = s$) is strictly increasing in the prior opinion, while the expected utility from not sharing the article without first inspecting it (i.e., action $\alpha_{it} = n$) is strictly decreasing. By contrast, the expected utility of inspecting the article before making a sharing decision (i.e., action $\alpha_{it} = c$) is independent of the prior opinion. These observations give rise to the pattern described in Proposition 1: if the agent's prior opinion is strongly opposed to the article's content (i.e., $b_{i0} \le z_1^l$), the agent simply dismisses the article as fake (and chooses not to share it without bothering to inspect it); at the other extreme, if the agent's opinion is strongly aligned with the article's content (i.e., $b_{i0} \ge z_1^h$), then the agent accepts the article as truthful (and chooses to share it, again without bothering to inspect); by contrast, only agents with more moderate views ex ante are willing to incur the cost of inspecting the item before deciding whether or not to share it with the next agent.

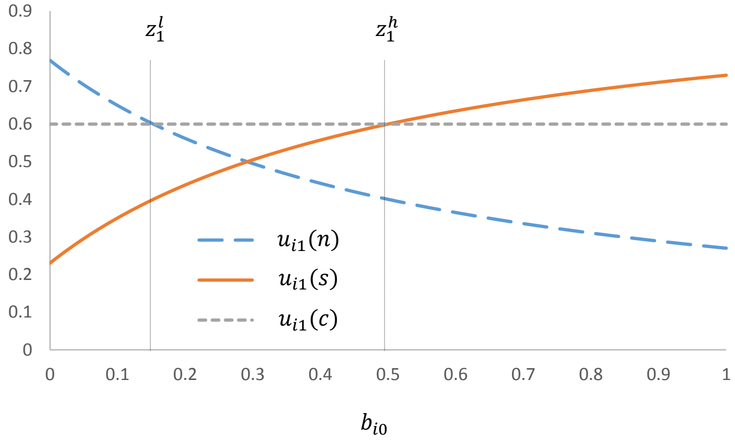

Figure 1 **The period-**1 **agent's expected utility from choosing action** $\alpha_{i1} \in \{s, n, c\}$**.**
Parameter values: $q_0 = K = 0.4$, $a = 0.9$.

We consider next the intertemporal dynamics of the agents' news-sharing process. To understand how the agents' behavior depends on the time at which they receive the news article, we first make the following observation with regards to the period-t agent's posterior belief q_{it}.

Lemma 2. *Suppose that an agent of prior opinion b_{i0} receives the article in period t. The agent's posterior belief that the article is fake, q_{it}, is weakly decreasing in time, t.*

If two agents with identical prior opinions receive the same article at two different times, the agent who receives it at the later time attaches lower probability to the article being fake. Rather than the timing per se, the key here is the fact that for an article to be received by the period-t agent, this means that the article has been shared $t - 1$ times in the past; and the more the article has been shared, the less likely it is that its content is fabricated. The mechanism through which this conclusion arises is based on the diversity and unobservability of the previous agents' inspection actions: the later an agent receives the article, the more likely it is that a false article would have been already detected through inspection and discontinued.

Using Lemma 2, it is straightforward to deduce that the thresholds z_t^l and z_t^h described in Proposition 1 are nonincreasing in time. In turn, this suggests that the sequential nature of the agents' news-sharing process gives rise to a phenomenon reminiscent of the "informational cascade" first identified by Bikhchandani et al. (1992) in their seminal herding paper.

Definition 1 (Sharing Cascade). We say that a sharing cascade is triggered at time T_c if in all periods $t \geq T_c$ the agents share the article without first inspecting it (i.e., $\alpha_{it} = s, \ \forall t \geq T_c$).

Proposition 2. *Suppose $z_1^l, z_1^h \in (0,1)$. There exists some finite T_c such that if a news article is shared in every period $t < T_c$, a sharing cascade is triggered in period T_c.*

Proposition 2 states that if an article survives for a critical number of periods T_c, it is subsequently shared by the agents in perpetuity (i.e., the article "goes viral"). The proof of the result relies on Lemma 1 to show that the threshold z_t^h decreases *strictly* over time until the entire support of agents' prior opinions finds it optimal to share the article without first inspecting it. From that point onwards, the propagation process enters a sharing cascade: the belief q_{it} remains constant over time, and each agent mimics the (sharing) action of her predecessor. Importantly, the critical number of shares T_c is a deterministic function of q_0, a, K, and can be calculated upfront for any parameter combination (see Proposition 3 and equation (4)). Finally, we note that by Proposition 1, if $z_1^h = 0$ the sharing process enters a cascade form the onset of the sharing process and we set $T_c = 1$, while if $z_1^l = 1$ the sharing process is discontinued form the onset of the sharing process.

In the sections that follow, we leverage the structure provided by Proposition 2 to study the inspection problem faced by a social media platform (§5), and to investigate the impact of fake news on the agents opinions over the state of the world θ (§6).

5. The Platform's Inspection Problem

5.1. Problem Description

We now imagine that the sharing of the news article between the agents is facilitated by a social-media platform (such as Facebook or Twitter). We assume that while the platform benefits from the sharing of content between its users (e.g., by collecting advertisement revenues), it is also interested in ensuring that the content being shared is truthful (e.g., to avoid loss of goodwill), and might therefore be willing to conduct inspection of an article itself in order to disclose the findings to its users (e.g., Financial Times 2017). Onto the propagation model described in §4.1, we superimpose the platform's decision-making process as follows.

In each period t, the platform moves before the period-t agent and decides whether to conduct an inspection of the article at a cost K_p. As before, we assume that inspection yields a perfect outcome, and the finding is immediately disclosed to the agents so that the article is either discontinued (if it is fake) or shared in perpetuity (if it is truthful). By contrast, if the platform chooses not to inspect the article, the period-t agent chooses whether to inspect and share the article, as described in the analysis of §4.2. If the article is shared by the period-t agent, the described process is repeated in the next period. In choosing an inspection policy, we assume that the platform (i) collects a reward r if an article is shared in period t (i.e., irrespective of whether the article is true), but also (ii) incurs a penalty p if the article being shared is fake. To exclude cases where the platform would never conduct an inspection, we further assume that the penalty incurred from the sharing of a fake news article is higher than the corresponding reward, $p > r$. The goal of the platform is to maximize its total expected discounted rewards over the infinite horizon, applying a discount factor of $\delta \in (0,1)$.

We assume that the platform is agnostic over θ, and adopts the agents' average prior opinion as its own, that is, $b_{p0} = \int_0^1 b_{i0} dF(b_{i0})$ (we note that this is not necessary for our analysis, provided b_{p0} is common knowledge). If the article survives up to period t, the platform's Bayesian belief that the article is fake is denoted by q_{pt}. We use $x_t \in X = \mathbb{Z}^+ \cup \{D\}$ to denote the state of the system, where $x_t = t \in \mathbb{Z}^+$ implies that the article has survived up until period t, while $x_t = D$ implies that either (i) sharing of the article has been discontinued by the agents, or (ii) the article has been inspected by the platform and its validity disclosed. The initial system state is $x_1 = 1$, and transitions between states are dictated by the platform's inspection policy and the agents' actions. An inspection policy is defined as $\pi : X \mapsto \{0,1\}$, with zero (one) denoting non-inspection (inspection); we use throughout the shorthand $\pi_{x_t} := \pi(x_t)$.

If the platform conducts an inspection in state $x_t = t$ (i.e., if $\pi_t = 1$), the state of the system in the next period is $x_{t+1} = D$. If the platform does not conduct an inspection in state $x_t = t$ (i.e., if $\pi_t = 0$), the system evolves on the basis of the period-t agent's action as described in §4.2. Let N_t, S_t, and C_t be the probabilities that the period-t agent takes actions n, s, and c respectively. These probabilities can be extracted from the agents' news-sharing process via the following result, which relies on calculating the thresholds z_t^l and z_t^h described in Proposition 1 recursively.

Proposition 3. *The probabilities N_t, S_t, C_t are given by the equations*

$$\begin{aligned}
N_t &= F\left(\frac{1}{2a-1}\left[\frac{\frac{1}{2}q_0 w_t K}{(1-q_0)(1-K)} - (1-a)\right]\right),\\
S_t &= 1 - F\left(\frac{1}{2a-1}\left[\frac{\frac{1}{2}q_0 w_t (1-K)}{(1-q_0)K} - (1-a)\right]\right),\\
C_t &= 1 - N_t - S_t,
\end{aligned}$$

where $w_1 = 1$ and $w_t \in [0,1]$ for $t \geq 2$ are calculated recursively via

$$w_t = \prod_{j=1}^{t-1} \frac{S_j}{S_j + C_j}.^{9} \tag{1}$$

Note that, consistent with the analysis of §4.2, the probabilities S_t (N_t) are non-decreasing (non-increasing) in time. Using the above probabilities and setting $N_D = 1$, $S_D = C_D = 0$, transitions between states $x_t \in X$ (assuming the platform does not conduct an inspection) can then be described as follows:

$$x_{t+1} = \begin{cases} t+1 & \text{w.p. } S_{x_t} + C_{x_t}(1 - q_{px_t}) \text{ (the article is shared by the period-}t\text{ agent)}, \\ D & \text{w.p. } N_{x_t} + C_{x_t} q_{px_t} \quad \text{(the article is not shared by the period-}t\text{ agent)}. \end{cases} \tag{2}$$

The platform's goal is to choose a policy π that maximizes the total expected discounted reward

$$E\left[\sum_{t=1}^{\infty} \delta^{t-1} g(x_t, \pi_{x_t})\right],$$

where the reward function $g(x_t, \pi_{x_t})$ is specified as

$$g(x_t, 0) = \begin{cases} r - p, & \text{w.p. } S_{x_t} q_{px_t}, \\ r, & \text{w.p. } (S_{x_t} + C_{x_t})(1 - q_{px_t}), \\ 0, & \text{w.p. } N_{x_t} + C_{x_t} q_{px_t}, \end{cases} \quad \text{and} \quad g(x_t, 1) = \begin{cases} -K_p, & \text{w.p. } q_{px_t}, \\ -K_p + \frac{r}{1-\delta}, & \text{w.p. } 1 - q_{px_t}, \end{cases} \tag{3}$$

for any $x_t \in X \setminus \{D\}$, and $g(D,0) = g(D,1) = 0$. In words, if the platform does not perform an inspection in state $x_t = t$, then it collects a reward $r - p < 0$ in the event that a fake article is shared by the period-t user; a reward of r in the event that a truthful article is shared; and zero in the event that the article is not shared. According to the platform's current belief in state x_t, q_{px_t}, these events occur with probability $S_{x_t} q_{px_t}$, $(S_{x_t} + C_{x_t})(1 - q_{px_t})$, and $N_{x_t} + C_{x_t} q_{px_t}$, respectively. On the other hand, if the platform performs an inspection in state $x_t = t$, then it incurs the inspection cost K_p and discloses its findings to the agents; if the article is found to be fake no further rewards are collected, while if it is found to be truthful, the platform's announcement triggers a sharing cascade resulting in a total discounted reward of $\frac{r}{1-\delta}$. According to the platform's belief q_{px_t}, the two events occur with probability q_{px_t} and $1 - q_{px_t}$, respectively.

5.2. Analysis

In this section, we derive the platform's optimal inspection policy as the solution to a dynamic program and analyze its main properties. In every period t, the platform chooses between performing its own inspection at cost K_p and allowing the agents' news-sharing process to continue

[9] Note that if $N_1 = 1$, the sharing process is trivially discontinued from the first period.

“organically” through the period-t agent’s inspect-and-share decision. Under a choice for the latter, if the agent does not share the article, the process is terminated and the platform collects no further rewards. If the agent shares the article, the platform collects the corresponding instantaneous reward and the process is repeated in the next period.

We begin our analysis by exploiting the structure of the agents’ news-sharing process in the absence of any platform intervention, and in particular the result of Proposition 2 pertaining to the occurrence of a sharing cascade in period T_c; note that the timing of a sharing cascade can be calculated via Proposition 3 as

$$T_c = \min\{t : S_t = 1\}. \tag{4}$$

Recall that once the agents’ news-sharing process enters a sharing cascade, all agents elect to share the article without any inspection in all subsequent periods. Assuming the platform has not taken action up until the period in which a cascade is triggered, Lemma 3 below suggests that it then makes an immediate once-and-for-all decision.

Lemma 3. *Suppose the platform has not conducted an inspection in periods $t < T_c$. If $q_{pT_c} > \frac{K_p(1-\delta)}{p-r}$, the platform inspects in period $t = T_c$. Otherwise, the platform never inspects the article.*

In a sharing cascade, the platform can no longer rely on the agents to perform an inspection that debunks a fake news article in some future period, nor can it learn anything further regarding the validity of the article by delaying its own inspection and observing the agents’ behavior. Therefore, in period T_c the platform uses its current validity belief q_{pT_c} to perform a calculation of the total expected reward associated with allowing the article to be shared in perpetuity without inspecting it. When the platform’s belief that the article is fake q_{pT_c} is sufficiently high, the platform conducts an inspection to make sure that the article is truthful before it is shared any further; in the opposite case, the platform allows the article to go viral without inspection.

Lemma 3 is particularly useful in solving the platform’s inspection problem, in that it reduces an infinite-horizon problem to a T_c-period finite-horizon problem. Using Proposition 3, T_c can be calculated upfront for any combination of our model parameters; then, employing Lemma 3, it suffices to specify the platform’s optimal policy in states $x_t \in [1, T_c]$. The solution to the platform’s inspection problem is described in the following result.

Theorem 1. *Let $T_c = \min\{t : \ S_t = 1\}$ and $v_{T_c} = \max\{\frac{(1-q_{pT_c})r}{1-\delta} - K_p, \frac{r-q_{pT_c}p}{1-\delta}\}$. Let v_t for $t < T_c$ be calculated recursively via*

$$v_t = \max\{\frac{(1-q_{pt})r}{1-\delta} - K_p, C_t(1-q_{pt})r + S_t(r - q_{pt}p) + [S_t + C_t(1-q_{pt})]\delta v_{t+1}\}, \tag{5}$$

where S_t and C_t are given in Proposition 3. Define $\tau^* = \{\min t : t \leq T_c, v_t = \frac{(1-q_{pt})r}{1-\delta} - K_p\}$. The optimal inspection policy is described as follows:

(i) If $\tau^* = \emptyset$, then the platform never inspects the article.

(ii) If $\tau^* \neq \emptyset$, then the platform inspects the article in period τ^*.

We henceforth adopt the convention $\tau^* = \infty$ for cases where the platform never conducts an inspection. Theorem 1 relies on the properties of the agents' news-sharing process as described by Propositions 2 and 3, and on the reduction of the platform's problem via Lemma 3. In any period t, the platform faces a choice between conducting an inspection which costs K_p and triggers a sharing cascade in the event that the article is found to be truthful, and allowing the period-t agent's action to dictate its current and future rewards. In turn, the actions of the period-t agent are characterized in Proposition 3, and the current expected reward can then be calculated via (3). The optimal inspection policy consists of the first time τ^* at which inspection is the platform's preferred option.

Obtaining the solution to the platform's inspection problem is easy computationally. However, the platform's optimal policy does not appear, at first glance, to exhibit much in the way of straightforward structure; to illustrate, we present the examples of Figure 2, which is typical of our numerical observations (note that shaded regions correspond to no inspection, i.e., $\tau^* = \infty$). The key in understanding the behavior of the optimal policy lies in understanding the interaction between the platform's objectives and the dynamics of the agents' news-sharing process. In order to explain the patterns observed in Figure 2, in the rest of this section we explore this interaction in detail.

We first consider cases where the agents' news-sharing process, in the absence of any platform intervention, is highly susceptible to the propagation of fake news. These are described in the following proposition.

Proposition 4. *Define*

$$Q^{al} = \frac{2(1-a)K}{2(1-a)K + (1-K)} \quad \text{and} \quad Q^p = \frac{2\gamma K_p(1-\delta)}{(p-r) + (2\gamma - 1)K_p(1-\delta)},$$

where $\gamma = ab_{p0} + (1-a)(1-b_{p0})$. Suppose $q_0 < Q^{al}$. Then:

(a) If $q_0 < Q^p$, the platform never performs an inspection.

(b) If $q_0 \geq Q^p$, the platform performs an inspection in the first period.

Proposition 4 characterizes the platform's optimal policy when $q_0 < Q^{al}$, where no agent is willing to incur the cost of conducting an inspection before sharing the article, because the probability that the article is fake is relatively low and/or the article's content is relatively mild. When this is the case, the article triggers a sharing cascade early on in the agents' sharing process, irrespective of its validity; indeed, using Proposition 3, we observe that the condition $q_0 < Q^{al}$ results in $S_1 = 1$. Since the agents are unwilling to inspect the article at any time, the platform is forced to decide in

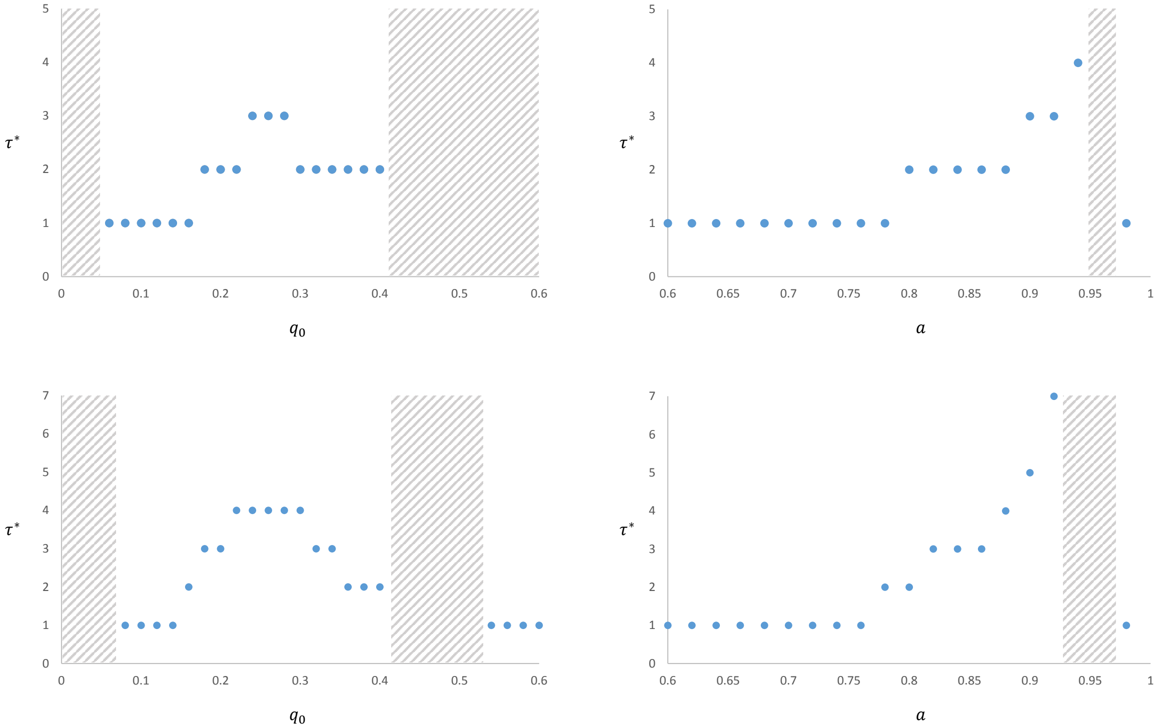

Figure 2 **Optimal inspection time τ^* as a function of the ex ante probability that the article is fake q_0 (left) and the article's persuasiveness a (right), for two different values of the platform's content-sharing reward, $r = 0.3$ (upper) and $r = 0.5$ (lower). Shaded regions denote no inspection ($\tau^* = \infty$). Parameter values: $K = 0.3$, $q_0 = 0.2$, $a = 0.85$, $K_p = 3$, $\delta = 0.99$, $p = 1$; F standard uniform.**

the first period whether or not to conduct an inspection. As Proposition 4 suggests, this decision depends on how q_0 compares to the threshold Q^p, which itself is a function of the platform's reward from content sharing relative to the penalty incurred from the sharing of a fake article: when q_0 is sufficiently low, the probability that the article is fake does not warrant the platform's inspection expenditure K_p. Proposition 4 is illustrated in the plots of Figure 2: in the left-hand side plots, the condition $q_0 < Q^{al}$ covers cases of $q_0 < 0.11$, while the condition $q_0 < Q^p$ covers cases of $q_0 < 0.04$ ($q_0 < 0.06$) in the upper plot where $r = 0.3$ (in the lower plot where $r = 0.5$); in the right-hand side plots, the condition $q_0 < Q^{al}$ covers cases of $a < 0.71$, none of which satisfy the condition $q_0 < Q^p$.

We consider next cases at the other extreme, where the agents' news-sharing process is highly robust to the propagation of fake news. These are characterized in the following result.

Proposition 5. *Define*

$$Q^{ah} = \frac{2aK}{2aK + (1-K)} \quad \textit{and} \quad N^p = \frac{K_p(1-\delta)[q_0 + 2\gamma(1-q_0)]}{2\gamma r(1-q_0)}.$$

Suppose $q_0 \geq Q^{ah}$. *Then:*

(a) If $N_1 < N^p$, the platform never performs an inspection.

(b) If $N_1 \geq N^p$, the platform performs an inspection in the first period.

Proposition 5 describes the platform's optimal policy when $q_0 \geq Q^{ah}$, where no agent is willing to share the article without first inspecting it, because the ex ante probability that the article is fake is relatively high, and/or because its content is not strong enough to generate "blind" support from agents whose opinions are ex ante aligned with its message; indeed, using Proposition 3, we observe that the condition $q_0 \geq Q^{al}$ results in $S_1 = 0$. It follows that in these cases, a sharing cascade can be triggered by the agents' news-sharing process only if the article is truthful. Interestingly, however, although a fake news article can never survive the agents' scrutiny, we observe that the platform may nevertheless opt to conduct an early inspection. In particular, the condition $N_1 > N^p$ captures cases where there is a high probability that an article will be rejected early on by the agents, even though its content is truthful. From the platform's perspective, such instances are costly, since the rewards collected from the sharing of a potentially truthful article are lost. To avoid this, the platform opts to conduct an early inspection, which triggers a sharing cascade in the event that the inspection finds the article to be truthful. The result of Proposition 5 is illustrated in the lower left-hand-side plot of Figure 2, where the condition $q_0 \geq Q^{ah}$ covers cases of $q_0 \geq 0.42$, while the condition $N_1 \geq N^p$ covers cases of $q_0 \geq 0.53$.

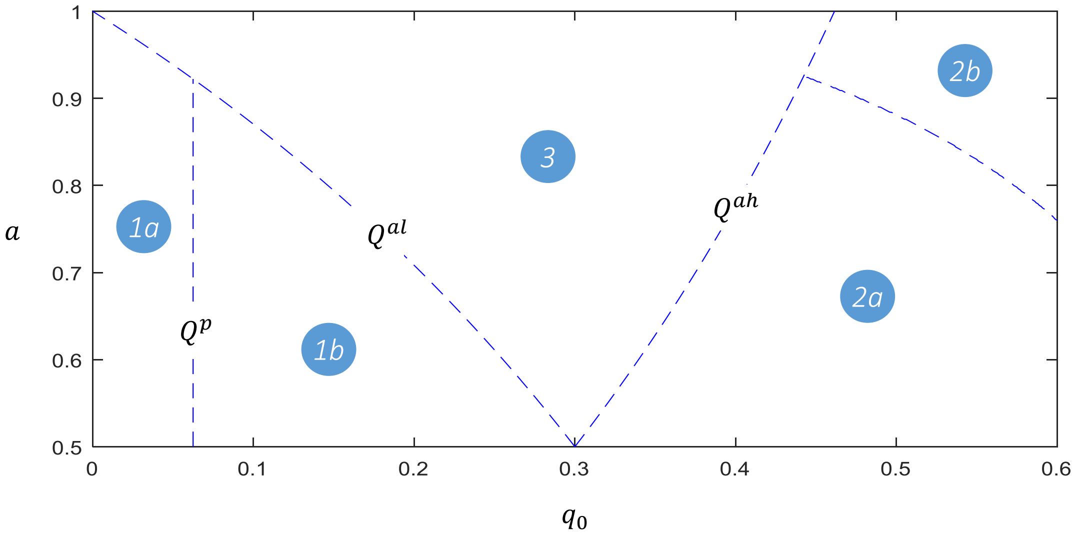

Figure 3 **Region plot of the platform's optimal inspection policy. Region 1a: No inspection (Proposition 4(i)); Region 1b: Inspection at $t = 1$ (Proposition 4(ii)); Region 2a: No inspection (Proposition 5(i)); Region 2b: Inspection at $t = 1$ (Proposition 5(ii)); Region 3: Inspection at $t = \tau^*$ (Theorem 1, Proposition 6). Parameters: $K = 0.3$, $K_p = 3$, $\delta = 0.99$, $p = 1$, $r = 0.5$; F standard uniform.**

Propositions 4 and 5 characterize the solution to the platform's inspection problem for combinations of our model parameters that satisfy $q_0 < Q^{al}$ and $q_0 \geq Q^{ah}$; the region plot of Figure 3 provides a visual representation of these results. We next consider the remaining parameter combinations, which satisfy $Q^{al} \leq q_0 < Q^{ah}$, and where the platform's optimal policy prescribes a "delayed inspection." The following result describes how the platform's policy in this parameter region depends on its reward structure.

PROPOSITION 6. *The optimal inspection time τ^* is nonincreasing in the penalty incurred from the sharing of fake news p and nondecreasing in the platform's inspection cost K_p.*

Proposition 6 verifies the intuition that (i) the higher the platform's penalty from the sharing of fake news, the earlier the platform opts to conduct an inspection, and (ii) the higher the platform's inspection cost, the more it prefers to delay its inspection. To draw additional insights into the platform's policy, and in particular on how this depends on the article's characteristics, we leverage the fact that cases that fall into the region $Q^{al} \leq q_0 < Q^{ah}$ can be viewed as interpolations between those characterized in Propositions 4 and 5.

Consider first the relationship between τ^* and q_0, which we observe is typically nonmonotonic (e.g., see Figure 2). As q_0 increases, the nature of the scrutiny faced by an article gradually shifts from long-run with low-intensity (at low q_0) to short-run with high-intensity (at high q_0). In the left-hand-side plots of Figure 2, when q_0 is relatively low the article faces very little scrutiny by the agents throughout the sharing process; in this case, a fake news article is unlikely to be detected by the agents, and at the same time the platform's ability to learn by observing the agents' sharing behavior is limited. Therefore, the platform prefers to make an inspection decision early on. On the other hand, when q_0 is relatively high, the article faces intense scrutiny by the agents in the early periods of the sharing process. However, because of its high intensity, this scrutiny quickly disappears: if a news article manages to survive early on, the agents' belief that the article is fake drops significantly, triggering an early sharing cascade. Here, following the short period of initial scrutiny by the agents, the platform is forced to decide early on whether to inspect the article. By contrast, at intermediate values of q_0, the article faces a relatively sustained level of moderate scrutiny which allows the platform to defer fact-checking to the agents for a greater period of time before intervening to conduct its own inspection.

Consider next the relationship between τ^* and the article's content a. From the platform's perspective, the important feature of the agents' news-sharing process here is that articles whose claimed persuasiveness is higher generate a more sustained level of opposition from agents whose prior opinions are opposed to the article's claims; thus, the higher the persuasiveness a is, the less likely a fake article is to survive long enough to trigger a sharing cascade. In the right-hand-side plots of Figure 2 we observe that as a increases, the platform tends to perform its inspection at a relatively later time and may eventually completely delegate inspection to the agents. However, observe that when the value of a is very high, the platform reclaims control of the inspection process and performs an early inspection. The rationale here is similar to that underlying Proposition 5(ii): when the opposition to an article becomes too high, the platform prefers to perform an early inspection to avoid losing the rewards associated with the sharing of a truthful article.

It is interesting to note that in cases where the platform's optimal policy prescribes a delayed inspection, we observe that this inspection typically occurs long before a cascade is triggered. For instance, while in the examples of Figure 2 the platform's delayed inspection occurs in periods $\tau^* \in \{2, ..., 7\}$, the time T_c at which a sharing cascade occurs is typically greater than 25. The reason is that although it can take a longer time for an article to trigger a cascade, the evolution of beliefs is slow after the initial periods of the sharing process. Therefore, by delaying inspection beyond these periods, the dynamics of the agents' sharing process do not change significantly, and the platform does not gain significant knowledge regarding the article's validity. However, at the same time the platform's penalties continue to accumulate in the event that the article is fake; this motivates the platform to take action well in advance of the cascade.

We conclude this section by considering how the agents' opinion distribution F affects the platform's optimal policy. Depending on the article's topic, agents may be more or less polarized in their ex ante opinions, and it is of interest to understand how this affects the platform's approach to conducting content inspections. In the example of Figure 4, we compare a setting where the agents are highly polarized and extreme in their opinions ($F \sim Beta(0.5, 0.5)$) against one where agents hold more moderate opinions ($F \sim Beta(2, 2)$). Observe that when the agents' opinions are more moderate, the platform conducts content inspections at an earlier time. To see why this is this case, Proposition 3 can be used to examine the dynamics of the agents' news-sharing process. Doing so reveals that when the agents are more polarized in their prior opinions, news articles face a more prolonged period of scrutiny relative to the case where the agents hold more moderate views (for instance, in the example of Figure 4 where $q_0 = 0.2$ and $a = 0.9$, the probability of inspection by the agents drops to around 10% after just two periods for $F \sim Beta(2, 2)$, while for $F \sim Beta(0.5, 0.5)$ the probability of inspection after two periods is around 30% and only drops to 10% after five periods). Given that an article faces scrutiny from the agents for a longer period of time when prior opinions are more polarized, the platform in such cases delays its own inspection more in order to save on the costs of inspecting fake articles that are likely to be detected by the agents through their own inspections.

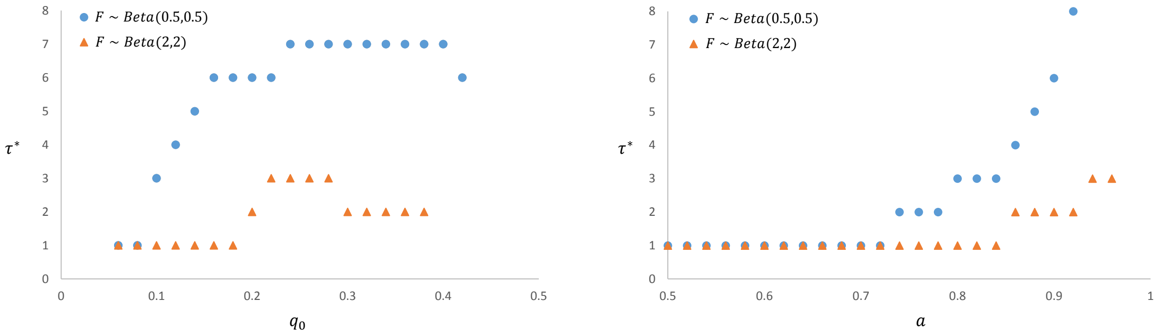

Figure 4 **Optimal inspection time τ^* as a function of the ex ante probability that the article is fake q_0 (left) and the article's persuasiveness a (right), for two different distributions F of the agents' ex ante opinions. Parameter values:** $K = 0.3$, $q_0 = 0.2$, $a = 0.9$, $K_p = 5$, $\delta = 0.99$, $p = 1$, $r = 0.1$.

6. The Impact of Fake News on Agent Opinions

The preceding analysis has focused on understanding the mechanics underlying the propagation of news articles in a society of rational agents (§4), and the properties of a social media platform's optimal inspection policy in the presence of fake news (§5). In this section, we turn our attention to the potential of fake news to sway the agents' opinions regarding the state of the world θ, which is arguably the main downside of the fake news phenomenon from a societal perspective. In doing so, we have in mind that the agents' opinions eventually translate into actions with surplus consequences (e.g., voting in a referendum). Thus, our investigation into the impact of fake news on the agents' ex post opinions can be viewed as an indirect investigation into the implications of fake news for the agents' surplus.

To conduct this investigation, let us define b_{it} as agent i's ex post opinion over the state of the world θ, after receiving the article in period t and choosing her inspect-and-share actions (if an agent does not receive the article, then $b_{it} = b_{i0}$, i.e., equal to her prior opinion). Recalling that the prior opinion b_{i0} captures the agent's ex ante belief that the state is $\theta = T$ and that our analysis focuses on articles whose message is $m =$ "T", we note first that

Lemma 4. *If agent i receives the article in period t, then her ex post opinion satisfies $b_{it} \geq b_{i0}$, where equality holds if and only if the agent conducts an inspection and finds the article to be fake.*

As long as the agent perceives a positive probability that the article's claims are truthful, her opinion is influenced in the direction supported by the article; it follows that the only way for this influence to be neutralized is if the agent conducts an inspection and concludes with certainty that the article is fake. Note that a direct consequence of Lemma 4 is that any article that survives long enough to trigger a cascade is then guaranteed to have a strictly positive influence on the opinions of all subsequent agents, that is, $b_{it} > b_{i0}$ for any $t \geq T_c$.

In the analysis that follows, we say that the *impact* of an article of type $v \in \{f, t\}$ is the overall shift it causes in agent opinions $b_{it} - b_{i0} \geq 0$, averaged across all agents in the population. According to this metric, it is useful to first establish a benchmark of "learning efficiency," which is described by the following two statements:

(i) The expected impact of a fake news article on the agents' opinions is $\psi^{f*} = 0$.

(ii) The expected impact of a truthful news article on the agents' opinions is

$$\psi^{t*} = \int_0^1 (b_i^{t*} - b_{i0}) dF(b_{i0}),$$

where b_i^{t*} is agent i's posterior opinion given knowledge that the article is truthful (i.e., $v = t$). In essence, the above benchmark describes the ideal learning environment where the validity of an article v is observable to the agents from the onset: if the article is fake, the agents simply discard it and the article has no impact whatsoever on the agents' opinions; if it is truthful, the article goes viral and each agent performs the appropriate Bayesian update to arrive at her posterior b_i^{t*}.

Let us consider first the case where there is no platform inspection. In this case, the impact of a news article can be calculated as follows.

Lemma 5. *In the absence of platform inspection, the expected impact of an article of validity $v \in \{f, t\}$ is given by*

$$\psi_n^v = \lambda_n^v \int_0^1 (b_{iT_c} - b_{i0}) dF(b_{i0}), \tag{6}$$

where λ_n^v is the ex ante probability that the article triggers a sharing cascade.

Lemma 5 suggests that the impact of an article is fully determined by the article's virality λ_n^v, the beliefs of the agent population once a cascade is triggered b_{iT_c}, and the distribution of prior agent opinions F. The result leverages the fact that an article either triggers a sharing cascade or is discontinued in finite time (Proposition 2): if the article is discontinued, then its average impact across all agents in the population is negligible; if it triggers a cascade, then the article's average impact is dominated by the shift in opinions it causes once a cascade is triggered. Our next result uses Lemma 5 to identify the ways in which the presence of fake news affects the agents' learning environment.

Proposition 7. *In the absence of platform inspection:*

(i) The expected impact of a fake article satisfies $\psi_n^f = \psi^{f*}$ if and only if

$$q_0 \geq \frac{2aK}{2aK + (1-K)}.$$

(ii) The expected impact of a truthful article satisfies $\psi_n^t = \psi^{t*}$ if and only if

$$\frac{2aK}{2aK + (1-K)} \leq q_0 < \frac{2(1-a)(1-K)}{2(1-a)(1-K) + K}.$$

Proposition 7 parses the detrimental impact of fake news into two distinct effects. The first is the *direct* effect, which refers to the propagation of fake news: if the agents fail to recognize an article as fake and the article goes viral, its content causes an erroneous shift in the agents' opinions. Since conditional on a fake article going viral we have $b_{it} = b_{iT_c} > b_{i0}$, the expected impact of a fake news article is zero (i.e., at the efficient level) if and only if it is impossible for the article to go viral without first being inspected by the agents; using Proposition 3, it can be deduced that this is the case only when q_0 is sufficiently high (i.e., $q_0 \geq \frac{2aK}{2aK+(1-K)}$).

The second is the *indirect* effect, which refers to the impact of fake news on the propagation of truthful articles: the presence of fake news in the news environment elevates the agents' suspicions about the validity of truthful articles (especially when these contain information that opposes their prior opinion), which in turn causes the impact of truthful articles to be diminished. In particular, observe from Lemma 5 that the impact of a truthful news article can be diminished for two reasons. First, the article may fail to go viral (i.e., $\lambda_n^t < 1$), which happens when q_0 is high enough so that the agents are highly suspicious of articles that contradict their own opinions, and are therefore prone to rejecting truthful articles; using Proposition 3, it can be shown that a truthful article is guaranteed to go viral only when $q_0 < \frac{2(1-a)(1-K)}{2(1-a)(1-K)+K}$. Second, even in environments where a truthful article is guaranteed to go viral, the article may fail to influence the agents' opinions to the extent warranted by the information it carries (i.e., $b_{iT_c} < b_i^{t*}$). To avoid this, the agents must be sure that an article could not have triggered a sharing cascade without first undergoing an inspection; using Proposition 3, this is ensured when $q_0 \geq \frac{2aK}{2aK+(1-K)}$.

Having identified the channels through which the presence of fake articles in the news environment affects the agents' opinions about the state of the world, we next consider whether and to what extent the platform's inspection policy is effective in safeguarding the agents from this impact. The following result captures the main structural properties of this effectiveness.

PROPOSITION 8. *In the presence of platform inspection, the expected impact of a news article of validity $v \in \{f, t\}$ is given by*

$$\psi^f = \begin{cases} \psi^{f*} & \text{if } \tau^* \leq T_c, \\ \psi_n^f & \text{otherwise,} \end{cases} \quad \text{and} \quad \psi^t = \begin{cases} \psi^{t*} & \text{if } \tau^* = 1, \\ \lambda^t \psi^{t*} & \text{if } 1 < \tau^* \leq T_c, \\ \psi_n^t & \text{otherwise,} \end{cases}$$

where $\lambda^t \geq \lambda_n^t$, and ψ_n^f, ψ_n^t are given in Lemma 5.

Proposition 8 highlights the asymmetric nature of the platform's inspection policy with respect to combating the direct and indirect effects of fake news. Consider first the direct effect, which is captured through the expected impact of a fake news article ψ^f. If the platform's inspection policy prescribes an inspection in any period up to and including period T_c, a fake news article cannot go viral and, as a result, the direct effect of fake news is largely mitigated—in our model, because the number of agents that receive the article before the platform conducts an inspection is relatively small, the article's impact across all agents becomes negligible. On the other hand, if the optimal policy does not prescribe inspection at any time, the magnitude of the direct effect of fake news is as characterized in Lemma 5, since this now depends exclusively on the dynamics of the agents' news-sharing process.

Consider next the indirect effect, which is captured through the expected impact of a truthful news article ψ^t. Proposition 8 suggests that the indirect effect of fake news is mitigated completely by the platform's policy only if the platform conducts an inspection from the onset of the sharing process. In these cases, the platform's inspection not only guarantees that the truthful article goes viral, it also reassures the agents that its content is valid, thus ensuring that the article's influence on the agents' opinions is at the efficient level b_i^{t*}. On the other hand, if the platform conducts a delayed inspection, the indirect effect of fake news is only partially mitigated. On one hand, assuming the truthful article has survived up to the period of the platform's inspection, the article is then guaranteed to go viral and its influence on the agents' beliefs is restored to the efficient level. On the other, however, the more the platform delays the inspection, the higher the risk that the article's sharing is prematurely discontinued by the agents. As in the case of the direct effect, if the platform opts not to conduct an inspection, then the magnitude of the indirect effect is determined by Lemma 5.

A natural next question is how the platform's effectiveness in safeguarding the agents' learning environment depends on the platform's incentives, captured in our model through the content-sharing reward r and the fake news penalty p. To investigate this question, we combine Propositions 7 and 8 with the analysis of §5. Proposition 9 below draws an important distinction between environments characterized by low and high prevalence of fake news.

Proposition 9. *For content-sharing reward r and fake news penalty p, define*

$$\mathcal{M}_l^{r,p} = \{(q_0, a) : q_0 \leq Q^{al}, \psi^f = \psi^{f*}, \psi^t = \psi^{t*}\} \quad \textit{and} \quad \mathcal{M}_h^{r,p} = \{(q_0, a) : q_0 \geq Q^{ah}, \psi^f = \psi^{f*}, \psi^t = \psi^{t*}\}.$$

For any $r_1 > r_2$, the following statements hold:

(i) $\mathcal{M}_l^{r_1,p} \subseteq \mathcal{M}_l^{r_2,p}$.

(ii) $\mathcal{M}_h^{r_1,p} \supseteq \mathcal{M}_h^{r_2,p}$.

For any $p_1 > p_2$, the following statements hold:

(i) $\mathcal{M}_l^{r,p_1} \supseteq \mathcal{M}_l^{r,p_2}$.

(ii) $\mathcal{M}_h^{r,p_1} = \mathcal{M}_h^{r,p_2}$.

The set $\mathcal{M}_l^{r,p}$ ($\mathcal{M}_h^{r,p}$) captures the cases of articles where the platform's inspection policy restores efficiency in the agent's learning process, in environments characterized by a low (high) prevalence of fake news. Observe that the set $\mathcal{M}_l^{r,p}$ expands (indicating an increase in the effectiveness of the platform's policy) when the content-sharing reward r decreases and when the penalty from the sharing of fake news p increases. By contrast, the set $\mathcal{M}_h^{r,p}$ expands when the reward r increases, and is unchanged when the penalty p increases. Proposition 9 therefore suggests that in environments where the proportion of fake news in circulation is low, lower rewards and/or higher penalties result in better alignment of the platform's incentives with those of a social planner who is interested in ensuring efficiency in the agents' learning process; by contrast, in environments where the proportion of fake news is high, penalties are ineffective, and the platform's inspection effectiveness increases when its rewards from content sharing are higher.

To explain why this occurs, it is instructive to first highlight the difference between environments of low and high fake news prevalence, with regards to the source of the inefficiency in the agents' learning process. When the amount of fake news in circulation is low, the agents conduct fewer inspections and are more likely to share both fake and truthful news articles. Thus, in this case the main source of inefficiency is the high virality of fake news articles. By contrast, when the amount of fake news in circulation is high, the agents conduct more inspections so that fake news articles are unlikely to go viral; however, the agents are also more likely to reject truthful articles without first inspecting them. Therefore, in this case the main source of inefficiency is the "non-virality" of truthful news. Consider next the platform's perspective. For the platform, the virality of fake news articles becomes more detrimental when the content-sharing reward r is lower and/or when the penalty from the sharing of fake news p is higher. On the other hand, the losses incurred from the non-virality of truthful are higher when the reward r is higher, and these losses are independent of the penalty p. Combining the above, it follows that when the prevalence of fake news is low, lower rewards and higher penalties provide an effective motivation for the platform to combat the virality of fake news, and in doing so to restore efficiency in the agents' learning process. By contrast, when the prevalence of fake news is high, higher rewards provide the appropriate motivation for the platform to ensure virality of truthful news articles, thereby restoring efficiency in the learning process, while penalties are redundant.

To illustrate, we present the example of Figure 5, where we plot the impact of a fake and a truthful article relative to an article with the same content but which is known to be true. Efficiency in the agents' learning process in this example is thus achieved when $\frac{\psi^f}{\psi^{t*}} = 0$ in the left-hand-side plot and at the same time $\frac{\psi^t}{\psi^{t*}} = 1$ in the right-hand-side plot. Consistent with the discussion above, observe that at relatively lower values of q_0, the platform's policy is more effective under the lower reward $r = 0.1$, as the lower reward motivates the platform to ensure the impact of fake news is limited; by contrast, at relatively higher values of q_0, the platform's policy is more effective under the higher reward $r = 0.4$, as the higher reward motivates the platform to ensure the impact of truthful articles is at the efficient level.

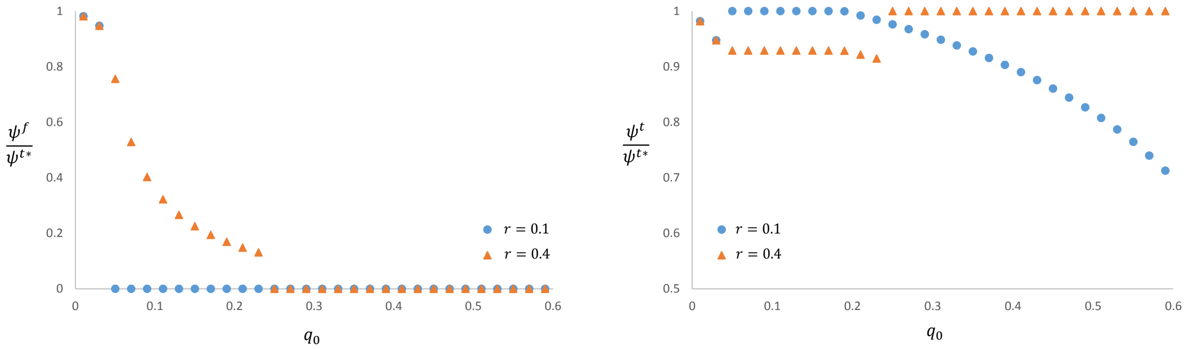

Figure 5 **Impact of a fake (left) and a truthful (right) article relative to that of an article known to be true, for two different values of the platform's content-sharing reward r. Parameter values: $K = 0.3$, $a = 0.95$, $K_p = 3$, $\delta = 0.99$, $p = 1$; F standard uniform.**

7. Model Extension: Partisan Sharing Behavior

We have assumed in our main analysis that agents are socially-responsible, to the extent that they would not knowingly share articles they know to be fake. We now consider an alternative mode of agent behavior, where the agents are partisan. In the context of our model, this implies that agents are willing to share an article as long as the article supports their own beliefs about the state of the world θ (i.e., irrespective of whether the article is true or not). In this setting, agents will inspect an article if and only if their sharing action depends on the outcome of the inspection; in turn, this implies that agents will inspect if and only if knowing that the article is truthful or fake has an impact on their assessment of which realization of θ is more likely.

Expanding on the analysis of §4, let us consider how agents take inspect-and-share actions in this case. It is straightforward to deduce that agents with a high prior belief b_{i0} who receive an article whose message is "T" will choose action $\alpha = s$, that is, they will choose (i) not to inspect the article, and (ii) to share the article. To see why, note that even if these agents were to discover through inspection that the article is fake, this would not change their sharing action, since their posterior belief would still be consistent with the message conveyed by the article. At the other extreme, agents with low b_{i0} will choose action $\alpha = n$, that is, they will choose (i) not to inspect the article, and (ii) not to share the article. This case describes agents whose beliefs would indeed increase if they were to find out that the article is true, but not to the extent that the article would change their mind over which state of the world is the more likely – thus, even if inspection resulted in learning that the article is true, they would nevertheless choose not to share it. In contrast, those agents' whose beliefs are intermediate would choose action $\alpha = c$, that is, they would (i) inspect the article, and (ii) share it only if they found it to be truthful; for these agents, whether or not sharing the article supports their own beliefs depends on whether the article is truthful or not – thus, as along as the inspection cost is not prohibitively high, they would opt to inspect the article before taking a sharing action.

It follows form the above discussion that while the incentives of the agents here are quite different from those in our main model, the agents' action profile with respect to the prior belief b_{i0} is qualitatively similar. Moreover, it can also be shown that the agents' beliefs that the article is fake q_{it} in this case will also decrease over time, so that the segment of agent opinions that lead to sharing the article without first inspecting it (not sharing the article without first inspecting it) is increasing (decreasing) in time, consistent with our main analysis. It then follows that the analysis of the platform's inspection problem and the impact of fake news on the agents' beliefs is expected to remain qualitatively similar to that in the main text.

8. Discussion

This paper represents an attempt at understanding the economic drivers underlying the propagation and detection of fake news articles when these are shared sequentially from agent to agent in a social media platform. In this section, we provide a brief discussion of our main results along with some of their potential implications.

We have shown in §4 that in the absence of any platform intervention, fake news articles are likely to proliferate through sequential sharing even when agents are intent on sharing only truthful articles. In particular, our model suggests that agents are rationally more likely to share an article the more it has been previously shared by their peers, resulting in fake news articles going viral after receiving a sufficiently high number of shares. Based on this observation, one potential platform intervention that has not been explored explicitly in this paper is that of concealing an article's past sharing statistics. By carefully choosing how to disclose past sharing information, the platform might be successful in inducing an increased level of alertness by the agents, thereby limiting the virality of potentially fake news articles. Furthermore, our model also suggests that agents of different prior opinions are likely to differ in their inspect-and-share actions in predictable ways. This then implies that the platform could use the past sharing behavior of individual agents in order to improve its inference about the likely validity of articles being circulated in the platform (e.g., by assigning appropriate "weights" to the sharing actions of agents whose past behavior suggests particular ideological preferences).

With respect to the optimal platform inspection policy, we have shown in §5 that this is non-monotone in the ex ante probability that an article is fake. Specifically, our analysis suggests that when this probability is very low, the platform should not conduct an inspection; when it is moderately low, the platform should conduct an inspection from the onset of the sharing process; when it is moderate, the platform should conduct a delayed inspection; and when it is high, the platform should either conduct an inspection from the onset or no inspection at all. Given the massive amount of content in circulation at any time, developing an understanding of when the platform can defer fact-checking to the agents either partially or completely can result in significant gains both in terms of fact-checking costs and in terms of efficiency in combating fake news. One interesting byproduct of our analysis of the platform's optimal policy is that revenues can be significantly decreased not only through the propagation of fake news articles, but also through the detrimental effect of fake news on the sharing of truthful articles, which are erroneously rejected by the agents due to suspicions that their content is fake. While much attention in practice is currently attached to "flagging" articles that contain fake content, this observation suggests that the platform might benefit from actively disclosing positive inspection outcomes as well; in this way, agents may be more receptive to information that opposes their own beliefs, rather than rejecting such information as fake.

We have also considered in §6 how the platform's content-sharing incentives affect the effectiveness of its inspection policy in safeguarding the agents' beliefs from the detrimental impact of fake news. We have shown that the incentives that lead to the platform's policy being more effective depend to a great extent on the characteristics of the news environment. Notably, when the prevalence of fake news is low (high), our analysis suggests that the platform's policy is more effective when its rewards from content sharing are low relative to the penalties incurred from the propagation of fake news (when the rewards from content sharing are high in absolute terms). This result suggests that penalties are not always effective in motivating the platform to safeguard its users, especially in cases where high volumes of fake news articles are in circulation. Moreover, the observation that high absolute rewards from content sharing are more effective in such cases may suggest that a platform acting with monopoly power can be beneficial from the standpoint of combating the fake news phenomenon.

In closing, we point out that one significant aspect of the fake news phenomenon that is not explored in our model is the process by which fake news is generated. In particular, in our paper fake news is assumed to be generated exogenously and independently of the platform's inspection policy; however, fake news articles in reality are generated by economic agents who seek to maximize either the virality or the influence of these articles. Investigations into the strategic interaction between the platform and the agents on one side, and the source of fake news on the other, will likely lead to additional and/or more nuanced insights into optimal detection policies, as well as on the implications of these policies regarding the impact of fake news on the society.

Appendix

A. Proofs

Proof of Lemma 1

Note first that beliefs in our model are defined over the set $G := \{(v,\theta)\}_{v\in\{t,f\},\theta\in\{T,C\}}$. According to our model specification, agent i's prior belief is given by

$$P_{i0}(f,T) = q_0 b_{i0},\quad P_{i0}(f,C) = q_0(1-b_{i0}),\quad P_{i0}(t,T) = (1-q_0)b_{i0},\quad P_{i0}(t,C) = (1-q_0)(1-b_{i0}). \tag{7}$$

Suppose an agent receives the article in period t. The agent uses two pieces of information to update her belief: (i) the article's content, $m =$ "T", and (ii) the fact that the article has been shared by all preceding agents; let $\mathcal{H}_t$ denote this event. For $g \in G$, the agent's posterior belief is

$$P_{it}(g \mid \mathcal{H}_t, \text{“}T\text{”}) = \frac{P(\mathcal{H}_t, \text{“}T\text{”} \mid g)P(g)}{\sum_{w\in G} P(\mathcal{H}_t, \text{“}T\text{”} \mid w)P(w)} = \frac{P(\mathcal{H}_t \mid \text{“}T\text{”}, g)P(\text{“}T\text{”} \mid g)P(g)}{\sum_{w\in G} P(\mathcal{H}_t \mid \text{“}T\text{”}, w)P(\text{“}T\text{”} \mid w)P(w)}.$$

Consider the probabilities in the last expression. $P(g)$ is the agent's prior belief given in (7). For $P(\text{“}T\text{”} \mid g)$, according to the signal structure described in §4.1, we have

$$P(\text{“}T\text{”} \mid (t,T)) = a,\ P(\text{“}T\text{”} \mid (t,C)) = 1-a,\ P(\text{“}T\text{”} \mid (f,C)) = P(\text{“}T\text{”} \mid (f,T)) = \rho_T = 0.5. \tag{8}$$

Next, let S_t, N_t, C_t denote the probability that the period-t agent chooses action s, n, c, respectively, conditional on receiving an article of type $m =$ "T" (i.e., $S_t = P(\alpha_{it} = s \mid \text{“}T\text{”}, \mathcal{H}_t)$, etc.). Note that if an agent inspects the item, she then shares it with the next agent if and only if she finds it to be truthful. Therefore,

$$P(\mathcal{H}_t \mid \text{“}T\text{”}, (f,T)) = P(\mathcal{H}_t \mid \text{“}T\text{”}, (f,C)) = \prod_{k=1}^{t-1} P(\alpha_{ik} = s \mid \text{“}T\text{”}, \mathcal{H}_k) = \prod_{k=1}^{t-1} S_k,$$
$$P(\mathcal{H}_t \mid \text{“}T\text{”}, (t,T)) = P(\mathcal{H}_t \mid \text{“}T\text{”}, (t,C)) = \prod_{k=1}^{t-1} \big(P(\alpha_{ik} = s \mid \text{“}T\text{”}, \mathcal{H}_k) + P(\alpha_{ik} = c \mid \text{“}T\text{”}, \mathcal{H}_k)\big) = \prod_{k=1}^{t-1} (S_k + C_k)$$

where we have used that (i) conditional on the message m, the agents' sharing actions depend only on v, and (ii) a fake (truthful) news article is shared only by agents who choose action s (is shared by agents who choose either action s or action c). Using the above probabilities, it can then be shown that

$$q_{it} = P(f \mid \mathcal{H}_t, \text{“}T\text{”}) = \frac{0.5q_0 w_t}{0.5q_0 w_t + [ab_{i0} + (1-a)(1-b_{i0})](1-q_0)}, \tag{9}$$

where we define $w_t := \frac{\prod_{k=1}^{k=t-1} S_k}{\prod_{k=1}^{k=t-1}(S_k + C_k)} = \prod_{k=1}^{k=t-1} \frac{S_k}{(S_k + C_k)}$ (note that for $t = 1$, where there are no previous-sharing observations, we set $w_t = 1$). Finally, note that since $a \in (0.5, 1)$, q_{it} is strictly decreasing in b_{i0}.

Proof of Proposition 1

Recall that q_{it} is the period-t agent's belief that the article is fake, and that we have normalized the agent's utility from sharing (not sharing) a truthful (fake) article to one, and to zero otherwise. The period-t agent chooses the action which maximizes her expected utility which, given her belief q_{it}, is described for each action $\alpha_{it} \in \{n, s, c\}$ as follows:

$$u_{it}(\alpha_{it}) = \begin{cases} q_{it} & \text{for } \alpha_{it} = n, \\ 1 - q_{it} & \text{for } \alpha_{it} = s, \\ 1 - K & \text{for } \alpha_{it} = c. \end{cases} \tag{10}$$

Using (10) and Lemma 1, we have that $u_{it}(n) = q_{it}$ is strictly decreasing in b_{i0}, which implies that $u_{it}(s) = 1 - q_{it}$ is strictly increasing in b_{i0}. Moreover, note also that $u_{it}(s) + u_{it}(n) = 1$, and that $u_{it}(c) = 1 - K > 0.5$ (by virtue of our assumption $K < 0.5$). Therefore, the period-t agent will choose action n if $q_{it} > 1 - K$; action s if $1 - q_{it} > 1 - K$; and action c if $1 - K \geq \max\{q_{it}, 1 - q_{it}\}$. Existence of the thresholds on b_{i0} stated in the result follows from the strict monotonicity of q_{it} in b_{i0}.

Proof of Lemma 2

From the proof of Lemma 1, we have $q_{it} = \frac{0.5q_0 w_t}{0.5q_0 w_t + [ab_{i0} + (1-a)(1-b_{i0})](1-q_0)}$, where $w_t = \prod_{k=1}^{k=t-1} \frac{S_k}{(S_k + C_k)}$. Consider the likelihood function

$$\frac{q_{it}}{1 - q_{it}} = \frac{0.5q_0 w_t}{[ab_{i0} + (1-a)(1-b_{i0})](1-q_0)}.$$

Note that the likelihood function is strictly increasing in w_t. Since the likelihood function is strictly increasing in q_{it}, this implies that q_{it} is also strictly increasing in w_t. Finally, note that since $S_t, C_t \in [0,1]$, we have that $w_t \in [0,1]$ and is nonincreasing in t, completing the claim.

Proof of Proposition 2

Let S_t, N_t, C_t denote the probability that the period-t agent chooses action s, n, c, respectively conditional on the article being of type $m =$ "T". The result makes use of the following lemma.

Lemma 6. *If $S_t, N_t > 0$, then $C_t > 0$.*

Proof. If $S_t, N_t > 0$, then by the period-t agent's expected utility given in (10), and the monotonicity of q_{it} in b_{i0} given in Lemma 1, there exists some $\bar{b} \in (0,1)$ such that $u_{it}(s; b_{i0} = \bar{b}) = u_{it}(n; b_{i0} = \bar{b}) = 0.5$ (i.e., an agent with $b_{i0} = \bar{b}$ is indifferent between actions s, n). However, note that $u_{it}(c) = 1 - K > 0.5$ for all b_{i0}, which implies that there exists a positive segment of agent opinions around $\bar{b}$ that prefer action c over actions s, n, so that $C_t > 0$.□

We now return to the proof of the proposition. Using the monotonicity of q_{it} in b_{i0} and expression (9) from Lemma 1, along with the agent's expected utility given in (10), it suffices to show that there exists some T_c such that an agent with $b_{i0} = 0$ has a utility from sharing without inspection $u_{iT_c}(s) > 1 - K$ which then implies that the period T_c-agent chooses action s with probability one. Once this occurs, every period is a repetition of the previous, and we say that a sharing cascade is triggered.

Suppose that in some period t an agent with $b_{i0} = 0$ chooses action n so that $N_t > 0$ while an agent with $b_{i0} = 1$ chooses action s (note that in period $t = 1$ this is ensured by $z_1^l, z_1^h \in (0,1)$). By Lemma 6, this means that $C_t > 0$. If the article is shared in period t, then since $w_t = \prod_{k=1}^{k=t-1} \frac{S_k}{(S_k + C_k)}$, we have $w_{t+1} < w_t$. This then implies that $q_{it+1} < q_{it}$, so that $S_{t+1} > S_t$ and $N_{t+1} < N_t$. As long as the article is being shared, the described process is repeated until some period $k > t$ where $N_k = 0$. If $N_k = 0$, we have that in period k an agent with $b_{i0} = 0$ chooses either action s or action c. If she chooses action s then this means that $N_k = C_k = 0$ and $S_k = 1$ so that a cascade has been triggered in period $T_c = k$. Alternatively, if the agent with $b_{i0} = 0$ chooses action c in period k, then this implies $C_k, S_k > 0$. If the article is shared further in period k, then we have $w_{k+1} < w_k$, $S_{k+1} > S_k$, $C_{k+1} < C_k$, $N_{k+1} = 0$, and the last process is repeated until some period $m > k$ where $S_m = 1$ and $C_m = N_m = 0$, at which time a cascade is triggered.

Proof of Proposition 3

Let N_t, S_t, C_t denote the probabilities that the period-t agent chooses action n, s, c respectively, conditional on the article's content $m =$ "T". Then, we have from Lemma 1 that

$$q_{it} = \frac{0.5 q_0 w_t}{0.5 q_0 w_t + (1 - q_0)[a b_{i0} + (1 - a)(1 - b_{i0})]}, \tag{11}$$

where w_t can be expressed as $w_t := \prod_{j=1}^{t-1} \frac{S_j}{S_j + C_j}$ with $w_1 = 1$. Starting in period $t = 1$ and for every subsequent period, we calculate first w_t and then the probabilities N_t, S_t, C_t in accordance with Proposition 1 via $N_t = P(q_{it} > 1 - K)$, $S_t = P(1 - q_{it} > 1 - K)$, $C_t = 1 - S_t - N_t$. Using (11), these probabilities can then be expressed in terms of the distribution of prior opinions F, as stated in the proposition. Note that in the case $N_1 = 1$ the sharing process is trivially discontinued from the first period.

Proof of Lemma 3

If a cascade occurs at time T, then $q_{pt} = q_{pT}$ for all $t \geq T$. If $(1 - q_{pT})(r + \delta r + \delta^2 r + \ldots) - K_p = \frac{r(1 - q_{pT})}{1 - \delta} - K_p \leq [q_{pT}(r - p) + (1 - q_{pT})(r)] + \delta[q_{pT}(r - p) + (1 - q_{pT})(r)] + \delta^2[q_{pT}(r - p) + (1 - q_{pT})(r)] + \ldots = \frac{r - q_{pT} p}{1 - \delta}$ then under the assumption that no inspection will occur in subsequent periods, it is suboptimal to inspect in period $t = T$. If $\frac{r(1 - q_{pT})}{1 - \delta} - K_p > \frac{r - q_{pT} p}{1 - \delta}$, or equivalently $K_p < \frac{(p - r) q_{pT}}{1 - \delta}$ then assuming no inspection occurs in subsequent periods, inspection is optimal at time T. It remains to establish that it cannot be optimal to delay inspection until some future period. Suppose that the optimal policy inspects at some time $t > T$. Then it must be that in period $t - 1$ inspection is suboptimal, which implies that $\frac{r(1 - q_{pT})}{1 - \delta} - K_p \leq (r - q_{pT} p) + \delta(\frac{r(1 - q_{pT})}{1 - \delta} - K_p)$ or, equivalently, that $K_p \geq \frac{(p - r) q_{pT}}{1 - \delta}$, which leads to a contradiction.

Proof of Theorem 1

The result relies on Propositions 2 and 3, and Lemma 3. We solve the firm's problem via dynamic programming. Note first that according to Lemma 3, the problem is reduced to a finite T_c-period-horizon problem, with a terminal value $v_{T_c} = \max\{\frac{(1 - q_{pT}) r}{1 - \delta} - K_p, \frac{r - q_{pT_c} p}{1 - \delta}\}$, where the max operator captures whether or not the platform finds it optimal to inspect in period T_c, as described in Lemma 3.

Next, in any period $t \in [1, T_c - 1]$, the platform has two options:

(i) Inspection: if it performs an inspection, the platform incurs the inspection cost K_p, and a sharing cascade is triggered via its announcement if and only if the article is truthful; since the latter occurs with probability $(1 - q_{pt})$ the total expected discounted reward from inspection is $(1 - q_{pt}) \frac{r}{1 - \delta} - K_p$.

(ii) No inspection: if the platform elects not to conduct an inspection, then it collects an instantaneous reward based on the article's validity and the period-t agent's action, and future rewards if the period-t agent shares the article. Specifically, if the period-t agent chooses action n, then the article is discontinued and no rewards are collected; if the agent chooses action c, then a reward r is collected in period t and the article is shared if and only if the article is truthful (this occurs with probability $(1-q_{pt})$); if the agent chooses action s, then a reward of r is collected if the article is truthful, a reward $r-p$ is collected if the article is fake, and the article is shared irrespective of its validity. The probabilities that the period-t agents chooses action n, c, s are given by Proposition 3 as N_t, C_t, S_t, respectively. Therefore, the total expected discounted reward from no inspection is

$$C_t(1-q_{pt})(r+\delta v_{t+1}) + S_t(1-q_{pt})(r+\delta v_{t+1}) + S_t q_{pt}(r-p+\delta v_{t+1}).$$

Rearranging this expression gives $C_t(1-q_{pt})r + S_t(r-q_{pt}p) + [S_t + C_t(1-q_{pt})]\delta v_{t+1}$.

To identify the optimal inspection time τ^*, we apply backward induction starting from period T_c to find the earliest period at which inspection becomes optimal for the platform. If such a period does not exist, it is optimal for the platform never to conduct an inspection.

Proof of Proposition 4

Observe first that from Proposition 3, we have that if $S_1 = 1$, then $S_t = 1$ for all t. Observe next that for any $q_0 < Q^{al}$, we have $S_1 = 1$, implying that a cascade is triggered from the onset of the agent's sharing process. Applying Lemma 3, we have that the platform in this parameter region makes a once-and-for-all inspection decision in the first period. Furthermore, we have also from Lemma 3 that the platform will perform an inspection in the first period if and only if

$$q_{p1} = \frac{0.5q_0}{0.5q_0 + \gamma(1-q_0)} \geq \frac{K_p(1-\delta)}{(p-r)}.$$

Rearranging, the above is equivalent to the condition $q_0 \geq Q^p$ stated in the result.

Proof of Proposition 5

Observe that using Proposition 3, for any $q_0 \geq Q^{ah}$, we have $S_1 = 0$. Moreover, observe also that if the article is shared in the first period, then $S_t = 1$ for all $t \geq 2$ since an article reaches the second period if and only if it is truthful. In particular, if the first-period agent takes action $\alpha = n$ (this occurs with probability N_1 given in Proposition 3) the article is discontinued even though it may be truthful, while if the first-period agent takes action $\alpha = c$ (this occurs with probability C_1 given in Proposition 3) the article triggers a sharing cascade in the second period if and only if the agent finds the article to be truthful and therefore decides to share it. It follows from the above that the platform must make an inspection decision in the first period, since delaying results in either the article being discontinued in the first period, or being revealed to be truthful (subsequently triggering a sharing cascade). If the platform inspects the article, its total expected reward is equal to $R_I = \frac{(1-q_{p1})r}{1-\delta} - K_p$. If it does not inspect the article, the total expected reward is $R_{I'} = (1-N_1)\frac{(1-q_{p1})r}{1-\delta}$. Thus, the platform performs an inspection in the first period if and only if $R_I \geq R_{I'}$ which is equivalent to the condition $N_1 \geq N^p$ stated in the main text.

Proof of Proposition 6

To prove that τ^* is nonincreasing in p, let R_t^I ($R_t^{I'}$) denote the platform's total expected reward if an inspection is conducted (is not conducted) in period t. From Theorem 1 we have $R_t^I = \frac{(1-q_{pt})r}{1-\delta} - K_p$ and $R_t^{I'} = C_t(1-q_{pt})r + S_t(r - q_{pt}p) + [S_t + C_t(1-q_{pt})]\delta v_{t+1}$. Notice that R_t^I is independent of p, so that to prove the claim it suffices to show that $R_t^{I'}$ is nonincreasing in p. In turn, a sufficient condition for this to hold is that v_{t+1} is nonincreasing in p. We prove by induction that v_t is nonincreasing in p for all $t \in [1, T_c]$. From Theorem 1 we have that $v_{T_c} = \max\{\frac{(1-q_{pT_c})r}{1-\delta} - K_p, \frac{r - q_{pT_c}p}{1-\delta}\}$, so that v_{T_c} is nonincreasing in p. We next show that if v_t is nonincreasing in p, then v_{t-1} is also nonincreasing in p. We have

$$v_{t-1} = \max\{\frac{(1-q_{pt-1})r}{1-\delta} - K_p, C_{t-1}(1-q_{pt-1})r + S_{t-1}(r - q_{pt-1}p) + [S_{t-1} + C_{t-1}(1-q_{pt-1})]\delta v_t\},$$

which is nonincreasing in p by the inductive hypothesis.

We next prove that τ^* is nondecreasing in K_p. Suppose first that for some $K_p = K_1$ the platform never conducts an inspection. Then we have from Theorem 1 that inspection is never optimal for any $K_p \geq K_1$. Suppose next that for $K_p = K_1$ it is optimal to inspect in period τ. We show that for any $K_p = K_2 < K_1$, if the platform has not conducted an inspection up to period τ, then it must be optimal to inspect in period τ. The statement holds trivially for $\tau = T_c$ by Lemma 3. For $\tau < T_c$, using the notation in the previous paragraph, we show that for any two platform inspection costs that satisfy $K_1 > K_2$, if (i) $R_\tau^I(K_1) > R_\tau^{I'}(K_1)$, then (ii) $R_\tau^I(K_2) > R_\tau^{I'}(K_2)$. Statement (i) is equivalent to

$$\begin{aligned}\frac{(1-q_{p\tau})r}{1-\delta} - K_1 &> C_\tau(1-q_{p\tau})r + S_\tau(r - q_{p\tau}p) + [S_\tau + C_\tau(1-q_{p\tau})]\delta v_{\tau+1}(K_1) \\ M_1 &> K_1 + M_2\delta v_{\tau+1}(K_1), \qquad (12)\end{aligned}$$

where in the last inequality we define $M_1 = \frac{(1-q_{p\tau})r}{1-\delta} - [C_\tau(1-q_{p\tau})r + S_\tau(r - q_{p\tau}p)]$ and $M_2 = [S_\tau + C_\tau(1-q_{p\tau})]$. Similarly, statement (ii) is equivalent to

$$M_1 > K_2 + M_2\delta v_{\tau+1}(K_2). \qquad (13)$$

In order to show that (12) implies (13), it suffices to show that

$$K_1 - K_2 > M_2\delta[v_{\tau+1}(K_2) - v_{\tau+1}(K_1)],$$

and since M_2 is a probability and $\delta \in (0,1)$, the last inequality holds provided

$$v_{\tau+1}(K_2) - v_{\tau+1}(K_1) \leq K_1 - K_2.$$

We prove by induction that $v_t(K_2) - v_t(K_1) \leq K_1 - K_2$ for any $t \in [\tau+1, T_c]$. By Theorem 1, we have that

$$\begin{aligned}v_{T_c}(K_2) - v_{T_c}(K_1) &= \max\{\frac{(1-q_{pT_c})r}{1-\delta} - K_2, \frac{r - q_{pT_c}p}{1-\delta}\} - \max\{\frac{(1-q_{pT_c})r}{1-\delta} - K_1, \frac{r - q_{pT_c}p}{1-\delta}\} \\ &= \max\{I_{T_c}(K_2), 0\} - \max\{I_{T_c}(K_1), 0\},\end{aligned}$$

for $I_{T_c}(K) = \frac{(1-q_{pT_c})r}{1-\delta} - \frac{r - q_{pT_c}p}{1-\delta} - K$. Since $K_2 < K_1$ implies $I_{T_c}(K_2) > I_{T_c}(K_1)$, there are three cases: (i) if $I_{T_c}(K_2), I_{T_c}(K_1) > 0$ then $v_{T_c}(K_2) - v_{T_c}(K_1) = K_1 - K_2$; (ii) if $I_{T_c}(K_2), I_{T_c}(K_1) \leq 0$ then $v_{T_c}(K_2) - v_{T_c}(K_1) = 0$; (iii) if $I_{T_c}(K_2) > 0$ and $I_{T_c}(K_1) \leq 0$, then $v_{T_c}(K_2) - v_{T_c}(K_1) = \frac{(1-q_{pT_c})r}{1-\delta} - \frac{r - q_{pT_c}p}{1-\delta} - K_2 \leq K_1 - K_2$ (by

$I_{T_c}(K_1) \le 0$). Therefore, we have $v_{T_c}(K_2) - v_{T_c}(K_1) \le K_1 - K_2$. We next show that if $v_t(K_2) - v_t(K_1) \le K_1 - K_2$, then $v_{t-1}(K_2) - v_{t-1}(K_1) \le K_1 - K_2$. We have from Theorem 1 that

$$\begin{aligned} v_{t-1}(K_2) - v_{t-1}(K_1) = &\max\{\frac{(1-q_{pt-1})r}{1-\delta} - K_2, C_{t-1}(1-q_{pt-1})r + S_{t-1}(r - q_{pt-1}p) + [S_{t-1} + C_{t-1}(1-q_{pt-1})]\delta v_t(K_2)\} \\ &- \max\{\frac{(1-q_{pt-1})r}{1-\delta} - K_1, C_{t-1}(1-q_{pt-1})r + S_{t-1}(r - q_{pt-1}p) + [S_{t-1} + C_{t-1}(1-q_{pt-1})]\delta v_t(K_1)\} \\ =&\max\{A - K_2, B + C\delta v_t(K_2)\} - \max\{A - K_1, B + C\delta v_t(K_1)\} \\ =&C\delta[v_t(K_2) - v_t(K_1)] + \max\{A - B - C\delta v_t(K_2) - K_2, 0\} - \max\{A - B - C\delta v_t(K_1) - K_1, 0\} \\ =&C\delta[v_t(K_2) - v_t(K_1)] + \max\{W_t(K_2), 0\} - \max\{W_t(K_1), 0\}, \end{aligned}$$

for A, B, C and $W_t(\cdot)$ appropriately defined. Note first that $W_t(K_2) > W_t(K_1)$, since

$$A - B - C\delta v_t(K_2) - K_2 > A - B - C\delta v_t(K_1) - K_1 \quad \Leftrightarrow \quad K_1 - K_2 > C\delta[v_t(K_2) - v_t(K_1)]$$

and the last inequality holds by the inductive hypothesis and $C\delta \in (0,1)$. Therefore, with respect to $v_{t-1}(K_2) - v_{t-1}(K_1)$, there are three possible cases: (i) if $W_t(K_2), W_t(K_1) > 0$, then $v_{t-1}(K_2) - v_{t-1}(K_1) = K_1 - K_2$; (ii) if $W_t(K_2), W_t(K_1) \le 0$, then $v_{t-1}(K_2) - v_{t-1}(K_1) = C\delta[v_t(K_2) - v_t(K_1)] < K_1 - K_2$ (by the inductive hypothesis and $C\delta \in (0,1)$); (iii) if $W_t(K_2) > 0$ and $W_t(K_1) \le 0$, then $v_{t-1}(K_2) - v_{t-1}(K_1) = A - B - K_2 - C\delta v_t(K_1) \le K_1 - K_2$ (by $W_t(K_1) \le 0$). We conclude that $v_{t-1}(K_2) - v_{t-1}(K_1) \le K_1 - K_2$.

Proof of Lemma 4

We use the same notation as in the proof of Lemma 1. Suppose the agent receives the article and does not conduct an inspection. Then

$$\begin{aligned} b_{it} = P(T \mid \mathcal{H}_t, \text{“}T\text{”}) &= P((f,T) \mid \mathcal{H}_t, \text{“}T\text{”}) + P((t,T) \mid \mathcal{H}_t, \text{“}T\text{”}) \\ &= \frac{[0.5q_0 w_t + a(1-q_0)]b_{i0}}{0.5q_0 w_t + (1-q_0)[ab_{i0} + (1-a)(1-b_{i0})]} \end{aligned}$$

To see that in this case $b_{it} > b_{i0}$, note that

$$\frac{b_{it}}{1-b_{it}} = \frac{[0.5q_0 w_t + a(1-q_0)]}{[0.5q_0 w_t + (1-q_0)(1-a)]}\frac{b_{i0}}{1-b_{i0}} > \frac{b_{i0}}{1-b_{i0}},$$

since $a \in (0.5,1)$. Now suppose the agent conducts an inspection. If the agent finds the article to be fake, it follows that $b_{it} = b_{i0}$ since the signal $m = \text{“}T\text{”}$ is uninformative; by contrast, if she finds the article to be truthful, then $b_{it} > b_{i0}$ since the signal $m = \text{“}T\text{”}$ is informative.

Proof of Lemma 5

Fix a sequence of agents. In the absence of platform inspection, we define the impact of a news article over the first N agents as $\Psi_N = \frac{1}{N}\sum_{t=1}^N (b_{it} - b_{i0})$, where $b_{it} = b_{i0}$ if an agent does not receive the article. The impact over all agents is then defined as

$$\Psi_\infty = \lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^N (b_{it} - b_{i0}).$$

We are interested in $\psi_n^v = E[\Psi_\infty \mid v]$. Let L denote the lifetime of an article (i.e., the number of agents the article reaches before it is discontinued). Note that by Proposition 2, L is a random variable that takes values $l \in \mathcal{L} = \{1, 2, ..., T_c - 1\} \cup \{\infty\}$, for some finite T_c. We have

$$\psi_n^v = E[\Psi_\infty \mid v] = \sum_{l\in\mathcal{L}} E[\Psi_\infty \mid v, l]P(l \mid v).$$

Notice that for any $l \in \mathcal{L} \setminus \{\infty\}$, we have

$$E[\Psi_\infty \mid v, l] = E\left[\lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N}(b_{it}-b_{i0})\right] = E\left[\lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{l}(b_{it}-b_{i0})\right] = 0,$$

since $(b_{it}-b_{i0}) \leq 1$. Next, consider the case $l=\infty$. We have

$$\begin{aligned} E[\Psi_\infty \mid v, l=\infty] &= E\left[\lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}(b_{it}-b_{i0})\right] = E\left[\lim_{N\to\infty}\frac{1}{N}\left(\sum_{t=1}^{T_c-1}(b_{it}-b_{i0}) + \sum_{t=T_c}^{N}(b_{it}-b_{i0})\right)\right] \\ &= E\left[\lim_{N\to\infty}\frac{1}{N}\sum_{t=T_c}^{N}(b_{it}-b_{i0})\right] \end{aligned}$$

Observe that in a cascade (which is triggered at time $t=T_c$), the posterior opinion b_{it} is time-invariant, that is, $b_{it}=b_{iT_c}$ $\forall t \geq T_c$, so that $\sum_{t=T_c}^{N}(b_{it}-b_{i0}) = \sum_{t=T_c}^{N}(b_{iT_c}-b_{i0})$, where $(b_{iT_c}-b_{i0}) \leq 1$ is a nonnegative random variable. Therefore,

$$\begin{aligned} E[\Psi_\infty \mid v, l=\infty] &= \lim_{N\to\infty}\frac{\sum_{t=T_c}^{N} E[b_{iT_c}-b_{i0}]}{N} = \lim_{N\to\infty}\frac{(N-T_c+1)E[b_{iT_c}-b_{i0}]}{N} = E[b_{iT_c}-b_{i0}] \\ &= \int_0^1 (b_{iT_c}-b_{i0})dF(b_{i0}) \end{aligned}$$

Using the definition $\lambda_n^v = P(l=\infty \mid v)$, we then have $\psi_n^v = \lambda_n^v \int_0^1 (b_{iT_c}-b_{i0})dF(b_{i0})$.

Proof of Proposition 7

Consider first the expected impact of a fake news article; note that this is given by ψ_n^f in Lemma 5. By Lemma 4, we have that once the fake article triggers a cascade $b_{iT_c} > b_{i0}$ and therefore $\int_0^1 (b_{iT_c}-b_{i0})dF(b_{i0}) > 0$. It follows that $\psi_n^f = \psi^{f*} = 0$ if and only if $\lambda_n^f = 0$. Since $\lambda_n^f = \prod_t S_t$ and S_t is nondecreasing in t, we have that $\lambda_n^f = 0$ if and only if $S_1 = 0$, which by Proposition 3 occurs if and only if $q_0 \geq \frac{2aK}{2aK+(1-K)}$.

Consider next the expected impact of a truthful news article; note that this is given by ψ_n^t in Lemma 5. Observe that $\psi_n^t = \psi^{t*}$ if and only if both (i) $\lambda_n^f = \prod_t (S_t + C_t) = 1$ and (ii) $b_{iT_c} = b_i^{t*}$. Note that condition (i) requires $N_1 = 0$ (which by Proposition 3 then implies $N_t = 0$ for all t), while condition (ii) requires $S_1 = 0$ (which by Proposition 3 then implies $S_t = 0$ for all t). The former occurs if and only if $q_0 < \frac{2(1-a)(1-K)}{2(1-a)(1-K)+K}$, while the latter occurs if and only if $q_0 \geq \frac{2aK}{2aK+(1-K)}$.

Proof of Proposition 8

Consider first the case of fake news. If the platform's optimal policy is such that the platform does not conduct an inspection, the impact of fake news is equivalent to that described in Lemma 5. Now suppose that the platform conducts an inspection in some period $\tau^* \in [1, T_c]$. Since the inspection reveals that the article is fake, it follows that the article is discontinued. In turn, this implies that the article reaches at most a finite number $\tau^* - 1$ of agents before it is discontinued. It then follows from the proof of Lemma 5 that in this case $\psi_n^f = 0$.

Consider next the case of truthful news. If the platform's optimal policy is such that the platform does not conduct an inspection, the impact of truthful news is equivalent to that described in Lemma 5. If the platform's optimal policy is such that the platform conducts an inspection in the first period, the inspection reveals that the article is truthful, the platform announces that $v = t$, and a sharing cascade is triggered in the first period so that $\psi^t = \psi^{t*}$. Finally, in the case where the platform conducts a delayed inspection, there are two possibilities. Either the article survives long enough to be inspected by the platform, at which point a sharing cascade is triggered resulting in $\psi^t = \psi^{t*}$, or the article is discontinued by the agents before the inspection, in which case $\psi^t = 0$ (following the arguments in the proof of Lemma 5). The former event occurs with probability $\lambda^t = \sum_{t=1}^{\tau^*-1}(1 - N_t)$, where the probabilities N_t are given in Proposition 3, so that $\psi^t = \lambda^t \psi^{t*}$.

Proof of Proposition 9

Consider the statements regarding r_1 and r_2. For the first statement, note that the condition $q_0 < Q^{al}$ is independent of r by Proposition 4. Next, notice that by Propositions 7 and 8 we have $\psi^f = \psi^{f*}$ and $\psi^t = \psi^{t*}$ if and only if the platform conducts an inspection at time $\tau^* = 1$. By Proposition 4, this occurs if and only if $q_0 \geq Q^p$. Observe next that Q^p is strictly increasing in r. This implies that if the platform conducts an inspection of an article with characteristics (q_0, a) in the first period for some content-sharing reward r_1, then it also does so for any r_2 such that $r_1 > r_2$. In turn, this implies $\mathcal{M}_l^{r_1,p} \subseteq \mathcal{M}_l^{r_2,p}$.

Consider next the second statement. Note first that the condition $q_0 \geq Q^{ah}$ is independent of r by Proposition 5. Next, notice that by Propositions 7 and 8 we have we have $\psi^f = \psi^{f*}$ and $\psi^t = \psi^{t*}$ if and only if the platform conducts an inspection at time $\tau^* = 1$. By Proposition 5, this occurs if and only if $N_1 \geq N^p$. Next, note that N^p is strictly decreasing in r. This implies that if the platform conducts an inspection of an article (q_0, a) in the first period at some content-sharing reward r_2, then it also conducts an inspection for any r_1 such that $r_1 > r_2$. In turn, this implies $\mathcal{M}_h^{r_1,p} \supseteq \mathcal{M}_h^{r_2,p}$.

The statements regarding p_1 and p_2 follow similarly, by noticing that (i) Q^{al} is independent of p and Q^p is strictly decreasing in p, and (ii) Q^{ah} is independent of p and N^p is also independent of p.

References

Acemoglu, D., M. Dahleh, I. Lobel, A. Ozdaglar. 2011. Bayesian learning in social networks. *The Review of Economic Studies* **78**(4) 1201–1236.

Allcott, H., M. Gentzkow. 2017. Social media and fake news in the 2016 election. *Journal of Economic Perspectives* **31**(2) 211–236.

Allon, G., D. Zhang. 2017. Managing service systems in the presence of social networks. *Working paper, Northwestern University* .

Banerjee, A., R. Somanathan. 2001. A simple model of voice. *The Quarterly Journal of Economics* **116**(1) 189–227.

Banerjee, A.V. 1992. A simple model of herd behavior. *The Quarterly Journal of Economics* **107**(3) 797–817.

Bikhchandani, S., D. Hirshleifer, I. Welch. 1992. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy* **100**(5) 992–1026.

Buzzfeed. 2016. This analysis shows how viral fake election news stories outperformed real news on facebook. (Dec. 15th).

Candogan, O., K. Bimpikis, A. Ozdaglar. 2012. Optimal pricing in networks with externalities. *Operations Research* **60**(4) 883–905.

Candogan, O., K. Drakopoulos. 2017. Optimal signaling of content accuracy: Engagement vs. misinformation .

Chakraborty, S., R. Swinney. 2017. Signaling to the crowd: Private quality information and rewards-based crowdfunding *Working Paper, Duke University.*

Che, Y.-K., N. Kartik. 2009. Opinions as incentives. *Journal of Political Economy* **117**(5) 815–860.

Crapis, D., B. Ifrach, C. Maglaras, M. Scarsini. 2017. Monopoly pricing in the presence of social learning *Management Science, forthcoming.*

Drakopoulos, K., A. Ozdaglar, J. Tsitsiklis. 2013. On learning with finite memory. *IEEE Transactions on Information Theory* **59**(10) 6859–6872.

Durbin, E., G. Iyer. 2009. Corruptible advice. *American Economic Journal: Microeconomics* **1**(2) 220–42.

Feldman, P., Y. Papanastasiou, E. Segev. 2018. Social learning and the design of new experience goods. *Management Science* .

Financial Times. 2017. Facebook to pay fact-checkers to combat fake news. (Apr. 6th).

Gao, F., X. Su. 2016. Omnichannel retail operations with buy-online-and-pick-up-in-store. *Management Science, forthcoming* .

Gentzkow, M., J. Shapiro. 2006. Media bias and reputation. *Journal of Political Economy* **114**(2) 280–316.

Hu, M., J. Milner, J. Wu. 2015. Liking and following and the newsvendor: Operations and marketing policies under social influence. *Management Science* **62**(3) 867–879.

Jun, Y., R. Meng, G.V. Johar. 2017. Perceived social presence reduces fact-checking. *Proceedings of the National Academy of Sciences* **114**(23) 5976–5981.

Manshadi, V. H., S. Misra. 2016. A generalized bass model for product growth in networks. *Working paper, Yale School of Management* .

Marinesi, S., K. Girotra, S. Netessine. 2017. The operational advantages of threshold discounting offers *Manufacturing & Service Operations Management, forthcoming.*

Momot, R., E. Belavina, K. Girotra. 2016. The use and value of social network information in selective selling *Working paper, University of Chicago.*

Morris, S. 1995. The common prior assumption in economic theory. *Economics & Philosophy* **11**(2) 227–253.

Morris, S. 2001. Political correctness. *Journal of political Economy* **109**(2) 231–265.

Ottaviani, M., P. Sørensen. 2006. Professional advice. *Journal of Economic Theory* **126**(1) 120–142.

Papanastasiou, Y. 2018. Newsvendor decisions with two-sided learning. *Available at SSRN 3241279* .

Papanastasiou, Y., N. Bakshi, N. Savva. 2016. Scarcity strategies under quasi-bayesian social learning. *Working Paper, London Business School.*

Papanastasiou, Y., K. Bimpikis, N. Savva. 2017. Crowdsourcing exploration *Management Science, forthcoming.*

Papanastasiou, Y., N. Savva. 2016. Dynamic pricing in the presence of social learning and strategic consumers. *Management Science* **63**(4) 919–939.

Qiu, X., D. Oliveira, A.S. Shirazi, A. Flammini, F. Menczer. 2017. Limited individual attention and online virality of low-quality information. *Nature Human Behaviour* **1**(0132).

Scharfstein, D., J. Stein. 1990. Herd behavior and investment. *The American Economic Review* **80**(3) 465–479.

Shin, D., A. Zeevi. 2017. Dynamic pricing and learning with online product reviews. *Working paper, Columbia University* .

Tech Times. 2017. Facebook replaces 'disputed flags' with 'related articles' to fight against fake news. (Dec. 21st).

Veeraraghavan, S., L. Debo. 2009. Joining longer queues: Information externalities in queue choice. *Manufacturing & Service Operations Management* **11**(4) 543–562.

Vosoughi, S., D. Roy, S. Aral. 2018. The spread of true and false news online. *Science* **359**(6380) 1146–1151.

Yu, M., L. Debo, R. Kapuscinski. 2015. Strategic waiting for consumer-generated quality information: Dynamic pricing of new experience goods. *Management Science* **62**(2) 410–435.