Department of Economics, University of Warwick

Monash Business School, Monash University

as part of

Monash Warwick Alliance

A Model of Online Misinformation with Endogenous Reputation

Andy Lau

Warwick-Monash Economics Student Papers

September 2023 **No: 2023/59**

ISSN 2754-3129 (Online)

https://warwick.ac.uk/fac/soc/economics/research/wmesp/

The Warwick Monash Economics Student Papers (WM-ESP) gather the best Undergraduate and Masters dissertations by Economics students from the University of Warwick and Monash University. This bi-annual paper series showcases research undertaken by our students on a varied range of topics. Papers range in length from 5,000 to 8,000 words depending on whether the student is an undergraduate or postgraduate, and the university they attend. The papers included in the series are carefully selected based on their quality and originality. WM-ESP aims to disseminate research in Economics as well as acknowledge the students for their exemplary work, contributing to the research environment in both departments.

Recommended citation: Lau, A. (2023). A Model of Online Misinformation with Endogenous Reputation. *Warwick Monash Economics Student Papers* 2023/59.

[1] Warwick Economics would like to thank Gianna Boero and Samuel Obeng for their contributions towards the selection process.

A Model of Online Misinformation with Endogenous Reputation

Andy Lau*

Abstract

Misinformation dissemination in social media has emerged as a critical contemporary issue. This paper augments existing models of online misinformation by incorporating endogenous reputation dynamics. In contrast to prior research, reputation plays a pivotal role in shaping agents' Bayesian-Nash equilibrium strategy through two key avenues: (i) the sharer's reputation positively impacts the likelihood of sharing, and (ii) agents with higher initial reputations are less willing to share compared to their counterparts with lower initial reputations. Furthermore, this paper provides insights into the formation of individuals' networks on social media. Surprisingly, individuals with high reputations are not universally favoured as network connections. Additionally, the paper examines relevant comparative statics, including the importance of interactions, and the implications of homophily. This research establishes a foundation for understanding the dynamics of reputation-based information sharing and network structure.

Keywords: Information sharing, misinformation, reputation, network, social media

JEL classification: C72, D83, D85

Email: andy200273@gmail.com. Online appendix.

*I would like to thank Kobi Glazer for his continuous guidance and encouragement. I am also grateful to Kirill Pogorelskiy for his invaluable support. I thank the participants at the 22nd Carroll Round conference at Georgetown University for insightful comments.

1 Introduction

Social media usage has surged worldwide. In the United States, for example, the proportion of citizens relying on social media for news skyrocketed from under 13% in 2008 to over 70% in 2019 (Levy, 2021). However, sharing of misinformation[1] on social media has raised concerns. The spread of fake news and misinformation is comparable to (Grinberg et al., 2019) or higher than (Vosoughi et al., 2018) the number of news articles verified to be true. Research has shown that misinformation on social media can lead to poor collective decision-making (Pogorelskiy and Shum, 2019), and hesitancy in COVID-19 vaccine acceptance (Loomba et al., 2021).

As online misinformation becomes a growing concern, it is critical to understand online sharing behaviour. Acemoglu et al. (2021) model online sharing behaviour as a sequential game in which an agent observes an article, then updates her belief about its veracity given the message and the reliability of the news, and subsequently decides whether to share, ignore, or dislike it.

In this context, reputation is both relevant and crucial, even though it is considered an exogenous variable in their model. The key research question of this paper is, how does reputation influence agents' sharing decisions in equilibrium? This paper introduces endogenous reputation to the theoretical framework, extending the model of online misinformation. In contrast to Acemoglu et al. (2021), this paper models agents who update their ex-post belief given the reputation of the initial sharer. Additionally, they take their own reputation into consideration when making sharing decisions. It lays the foundation for understanding how reputation influences information sharing decisions, and contributes to our understanding of its role in shaping social interactions on social media. It finds that reputation influences agents' sharing strategies in two main ways.

First, agents are more likely to share and less likely to dislike if the person who initially shares it has a high reputation, as the reputation of the sharer positively influences agents' perception of the article's veracity. For instance, a news article shared by a reputable person is more likely to be perceived as truthful, while if it is shared by someone known for spreading fake news, the agent

[1]Misleading and false news articles can be divided into various categories: misinformation (inaccurate or false information); disinformation (false information created to cause harm); fake news (information fabricated to mimic news media); and hyperpartisan news (misleading content with substantial partisan bias) (Pennycook and Rand, 2019; Lazer et al., 2018; Wardle, 2018). This paper will use the terms misinformation and fake news interchangeably, and assumes news articles to be either truthful or to contain misinformation.

may consider it as misinformation.

Second, agents with higher initial reputations are less willing to share, and agents with low initial reputations are more likely to share, as they care about their reputation, and value positive interactions from peers. Specifically, agents with higher initial reputations are cautious in sharing articles to maintain their high reputations, whereas, agents with low initial reputations are more willing to share as they have less to lose in reputation. Overall, an agent's sharing decision depends on both their initial reputation, and the sharer's reputation.

This paper also contributes to the understanding of the role of reputation in individuals' network formation on social media. Notably, agents with high reputations are not always preferred as network connections. Furthermore, agents have an asymmetric preference over their network. In particular, agents prefer highly reputable agents in their incoming network to observe truthful articles, and prefer agents with low reputations in their outgoing network to receive positive interactions.

Finally, this paper examines the impact of reputation in two comparative statics: the importance of interactions, and the effects of homophily. The first comparative statics explains why some individuals are reluctant to share even if they observe a truthful article, as they have a preference to avoid negative feedback. The second comparative statics shows that agents, regardless of reputation, prefer to stay in homophilic network as it encourages positive interactions, and discourages negative interactions.

The remainder of the paper is structured as follows. Section 2 provides a literature review. Section 3 introduces the model. Section 4 characterises the equilibrium and discusses how reputation influences equilibrium strategies. Section 5 explores implications on networks. Section 6 presents comparative statics results. Section 7 concludes. All proofs are available in the online appendix.

2 Related Literature

Recent work by Papanastasiou (2020) and Acemoglu et al. (2021) modelled online sharing decisions in a sequential setting. In Papanastasiou (2020), agents sequentially read an article and potentially share it, and their sharing decisions are modelled as strategic substitutes. In contrast,

Acemoglu et al. (2021) allow agents to dislike articles, and sharing decisions were modelled as strategic complementarities. Research suggests that strategic complements are more appropriate for online sharing behaviour as individual values positive social interactions (Eckles et al., 2016). Nevertheless, reputation, which is an important factor in sharing behaviour, remains exogenous in both settings.

Earlier papers have found that reputation impacts rational agents' decisions, especially in repeated situations (Kreps and Wilson, 1982). Kreps et al. (1982) showed that reputational concerns lead to cooperative behaviour in the repeated prisoners' dilemma. A similar result has been found in an experiment (Ettinger and Jehiel, 2020), where agents have the incentive to build a reputation in a multi-period sender-receiver game. Recent evidence shows that reputation is influential for investment bankers (Lyu et al., 2022), politicians (Bjørnå, 2021), and academic researchers (Petersen et al., 2014).

Regarding reputation, earlier work by Sobel (1985) examines the effect of reputation on strategic information transmission; from his repeated cheap talk model, a rational agent mimics the honest type to develop a reputation. Lunawat (2013a) further establishes that information sharing can be a tool for building a reputation in a repeated situation, and Lunawat (2013b) provides experimental evidence for building a reputation in a repeated investment/trust game. However, these models assume that all shared information is true, without taking into account any potential misinformation.

Individuals care about the veracity of the information they shared, as it impacts their reputation (Pennycook et al., 2021). In the experiment by Altay et al. (2022), they conclude sharing fake news will negatively impact the reputation of the sharer, and recovering from this reputational loss is difficult.

Bala and Goyal (2000) proposes the seminal paper of strategic network formation. Subsequently, many research have studied network formation under a theoretical framework. Zhang and Van Der Schaar (2015) connect network and reputation within a game-theoretical framework. In their paper, agents act under incomplete information, and they form networks based on reputations of other agents. The primary focus of this paper is on the learning process and the formation

of stable networks.

Overall, reputation plays an important role in strategies in repeated situations, making it a crucial factor in information sharing in social media. However, the existing literature on online misinformation sharing does not take into account the effects of reputation. This paper introduces endogenous reputation into the model of online misinformation. Specifically, it builds upon the theoretical model by Acemoglu et al. (2021) where agents receive an article, and decide whether to share, ignore, or dislike it. In contrast to their model, agents make strategic sharing decisions, taking into account both their initial reputations, and the reputation of the sharer.

This paper addresses a gap in online misinformation sharing by incorporating endogenous reputation into a sequential model. It establishes a foundation for understanding the dynamics of reputation based information sharing. This paper also provides crucial insights for the role of reputation in network structure. This paper contributes to the research area by studying the role of reputation in online misinformation sharing behaviour and network structure.

3 Model

There is an underlying state of the world[2], denoted as $\theta \in \{L, R\}$, which can be interpreted as the candidate's suitability for a political office (e.g. left-wing or right-wing). Agents hold heterogeneous ideological prior beliefs about the unknown state θ. Agent i's prior that $\theta = R$, denoted by b_i, is distributed according to the ex-ante distribution $H_i(\cdot)$. Agent i's prior b_i is private knowledge, whereas the ex-ante distribution $H_i(\cdot)$ is observable to all agents.

Sharing network. Assume there are N agents in the population, they share articles within their

[2]This paper follows some of the notations from Acemoglu et al. (2021), including the priors and the underlying state, sharing network, message and veracity of articles, and agents' actions. The reliability of articles has been simplified as it is not the main focus of this paper.

sharing network. The sharing network is denoted by

$$\mathbf{P} \equiv \begin{pmatrix} 0 & P_{12} & \cdots & P_{1N} \\ P_{21} & 0 & \cdots & P_{2N} \\ \cdots & \cdots & \cdots & \cdots \\ P_{N1} & P_{N2} & \cdots & 0 \end{pmatrix}$$

where P_{ij} represents the probability that agent i is linked to agent j. Let N_i be the set of agents attached to agent i with an outgoing link, and $|N_i|$ be the size of her neighbourhood.

News article. Each news article has four dimensional types (r, v, m, μ_h): $r \in [0, 1]$ denotes the reliability of the news, $v \in \{T, M\}$ indicates the veracity of the article (i.e. whether the article is truthful or contains misinformation), $m \in \{L, R\}$ is the article's viewpoint or the message from the article (e.g. in favour of left-wing or right-wing), and $\mu_h \in [0, 1]$ denote the reputation of agent h (i.e. the agent who shares the article before agent i). The type vector (r, v, m, μ_h) is drawn from the independent and identically distributed (i.i.d) process at the beginning of the game.

(i) The reliability of the article $r \in [0, 1]$ is distributed by a continuous function F with density f. When $r = 0$, the article is always misinformation; conversely, when $r = 1$ it is always true.

(ii) An article can either be $v = T$ (truthful) with probability $\mu_h r$, or $v = M$ (contains misinformation) with probability $1 - \mu_h r$.

(iii) If $v = T$ (the article contains truthful information), the probability that the message is generated as $m = \theta$ is $p \geq 1/2$. On the other hand, if $v = M$ (the article contains false information), the probability that the message is generated as $m = \theta$ is $q \leq 1/2$. Thus, when the article is true, the message more likely to match the underlying state.

(iv) Agent i receives article from agent h. The reputation of agent h, denoted as $\mu_h \in [0, 1]$, is drawn from a continuous distribution G with density g. When $\mu_h = 0$, agent h always shares misinformation, and when $\mu_h = 1$, she always shares truthful information[3]. A higher

[3]It is impossible for a perfectly reputable person to share misinformation, and for a perfectly untrustworthy individual to share truthful information (i.e. the combination of $(r, \mu_h) = (0, 1)$ or $(1, 0)$ does not exist).

sharer's reputation is defined as a first-order stochastically dominant shift of the distribution of reputation G to G' where $G' \succsim_{FOSD} G$.

Although v is unknown to all agents, the other dimensions r, m and μ_h are common knowledge (e.g. reliability r depends on some commonly observed factors such as sources, message m is directly observable to all agents in the network, and sharer's reputation μ_h depends on the numbers of shares and dislikes of previously shared articles). It is assumed that agents update their beliefs about v using Bayes' rule given their prior beliefs about the underlying state θ, and the characteristics (r, m, μ_h) of the article[4].

Actions. Time is infinite and discrete, $t = 1, 2, \ldots$. At $t = 1$, sharer h shares an article which is observed by agent i. At $t = 2$, agent i chooses from three possible actions $a_i \in \{S, I, D\}$.

(i) Share (S): Agent shares the article with other agents in her sharing network.

(ii) Ignore (I): Agent ignores the article, and does not engage with it.

(iii) Dislike (D): Agent dislikes the article, and expresses some level of disagreement.

If agent i shares an article, it will be received by all $j \in N_i$. On the other hand, if agent i ignores or dislikes the article, it will not be disseminated beyond agent i.

Payoffs and reputation. Assume the action of dislikes does not affect one's reputation, but one's reputation will be influenced by others' dislikes, since it is often easier to see the number of dislikes for an article on social media than to see what a person has previously disliked.

Let $S_i = \frac{|j \in N_i : a_j = S\}|}{|N_i|}$ denote the proportion of agents who share after i, and $D_i = \frac{|\{j \in N_i : a_j = D\}|}{|N_i|}$ denote the proportion of agents who dislike after i. Unlike Acemoglu et al. (2021), who measure interactions in terms of the number of agents who share or dislike after i, modelling them as proportions in agent i's network is more appropriate. This is because the number of agents depends on the size of networks which is exogenous. This distinction is crucial in eliminating trivial equilibria, as discussed in Lemma 1.

The initial reputation of agent i, denoted by $\mu_{i0} \in [0, 1]$, is drawn from a continuous distribution Y with density y. A higher potential for initial reputation of agent i is defined by a first-order

[4]The updating process is similar with the paper by Acemoglu et al. (2021), where the reputation of sharer is endogenous in this paper.

stochastically dominant shift of the distribution Y to Y' where $Y' \succsim_{FOSD} Y$. Let $\mu_i{}'$ denote the reputation of agent i after she shares. The difference between the actual and expected proportion of shares and dislikes is denoted by $\tilde{S}_i$ and $\tilde{D}_i$, respectively. Specifically, the functions $\phi_s(\cdot)$ and $\phi_d(\cdot)$ translate an agent's initial reputation into an expected proportion of agents in her network who share and dislike, $\phi_s(\mu_{i0}) = \mathbb{E}[S_i|\mu_{i0}]$, and $\phi_d(\mu_{i0}) = \mathbb{E}[D_i|\mu_{i0}]$, respectively, where $\phi_s(0) = \phi_d(1) = 0$, $\phi_s(1) = \phi_d(0) = 1$, $\phi_s'(\mu_{i0}) > 0$ and $\phi_d'(\mu_{i0}) < 0$ (i.e. higher initial reputation implies a higher expected proportion of shares, and a lower expected proportion of dislikes). Thus, $\tilde{S}_i = S_i - \mathbb{E}[S_i|\mu_{i0}]$ and $\tilde{D}_i = D_i - \mathbb{E}[D_i|\mu_{i0}]$.

When agent i shares an article, her reputation increases if the actual proportion of shares is greater than the expected proportion (i.e. $\tilde{S}_i > 0$), but decreases if the proportion of dislikes is larger than expected (i.e. $\tilde{D}_i > 0$). The reputation of agent i is characterised by the following equation:

$$\mu_i{}' = \begin{cases} 1 & \text{if} \quad \mu_{i0} + \alpha\gamma(\tilde{S}_i) - \beta\gamma(\tilde{D}_i) \geq 1 \\ \mu_{i0} + \alpha\gamma(\tilde{S}_i) - \beta\gamma(\tilde{D}_i) & \text{if} \quad 0 < \mu_{i0} + \alpha\gamma(\tilde{S}_i) - \beta\gamma(\tilde{D}_i) < 1 \\ 0 & \text{if} \quad \mu_{i0} + \alpha\gamma(\tilde{S}_i) - \beta\gamma(\tilde{D}_i) \leq 1 \end{cases} \tag{1}$$

where $\gamma(\cdot)$ is an increasing concave function, differentiable in $[-1, 1]$, $\gamma(1) = 1$, $\gamma(0) = 0$, and $\gamma(-1) = -1$. Additionally, $\alpha \in [0, 1]$ and $\beta \in [0, 1]$ represent the importance of positive feedback and negative feedback respectively.

Agents may react differently when an article is shared by a highly reputable agent or an agent with a low reputation (i.e. reputation may be updated in a Bayesian way given the reputation of the agent who subsequently shares or dislikes it). However, it is more reasonable to assume reputation depends on the proportion of agents who share or dislike the article. This is because, on social media, it is easier to observe the number of shares and dislikes of an article than to determine exactly which agents have shared or disliked it. Therefore, it is more appropriate to model reputation so that each instance of sharing or disliking contributes equally to the reputation, regardless of the specific agent who took the action.

The change in reputation after sharing an article is $\Delta\mu_i = \mu_i{}' - \mu_{i0} = \alpha\gamma(\tilde{S}_i) - \beta\gamma(\tilde{D}_i)$. Agent

i's expected utility is defined as:

$$U_i = \begin{cases} 0 & \text{if} \quad a_i = I \\ \tilde{u}1_{v=M} - \tilde{c} & \text{if} \quad a_i = D \\ u1_{v=T} - c1_{v=M} + k\Delta\mu_i & \text{if} \quad a_i = S \end{cases} \tag{2}$$

where 1 is the indicator function (equal to 1 if it is true, 0 otherwise), and $\tilde{u}, \tilde{c}, u, c, k$ are strictly positive parameters.

(i) Payoff from ignore, I, is normalised to 0: $U_i(I) = 0$.

(ii) Payoff from dislike, D, depends on the veracity of the information, $\tilde{u}$ denotes the utility from disliking misinformation (i.e. expressing dissatisfaction with misinformation), while $\tilde{c}$ denotes the cost associated with disliking, regardless of the truthfulness of information. It is assumed that $\tilde{u} > \tilde{c}$, implying that agents receive positive utility when disliking misinformation, but incur negative utility when disliking truthful articles.

(iii) Payoff from share, S, depends on both the veracity of the information and the change in reputation, u represents the utility derived from sharing truthful information, c signifies the cost incurred when sharing misinformation, and k denotes the importance of reputation. The utility from positive peer interactions (modelled by shares) or negative peer interactions (modelled by dislikes) is reflected in changes in reputation.

Summary of the sequential game. Initially, agent i has a prior about the underlying state θ. At $t = 1$, agent h shares an article, and agent i observes it. Agent i then updates her posterior belief about the veracity of the article v using Bayes' rule given her prior belief, and the characteristics (r, m, μ_h) of the article. At $t = 2$, she decides whether to share (S), ignore (I) or dislike (D) the article. Agent i's reputation $\mu_i{}'$ and payoff increase when more agents in her network share, but decreases when more agents dislike the article.

Trivial equilibria. Trivial equilibria exist if there is a dominant strategy for every agent (e.g. all agents always share every article for all ex-post belief, regardless of their initial reputation). The following lemma asserts the non-existence of such trivial equilibria.

Lemma 1. *There are no pure strategy equilibria in which all agents play a dominant strategy.*

This lemma states that share, ignore, and dislike cannot be a dominant strategy for every agent. This differs from the result in Acemoglu et al. (2021), where sharing can be a dominant strategy for all agents in their model. In this model, shares and dislikes are measured in proportion, and the utility from sharing depends not only on the ex-post beliefs but also on the agent's initial reputation. Some agents will be strictly worse off if their initial reputation is high but their ex-post belief is low (as $\Delta\mu_i$ can be negative), i.e. the expected cost of reputational loss is greater than the expected benefit from sharing truthful information. In summary, while sharing can be a dominant strategy for certain individuals, it cannot be the dominant strategy for all agents, thus negating the existence of the trivial equilibrium where all agents always share.

4 Equilibrium

This section illustrates the agent's (Bayesian-Nash) equilibrium strategy for a general network, $\mathbf{P}$, and characterises how reputation influences these equilibrium strategies. Without loss of generality, let the article's message $m = R$ for the rest of the paper.

4.1 Ex-post Belief

Agent i has a prior that $\theta = R : b_i$, and updates her ex-post belief about the veracity of the information when she observes an article. She updates ex-post belief π_i given reliability r, message m, and sharer h's reputation μ_h, that the article is truthful using Bayes' rule. Her ex-post belief is:

$$\pi_i = \frac{(pb_i + (1-p)(1-b_i))\mu_h r}{(pb_i + (1-p)(1-b_i))\mu_h r + (qb_i + (1-q)(1-b_i))(1-\mu_h r)} \tag{3}$$

which is increasing in μ_h, as agents are more likely to believe an article to be truthful if the person who shares it is more reputable. This suggests social media interactions exhibit strategic complementarities which match the result in Acemoglu et al. (2021) and Eckles et al. (2016).

In addition, reputations amplify the extent of strategic complementarities. In other words, when agents receive more shares in their network, their reputations increase, which in turn encourages more shares. Thus, agents are motivated to cohere with others' behaviour as they value interactions from peers.

4.2 Equilibrium Strategy

Bayesian-Nash equilibrium exists in the form of cutoff strategies, and the set of cutoffs $(\mathbf{b}^*, \mathbf{b}^{**})$ forms a complete lattice which has an infimum (i.e. the most-sharing equilibrium) and a supremum (i.e. the least-sharing equilibrium) (Acemoglu et al., 2021). Agent i's mixed strategy, denoted by σ_i, maps prior b_i to an element of the simplex $\Delta(\{D, I, S\})$. The vector of strategies of all agents in the sharing network is denoted by $\boldsymbol{\sigma}_{-i}$. Given any strategies $\boldsymbol{\sigma}_{-i}$, agent i's best response is a cutoffs strategy, with cutoffs (b_i^*, b_i^{**}), where agent i:

(i) Shares (S) if $b_i > b_i^{**}$

(ii) Ignores (I) if $b_i^* < b_i < b_i^{**}$

(iii) Dislikes (D) if $b_i < b_i^*$

Overall, the first cutoff is a function of reliability and sharer's reputation, $b_i^*(r, \mu_h)$, where the second cutoff is a function of reliability, sharer's reputation, own initial reputation, and importance of positive and negative interactions, $b_i^{**}(r, \mu_h, \mu_{i0}, \alpha, \beta)$. The following proposition presents the first result of this paper, which demonstrates how the reputation of the sharer influences an agent's equilibrium strategy.

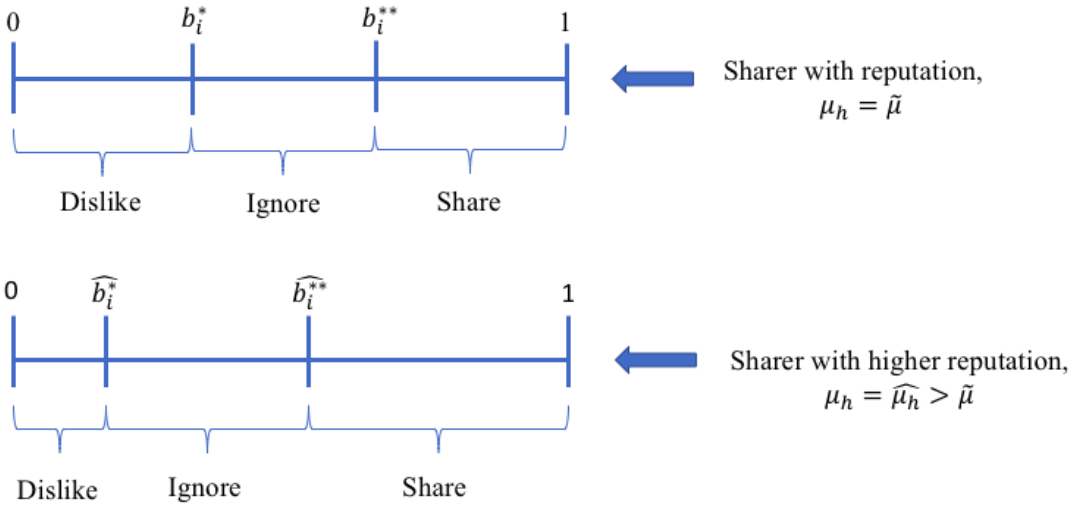
Proposition 1. *For any* $\widehat{\mu_h} > \mu_h$,

$$i.\ b_i^*(\widehat{\mu_h}) < b_i^*(\mu_h)$$

$$ii.\ b_i^{**}(\widehat{\mu_h}) < b_i^{**}(\mu_h)$$

This proposition states that, in equilibrium, agents are more likely to share and less likely to dislike an article when it is shared by a reputable person. The agents' decisions, therefore, depend on

Figure 1: Effects of a sharer with high reputation on agent's equilibrium strategy



the reputation of the sharer. When the sharer's reputation is high, agents' ex-post beliefs about the truthfulness of the article increase, subsequently increasing the expected utility from sharing, and decreasing the expected utility from disliking. This result is shown by a decrease in both cutoffs (b_i^*, b_i^{**}). Figure 1 shows that a sharer with higher reputation shifts the cutoffs leftward from (b_i^*, b_i^{**}) to $(\widehat{b_i^*}, \widehat{b_i^{**}})$, increasing the likelihood of shares and lowering the likelihood of dislikes.

This finding contrasts the results in Acemoglu et al. (2021), in which they did not model the impact of the reputation of the initial sharer. By introducing reputation, agents take into account the reputation of the sharer, and update their ex-post beliefs accordingly. This offers additional insights into situations where an individual's reputation influences others' belief in an article's truthfulness, and ultimately influencing their sharing decisions. This suggests an article's perceived truthfulness depends not only on a random draw but also on the reputation of the initial sharer, suggesting that an agent's reputation has a dynamic effect in repeated situations.

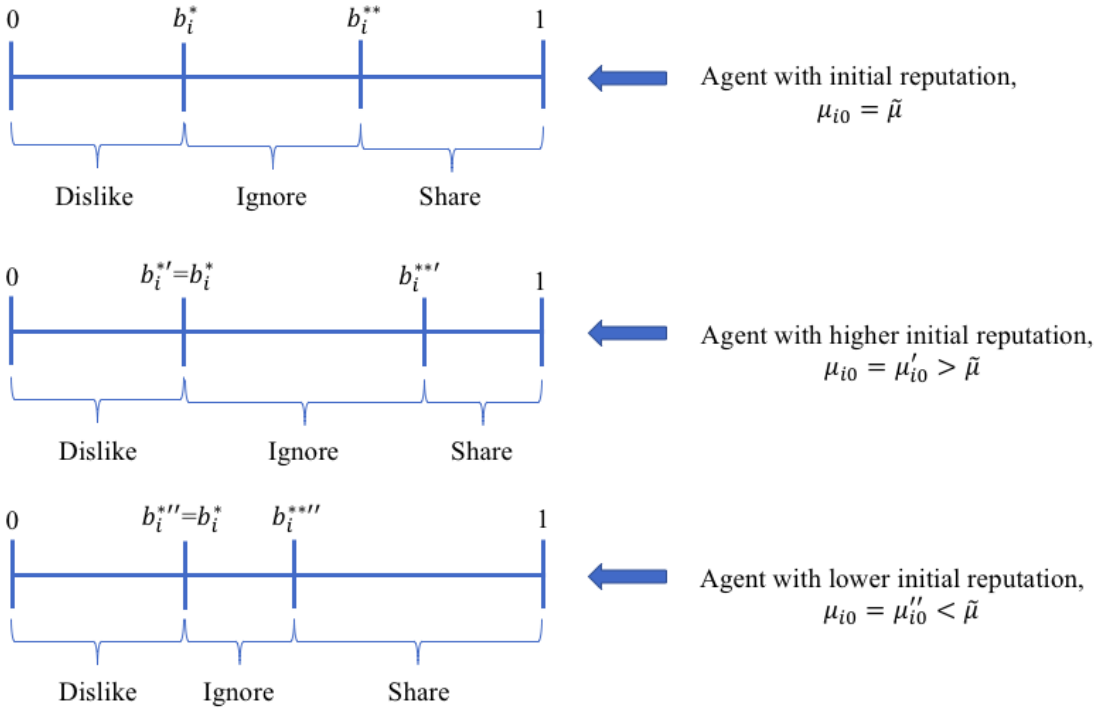
The second result of this paper identifies the relationship between own reputation and the decision to share.

Proposition 2. *For any* $\mu_{i0}'' < \mu_{i0} < \mu_{i0}'$,

$$i.\ b_i^*(\mu_{i0}) = b_i^*(\mu_{i0}') = b_i^*(\mu_{i0}'')$$

$$ii.\ b_i^{**}(\mu_{i0}') > b_i^{**}(\mu_{i0}) > b_i^{**}(\mu_{i0}'')$$

Figure 2: Effects of initial reputation on agent's equilibrium strategy



The proposition states that agents with higher initial reputations are less willing to share, while agents with lower initial reputations are more likely to share. Moreover, agents' concerns about reputation do not affect their decisions to dislike. Overall, an agent's reputation is only influenced by their decision to share, which, in turn, depends on their initial reputation.

Agents with high initial reputations are more cautious in sharing articles as the scope for improving their reputation is limited. On the other hand, agents with low initial reputations are more willing to share as they have less to lose in terms of reputation, and sharing is the only way for them to improve it. This is illustrated in Figure 2. The second cutoff increases with the level of reputation ($b_i^{**'} > b_i^{**} > b_i^{**''}$), while the first cutoff remains the same regardless of reputation ($b_i^* = b_i^{*'} = b_i^{*''}$). This finding suggests that agents with higher initial reputation only share articles with high priors and ex-post beliefs, while agents with lower initial reputation share articles with lower priors and ex-post beliefs.

The result has intriguing implications for agents with low initial reputation. When reputation becomes endogenous, agents with lower initial reputations become more willing to share articles. However, they may also share articles that they would not share if they had a higher initial

reputation. In other words, they share articles with lower priors and ex-post beliefs, since the utility is compensated by the potential gain in reputation. Consequently, they tend to share articles that are less likely to be truthful, making it even more difficult to recover their reputation, which is consistent with the empirical evidence in Altay et al. (2022).

Furthermore, since agents with higher initial reputations only share articles with high priors and ex-post beliefs, their reputations are likely to improve over time. As a result, reputations will tend to converge towards two extremes if repeated interaction is introduced. Specifically, highly reputable agents will observe improvements in their reputation over time, while agents with lower initial reputations will experience a decline. The exact cutoff can be formally defined after introducing repeated interactions. This implies that agents may transition from a continuous consideration of reputation to a more discrete approach over time. They may come to believe all information shared by highly reputable agents, while disbelieving articles shared by agents with low reputations.

Comparison between Proposition 1 and Proposition 2. While both sharer's reputation and the agent's initial reputation influence the equilibrium strategy, they affect it through different channels. The sharer's reputation influences cutoffs (b_i^*, b_i^{**}) through changes in agent i's ex-post belief π_i, which indirectly affects the expected payoff. Conversely, the agent's initial reputation directly impacts the equilibrium strategy (b_i^*, b_i^{**}) by modifying the expected payoff.

5 Implications on Network

Although this paper does not fully endogenize agents' networks, some implications about the network can be discussed from the results above. Since ex-post belief of the veracity of the article is increasing in the reputation of the sharer, agents prefer the sharer to have a higher reputation if they prefer to observe true articles. Additionally, agents prefer to have a higher reputation to obtain a higher utility. An intuitive question to ask is whether agents always prefer their peers in the network to have high reputations. This paper demonstrates an asymmetric result on agents' preference of peers' reputation.

It is crucial to note that an agent's network can be divided into two groups: agents with an

incoming link, and those with an outgoing link. Agents with an incoming link refers to the groups of agents from whom agent i receives articles, and agents with an outgoing link refers to the agents who observe articles shared by agent i. The people from whom an agent receives articles do not necessarily have to be the same people that they interact with, on social media. For instance, individuals may receive articles from journalists or politicians, while their interactions on social media are primarily with their friends.

There are two sets of agents: N_i' which denotes the set of agents attached to agent i with an incoming link, and N_i which denotes the set of agents attached to agent i with an outgoing link. These two sets can be disjoint. Thus, agent i's network is denoted as:

$$\mathbf{P_i} \equiv \begin{pmatrix} P_{1i} & P_{2i} & \cdots & P_{Ni} \\ P_{i1} & P_{i2} & \cdots & P_{iN} \end{pmatrix}$$

where the first row (P_{hi}) represents the probability that an agent is linked to agent i with an incoming link, and the second row (P_{ij}) is the probability that an agent is linked to agent i with an outgoing link. The following proposition describes an agent's asymmetric preference of her network.

Proposition 3. *Suppose $P_{hi}(\mu_h)$ and $P_{ij}(\mu_j)$, then*

i. If $\frac{\partial \pi_i}{\partial P_{hi}} > 0$, then $\frac{\partial P_{hi}}{\partial \mu_h} > 0$

ii. If $\frac{\partial U_i(S)}{\partial P_{ij}} > 0$, then $\frac{\partial P_{ij}}{\partial \mu_j} < 0$.

The first part of the proposition highlights that when agents select their incoming network based on agents' reputations $P_{hi}(\mu_h)$ with a preference for observing truthful articles, they prefer agents with higher reputations to be in their incoming link, expressed as $\frac{\partial P_{hi}}{\partial \mu_h} > 0$. This indicates a clear preference for highly reputable agents within their incoming network.

The second part of the proposition states that when agents select their outgoing network to increase their utility from sharing (which is true, as the outgoing network only impacts the utility from sharing, but not not ignoring or disliking), they prefer to have outgoing links with agents who

have low reputations. This is because low-reputation agents are more likely to share, as stated in Proposition 2.

This highlights that agents do not uniformly favour highly reputable individuals as their network connections, revealing an asymmetric preference over their network. When considering their peers' reputation, agents exhibit distinct preferences for their network structure. They prefer highly reputable individuals in their incoming network, emphasising the importance of observing truthful information. In contrast, agents prefer individuals with low reputations in their outgoing network. This asymmetric preference arises from the different roles played by incoming and outgoing networks. The incoming network determines how likely she is going to observe truthful articles, whereas the outgoing network dictates the interactions an agent engages in.

6 Comparative Statics

This section examines how the importance of interactions and homophily influences agents' reputations and equilibrium strategies.

6.1 Importance of interaction

Agents may feel differently when they receive positive or negative feedback (i.e. $\alpha \neq \beta$). For example, an agent may prefer to avoid dislikes rather than acquire shares. The following proposition describes how the importance of each type of interactions influences their sharing decisions.

Proposition 4. *If* $\alpha' > \tilde{\alpha} > \alpha''$, $\beta' > \tilde{\beta} > \beta''$, *and* $\tilde{\alpha} = \tilde{\beta}$, *then*

$$i.\ b_i^{**}(\alpha', \tilde{\beta}) < b_i^{**}(\tilde{\alpha}, \tilde{\beta}) < b_i^{**}(\alpha'', \tilde{\beta})$$

$$ii.\ b_i^{**}(\tilde{\alpha}, \beta') > b_i^{**}(\tilde{\alpha}, \tilde{\beta}) > b_i^{**}(\tilde{\alpha}, \beta'')$$

This result states that if an agent values positive interactions more than negative interactions, then the agent is more likely to share. Conversely, if an agent values negative interactions more, then she is less likely to share. For example, if $\alpha > \beta$ and the proportion of interactions are the

same, the gain from positive interaction is greater than the loss from negative feedback, thus, the expected payoff from share is higher, and causes a leftward shift in b_i^{**}. Also, the importance of interactions only impacts sharing decisions as there is no interaction when an agent ignores or dislikes.

This result suggests that agents share more to receive positive feedback or share less to avoid negative feedback. Therefore, for agents with a strong preference to avoid negative feedback, they are hesitant to share even when they observe articles with high priors and ex-post beliefs.

6.2 Homophily

Homophily is the phenomenon of people tending to form connections with others who share similar beliefs or characteristics. Assuming that other agents' priors are distributed with $H_{-i}^L(\cdot)$ or $H_{-i}^R(\cdot)$, where $H_{-i}^R(\cdot)$ implies the distribution are more likely to assign higher values than $H_{-i}^L(\cdot)$, i.e. $H_{-i}^R(\cdot) \gtrsim_{FOSD} H_{-i}^L(\cdot)$. Without loss of generality, assume a more homophilic network $\mathbf{P^H}$ (associated with the sharing network N_i^H) has a higher proportion of agents with $H_{-i}^R(\cdot)$ compared to the network $\mathbf{P}$. In other words, there is a greater proportion of agents with higher priors b_{-i} in N_i^H than in N_i. The level of homophily is increasing in the proportion of like-minded people (i.e. higher priors) in their network.

Proposition 5. *For network N_i^H and N_i,*

$$i.\ \mathbb{E}[S_i|N_i^H] > \mathbb{E}[S_i|N_i]$$

$$ii.\ \mathbb{E}[D_i|N_i^H] < \mathbb{E}[D_i|N_i]$$

This proposition establishes that homophily leads to more sharing, and less dislike regardless of reputation. In a highly homophilic network, agents with low reputation can effectively build their reputation by sharing articles, because their peers are more likely to share, and less likely to dislike the content. Similarly, agents with high reputation are more willing to share, because the likelihood of resulting reputational loss is smaller. Therefore, agents prefer to stay in a homophilic network.

This proposition suggests that positive interactions are encouraged, and negative interactions are discouraged in a homophilic network. This finding contrasts with Acemoglu et al. (2021), where they consider homophily in terms of content virality and platform design from the perspective of firms. Whereas, this paper focuses on information sharing and agents' sharing decisions. When reputation becomes endogenous, agents prefer to remain in a homophilic network where sharing is encouraged. This implies homophilic network is preferred by all agent when reputation becomes endogenous, suggesting a potential reason for the prevalence of homophilic network formation on social media.

7 Conclusion

This paper introduces endogenous reputation to the model of online misinformation presented by Acemoglu et al. (2021), thereby broadening its theoretical implications and providing new opportunities for discussion. It demonstrates that agents' sharing decisions are influenced by both their own reputations, and the reputation of the sharer. Additionally, the incorporation of endogenous reputation produces notable implications on networks that have not been previously discussed in the literature. Overall, this paper presents two key findings on how reputation shapes an agent's equilibrium sharing strategy.

First, the reputation of the person who initially shares the article influences agents' sharing decisions. Specifically, agents are more likely to share, and less likely to dislike when the sharer's reputation is higher. This result improves upon the model in Acemoglu et al. (2021), which did not account for the reputation of the initial sharer. This paper demonstrates that the reputation of the sharer influences agent's perception of the truthfulness of the article. It indicates that agents' reputations dynamically influence how others perceive the truthfulness of the article, particularly in repeated situations where the article is shared multiple times, rather than being determined by a random draw each time.

Second, the agent's initial reputation influences their equilibrium strategy in sharing. Agents with higher initial reputations are less willing to share, and those with lower initial reputations tend to share more frequently. This result has intriguing implications for those with low initial

reputations. This encourages them to share more articles, especially those with lower priors and ex-post beliefs, which they would refrain from sharing if they possessed a higher initial reputation. Consequently, they end up sharing articles that are less likely to be truthful, making it increasingly challenging for them to rebuild their reputation. In contrast, agents with higher initial reputations are more cautious in sharing, leading to an even stronger reputations. If repeated interaction is introduced, it will create a positive feedback loop for more reputable agent, and a negative feedback loop for those with lower initial reputations. As a result, agents' reputations are likely to converge to two extremes over time.

The broader implications for networks extend beyond the field of network economics. It highlights that individuals with high reputations are not universally favoured as network connections. Furthermore, agents demonstrate an asymmetric preference of their network peers: they favour highly reputable individuals in their incoming network but prefer less reputable ones in their outgoing network. This is because agents prefer to observe articles that are more likely to be true, and they seek more positive interaction from their peers.

This paper discusses two comparative statics: importance of interaction, and homophily. The first comparative statics addresses why individuals may hesitate to share truthful articles, due to a strong preference for avoiding negative feedback. The second comparative statics establishes that agents prefer to stay in a homophilic network, irrespective of their reputations. This provides an intriguing implication: homophilic network is preferred by all agent when reputation is endogenous. This insight suggests that the prevalence of homophilic networks on social media may be attributed, in part, to the endogenous nature of reputation.

There are two main limitations in this paper. First, this paper focuses only on the agent's strategic perspective. Future research can investigate how reputation in this context influences platform design (e.g. some social media platforms have removed the dislike function, or have hidden the number of dislikes), and implications to policies (e.g. provenance policy, censorship, and transparency). Second, the paper does not incorporate endogenous network dynamics, which can be a promising area for future research. Further research can introduce endogenous network, and explore the effects of an agent's reputation on the equilibrium distribution of reputation,

and the equilibrium network formation. Furthermore, future research can introduce asymmetry in information structure (i.e. different decay factors for different types of information) into this model of sharing behaviour, and examine its effect on the virality of articles. Some of the result of this paper can be tested empirically if data is available (e.g. survey data or data from experiments). Research can compare the coefficient for the willingness to share for an article received from individuals with different reputations. Some of the parameters (e.g. α and β) may be estimated numerically through calibration, or in an experimental setting.

References

Acemoglu, D. et al. (2021) "A Model of Online Misinformation," *NBER Working Papers*, 28884, National Bureau of Economic Research, Inc.

Altay, S. et al. (2022) "Why do so few people share fake news? it hurts their reputation," *New Media and Society*, 24(6), pp. 1303–1324.

Bala, V and Goyal, S. (2000) "A Noncooperative Model of Network Formation," *Econometrica*, 68(5), pp. 1181-1229.

Bjørnå, H. (2021) "Reputational assets for local political leadership," *Heliyon*, 7(8).

Eckles, D. et al. (2016) "Estimating peer effects in networks with peer encouragement designs," *Proceedings of the National Academy of Sciences, 113(27)*, pp. 7316–7322.

Ettinger, D. and Jehiel, P. (2020) "An experiment on deception, reputation and trust," *Experimental Economics*, 24(3), pp. 821–853.

Grinberg, N. et al. (2019) "Fake news on Twitter during the 2016 U.S. presidential election," *Science*, 363(6425), pp. 374–378.

Kreps, D.M. and Wilson, R. (1982) "Reputation and imperfect information," *Journal of Economic Theory*, 27(2), pp. 253–279.

Kreps, D.M. et al. (1982) "Rational cooperation in the finitely repeated prisoners' dilemma," *Journal of Economic Theory*, 27(2), pp. 245–252.

Lazer, D. et al. (2018) "The science of fake news," *Science*, 359(6380), 1094–1096.

Levy, R. (2021) "Social Media, news consumption, and polarization: Evidence from a field experiment," *American Economic Review*, 111(3), pp. 831–870.

Loomba, S. et al. (2021) "Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA," *Nature Human Behaviour*, 5(3), pp. 337–348.

Lunawat, R. (2013a) "The role of information in building reputation in an investment/Trust game," *European Accounting Review*, 22(3), pp. 513–532.

Lunawat, R. (2013b) "An experimental investigation of reputation effects of disclosure in an investment/Trust game," *Journal of Economic Behavior and Organization*, 94, pp. 130–144.

Lyu, H. et al. (2022) "Individual investment bankers' reputation concerns and bond yield spreads: Evidence from China," *Journal of Banking and Finance*, 140, p. 106508.

Papanastasiou, Y. (2020) "Fake news propagation and detection: A sequential model," *Management Science*, 66(5), pp. 1826–1846.

Pennycook, G. and Rand, D.G. (2019) "Fighting misinformation on social media using crowdsourced judgments of news source quality," *Proceedings of the National Academy of Sciences, 116(7)*, pp. 2521–2526.

Pennycook, G. et al. (2021) "Shifting attention to accuracy can reduce misinformation online," *Nature*, 592(7855), pp. 590–595.

Petersen, A. M. et al. (2014) "Reputation and impact in academic careers," *Proceedings of the National Academy of Sciences of the United States of America, 111(43)*, 15316–15321.

Pogorelskiy, K. and Shum, M. (2019) "News sharing and voting on social networks: An experimental study," *SSRN Electronic Journal*

Sobel, J. (1985) "A Theory of Credibility," *The Review of Economic Studies*, Vol. 52, No. 4 (Oct., 1985), pp. 557-573, 52(4), pp. 557–573.

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

Wardle, C. (2018) "Information disorder: The essential glossary," *Shorenstein Center on Media, Politics, and Public Policy, Harvard Kennedy School.*

Zhang, S. and Van Der Schaar, M. (2015) "Reputational Learning and Network Dynamics," *Papers 1507.04065, arXiv.org.*