UNIVERSITY OF CAMBRIDGE
Faculty of Economics

Cambridge Working Paper Economics

Cambridge Working Paper Economics: 1747

SPREADING LIES

Bartosz Redlicki

26 October 2017

The recent increase in partisan media has generated interest in what drives media outlets to become more partisan. I develop a model where a media outlet can report news with a partisan slant and the news then spread by word of mouth in a population of agents with heterogeneous preferences. The level of slant has an impact on whether the agents find the news credible and on their incentives to pass the news to others. The analysis elucidates how partisanship of media can be driven by political polarisation of the public and by the tendency of people to interact with people with similar political views. Extensions of the model shed light on the influences of social media and the fact that people with different political views tend to trust different media outlets.

Spreading Lies

Bartosz Redlicki*

October 24, 2017

Abstract

The recent increase in partisan media has generated interest in what drives media outlets to become more partisan. I develop a model where a media outlet can report news with a partisan slant and the news then spread by word of mouth in a population of agents with heterogeneous preferences. The level of slant has an impact on whether the agents find the news credible and on their incentives to pass the news to others. The analysis elucidates how partisanship of media can be driven by political polarisation of the public and by the tendency of people to interact with people with similar political views. Extensions of the model shed light on the influences of social media and the fact that people with different political views tend to trust different media outlets.

JEL Classification: D82, D83, L82

Keywords: media slant, partisan media, word of mouth, Bayesian persuasion

*Faculty of Economics, University of Cambridge. E-mail: bar43@cam.ac.uk. I am especially grateful to Sönje Reiche and Edoardo Gallo for their feedback. I would also like to thank Francis Bloch, Matthew Elliott, Robert Evans, Paweł Gola, Sanjeev Goyal, David Minarsch, Jakub Redlicki, Hamid Sabourian and many others, as well as seminar participants at the University of Cambridge, the World Congress of the Game Theory Society (Maastricht), Paris School of Economics, the Hurwicz Workshop (Warsaw), and the University of Bristol for their comments. I gratefully acknowledge financial support from the Economic and Social Research Council (UK).

1 Introduction

Empirical evidence from the United States and other countries shows that media outlets manipulate information with a partisan slant, i.e. in a way that systematically favours one side of the political spectrum or the other. The determinants of such media slant have been widely analysed by the theoretical literature; however, little attention has been given to the fact that the information reported by media can spread in the public by word of mouth. The aim of this paper is to develop a model of partisan media slant with diffusion by word of mouth, in order to better understand media outlets' motivations for becoming more partisan.

To set the scene, consider an example of a pro-Republican media outlet (e.g., a newspaper, a website, or a TV channel), whose objective is to make as many people as possible support the Republican Party. The outlet can manipulate the news which it reports by giving them a pro-Republican slant. The news reports are first seen by the outlet's Republican audience and then spread in the public by word of mouth. In the public, there are people with heterogeneous political views: Republicans and Democrats, which means they have their own incentives to share or not share the news with others.

The above example motivates the following model. There is a manipulator and a population of agents. The state of the world is binary: it is *good* or *bad*. During the course of the game, each agent in the population chooses an individual action; if her action is sufficiently high, then we say that she has been *persuaded*. The manipulator's objective is to maximise the number of persuaded agents. There are two types of agents in the population: high-type (H-type) and low type (L-type). Each type of agent prefers a higher action if the state is good than if the state is bad, but H-type agents are biased in favour of the manipulator: for a given state of the world, they prefer a higher action than L-type agents. Each agent wants other agents to take an action that is as close as possible to her own action.

The game unfolds as follows. First, the manipulator designs an *information policy*, which is a map from each possible state of the world (*good* and *bad*) to a probability distribution over possible news reports (*good* and *bad*). The news report generated by

the information policy is observed by a randomly selected H-type agent and then diffuses via a communication chain: each agent, upon receiving the report, meets another agent, observes her type, and chooses whether to pass it on to her or not. The assortativity of meetings determines how likely it is that an agent meets an agent of the same or the opposite type. Diffusion continues until (a) an agent fails to meet a successor, which happens with an exogenous positive probability to each agent, or (b) an agent decides not to pass the news report on. At that point, each agent in the population chooses her individual action and the game ends.

This paper studies the optimal information policy for the manipulator. In the optimal policy, the news report must be good whenever the state is good, so the key question is: how often should the news report be good when the state is bad? In other words, how often should the manipulator lie in his favour? I refer to the probability that the news report is good when the state is bad as the *slant* of the information policy. Thus, the slant captures here lying in the most direct form, i.e. negating the truth.

The manipulator needs to consider two effects of the information policy: (1) on the credibility of news reports and (2) on the agents' incentives to pass the news reports on. First, if the slant is too high, then a good report is not credible enough, and the agents are not persuaded by it (due to their bias, H-type agents can be persuaded under a higher slant than L-type agents). Second, more interestingly, if the slant is too high, then a good report does not spread so well, as some of the agents no longer have an incentive to pass it on (more specifically, L-type agents prefer not to pass it on to H-type agents). These two effects provide an incentive for the manipulator to lower the slant. On the other hand, he would like to keep the slant high so that the report is good as often as possible. This creates a trade-off that shapes the optimal information policy.

The model identifies a spectrum of available information policies. At one extreme is a "mainstream" policy, which aims to spread information among and persuade both types of agents. It requires a relatively low slant because the report must be credible enough for both types of agents and incentives must be provided for L-type agents to pass a good report on to H-type agents. At the other extreme is a "partisan" policy, which aims to diffuse the information primarily among H-type agents and to persuade

only them. Hence, it can be achieved with a relatively high slant.

The presence of a spectrum of policies leads to the following question: what features of the environment make the manipulator choose a particular policy, for example a mainstream one or a partisan one? My analysis puts emphasis on two features of the environment: (i) the polarisation of the preferences of H-type and L-type agents, measured by the upward bias of the former, and (ii) the assortativity of meetings in the communication chain. These two features correspond to two important trends in modern societies: political polarisation, for example between Republicans and Democrats in the US, and homophily, which is a tendency of individuals to interact with those who are similar to them.

The analysis elucidates how polarisation and assortativity can make the partisan policy optimal for the manipulator. This occurs through three channels. First, as polarisation increases, L-type agents become more and more difficult to persuade relative to H-type agents, as the former require information of higher and higher credibility relative to the latter. Second, as polarisation and assortativity increase, providing incentives for L-type agents to pass the information on to H-type agents becomes more difficult, i.e. the slant required for such diffusion becomes lower. Finally, as assortativity increases, it becomes more likely that only H-type agents appear in the communication chain, so providing incentives for L-type agents to pass the information on to H-type agents has actually little effect on the diffusion of information.

I then explore two extensions of the model. In the first extension, I modify the communication between agents by assuming that each agent does not observe the type of her successor in the communication chain; however, she is still aware of the assortativity of meetings. This extension recognises that diffusion by social media such as Facebook or Twitter differs from conventional word of mouth; in particular, people can post information to their friends or followers but they do not know who will actually read the information and potentially pass it further on. I show that, under unobservable types of successors, higher assortativity makes maximal diffusion easier to induce, unlike under observable types of successors. This suggests that the growing role of social media makes media outlets less constrained by diffusion.

In the second extension, I assume that agents misestimate the slant chosen by the

manipulator: L-type agents overestimate it and H-type agents underestimate it. Thus, this extension studies the impact of the tendency of people to trust those media outlets which match their views and distrust those which do not. The analysis reveals that such misestimation of the slant increases the chances that the manipulator chooses a partisan policy.

The paper is organised as follows. The rest of this section discusses the empirical motivation for the model and the related literature. Section 2 presents the model. Section 3 characterises the equilibrium in the communication chain for a given information policy. Section 4 analyses the optimal information policy for the manipulator. Section 5 considers the extensions of the model. Section 6 concludes.

1.1 Empirical Motivation

Although the model is stylised, the main ingredients of the modelling approach are supported by empirical evidence. I discuss these key ingredients below.

The first main ingredient of the model is that *media slant is supply-driven*, as it is driven by the manipulator's internal incentives to influence the agents' actions. These incentives could arise directly from the preferences of media owners or indirectly from the preferences of editors or journalists. This contrasts with demand-driven media slant, in which case the driver of the slant is the demand from consumers (e.g., for news which confirm their views). Empirical evidence shows that both supply-side factors and demand-side factors influence media slant. Here, I briefly discuss papers which suggest that media slant is supply-driven. Larcinese, Puglisi and Snyder (2011) and Durante and Knight (2012) provide case studies (*Los Angeles Times* in the former and Italian TV station TG1 in the latter) where a change of the owner of a media outlet led to a rapid change in the news content of the outlet. Ansolabehere, Lessem and Snyder (2006) find a discrepancy between the slant of newspapers in the US in the 20th century and the political preferences of people living in the market areas of these newspapers, which suggests that demand-side factors cannot fully explain the observed slant. In a similar vein, Martin and Yurukoglu (2017) discover that the observed slant of Fox News

is much more pro-Republican than the viewership-maximising slant.[1]

The second main ingredient is that *individuals' beliefs and behaviour can be changed by what they see in the media*. Numerous studies have found evidence of media's influence on people's beliefs and behaviour. DellaVigna and Kaplan (2007) discover that availability of Fox News increased the Republican vote share in the 2000 US presidential elections. Gerber, Karlan and Bergan (2009) identify a positive effect of subscriptions to the Washington Post on the support for the Democratic candidate in the 2005 Virginia gubernatorial election. Enikopolov, Petrova and Zhuravskaya (2011) find that availability of an independent TV station, NTV, had a positive effect on the vote on opposition parties in the 1999 Russian parliamentary elections. Chiang and Knight (2011) use daily survey data on voting intentions before the 2000 and 2004 US presidential elections to find that people are more likely to support a candidate after a publication of a newspaper endorsement for that candidate. Martin and Yurukoglu (2017) exploit cable channel positions as exogenous shifters of their viewership to show that watching Fox News increases Republican vote shares.[2]

The third main ingredient is that *people detect and discount the media slant*. Evidence suggests that people do not blindly trust news reports, but they are aware of media slant and they attempt to filter it out. A survey by YouGov from June 2017 reports that 70% of Americans think that media outlets "tend to provide only one side of the story depending on who owns them or funds them".[3] Chiang and Knight (2011) find that endorsements for Democratic candidates from left-leaning newspapers are less effective than those from neutral or right-leaning newspapers, and vice versa for Republican candidates. Gentzkow, Shapiro and Sinkinson (2011) analyse a data set of US daily newspapers from 1869 to 2004 and find that the persuasive effect of partisan newspapers is limited, which they argue is consistent with people filtering partisan information.

[1]Papers which show that media slant can be driven by demand-side factors include Gentzkow and Shapiro (2010), Puglisi and Snyder (2011) and Larcinese, Puglisi and Snyder (2011). For a survey of empirical literature on the demand-side and supply-side factors of media slant, see Puglisi and Snyder (2015).

[2]For summaries of evidence on these effects, see DellaVigna and Gentzkow (2010), Prior (2013), and Puglisi and Snyder (2015).

[3]https://today.yougov.com/news/2017/06/20/Americans-agree-media-is-biased/

Finally, the fourth main ingredient of the model is that *information spreads through assortative but exogenous meetings*. There is large amount of evidence on assortativity in social networks, also referred to as homophily. McPherson, Smith-Lovin and Cook (2001) provide a survey of studies on homophily in a wide range of characteristics such as race, ethnicity, sex, gender, age, religion, education, occupation, social class, and others. When it comes to political views more specifically, a 2014 survey by Pew Research Center shows that the majority of people with consistently conservative or consistently liberal views say that most of their close friends share their views on government and politics.[4] Evidence also shows that the choice of whom to discuss politics with is exogenous, i.e. people do not consciously choose individuals to discuss politics, but they simply tend to discuss politics with the same people with whom they discuss other important matters in their lives (Klofstad, McClurg and Rolfe, 2009).

The first extension of the model is motivated by the growing role of social media as a source of news for people. A 2017 study by Pew Research Center shows that 67% of Americans now report that they get news from social media. Among social media sites, Facebook is the dominant leader, with 45% of Americans saying that they get news from it, while YouTube is second (18%) and Twitter is third (11%).[5] An important characteristic of social media, especially Facebook and Twitter, is that people often do not share information with a specific person (like in traditional word of mouth) but with their social network.

The motivation for the second extension is the observation that people with different ideological views tend to trust (and distrust) different sources of news; in particular, they trust more those sources which match their views and distrust those which do not. A 2014 report by Pew Research Center finds stark differences in the sources that Americans trust. For example, Fox News—generally considered as Republican-leaning—is trusted by 88% of Americans with consistently conservative views but is distrusted by 81% of those with consistently liberal views. When it comes to CNN, which is often viewed as having a liberal slant, 61% of consistent conservatives distrust

[4]Pew Research Center, October 2014, "Political Polarization and Media Habits".

[5]Pew Research Center, September 2017, "News Use Across Social Media Platforms 2017".

it, while 56% of consistent liberals trust it.[6]

1.2 Related Literature

My paper contributes to the research on the economics of media, in particular to the study of media slant (often also referred to as media bias). Empirical studies of media slant have analysed the measurement of media slant (e.g., Ansolabehere, Lessem and Snyder, 2006; Groseclose and Milyo, 2005; Gentzkow and Shapiro, 2010), the determinants of media slant (e.g., Gentzkow and Shapiro, 2010; Puglisi and Snyder, 2011; Larcinese, Puglisi and Snyder, 2011), and the effects of media slant on political behaviour (e.g., DellaVigna and Kaplan, 2007; Gerber, Karlan and Bergan, 2009; Chiang and Knight, 2011; Gentzkow, Shapiro and Sinkinson, 2011; Enikopolov, Petrova and Zhuravskaya, 2011). Puglisi and Snyder (2015) provide an excellent survey of the empirical literature on media slant. Theoretical studies have developed models of supply-driven media slant (e.g., Baron, 2006; Anderson and McLaren, 2012; Gehlbach and Sonin, 2014) and demand-driven media slant (e.g., Mullainathan and Shleifer, 2005; Gentzkow and Shapiro, 2006; Stone, 2011). An excellent survey of the theoretical literature is in Gentzkow, Shapiro and Stone (2015). My paper adds to the theoretical literature by studying the role of diffusion of information as a determinant of supply-driven media slant.

My paper contributes also to the economic study of strategic communication, in particular to the literatures on information design and information diffusion. It bridges these two strands by showing that, on the one hand, the design of an information structure can influence the diffusion of information and, on the other hand, that diffusion can be a factor that affects the optimal information structure for the designer.

The literature on information design has grown rapidly and has focused on Bayesian persuasion, following the work of Kamenica and Gentzkow (2011). To the best of my knowledge, my paper is the first that studies Bayesian persuasion followed by diffusion. My paper is related to those extensions of Kamenica and Gentzkow (2011) which take account of multiple receivers: Alonso and Câmara (2016), Taneva (2016), and Laclau

[6]Pew Research Center, October 2014, "Political Polarization and Media Habits".

and Renou (2017), but none of these papers feature strategic communication between the receivers. Like me, Gehlbach and Sonin (2014) apply the Bayesian persuasion framework to study information manipulation by media; however, in their paper the additional effect of the way the news are manipulated—beyond the effect on people's beliefs—is that it influences whether people decide to read the news, whereas in my paper the additional effect is that it influences how well the news spread by word of mouth.

The literature on information diffusion is very large. The most closely related papers to mine are those which study cheap-talk communication chains (Ambrus, Azevedo and Kamada, 2013; Anderlini, Gerardi and Lagunoff, 2012) and the diffusion of rumours (Bloch, Demange and Kranton, 2017). Apart from the communication protocol (cheap talk instead of verifiable disclosure), the communication chain in Ambrus, Azevedo and Kamada (2013) differs from mine in that it is finite and only the final receiver takes an action, while the chain in Anderlini, Gerardi and Lagunoff (2012) differs in that the incentive to communicate strategically is created by the fact that each agent in the chain prefers to take a lower action than his predecessors would like, whereas in my paper it is created by the heterogeneity of agents' preferences. Bloch, Demange and Kranton (2017) is similar to mine in that a rumour starts when one agent learns the true state of the world and it then spreads among biased and unbiased agents; however, the agents are arranged in an exogenous network and the focus of the paper is on identifying paths in the network along which rumours can diffuse. My paper adds to this literature, first, by analysing diffusion via a communication chain in a population of agents with heterogeneous preferences, and second, by investigating the impact of the information structure on that diffusion.

2 Model

Players. There is a manipulator, M, and an infinite population of agents, $N = \{1, 2, 3, ...\}$. The population consists of two types of agents with different preferences: high-type (H-type) agents and low-type (L-type) agents. The set of types of agents is denoted by $T = \{H, L\}$.

State of the world. There are two possible states of the world, $\omega \in \{0, 1\}$. The state of the world is ex ante unknown both to the manipulator and to the agents. From the perspective of the manipulator, $\omega = 1$ can be interpreted as a *good* state and $\omega = 0$ as a *bad* state. All players have a common prior belief that $\omega = 1$ with probability p.

Timeline. The game proceeds as follows:

1. The manipulator chooses an information policy π, which maps each state $\omega \in \{0, 1\}$ to a distribution over possible signal realisations $s \in \{0, 1\}$. From the perspective of the manipulator, the realisation $s = 1$ can be interpreted as a *good* news report and $s = 0$ as a *bad* news report. All agents observe the information policy π.

2. The state of the world $\omega \in \{0, 1\}$ is realised, with the signal realisation s determined according to π.

3. The information about s diffuses via a communication chain C.

    - One randomly selected H-type agent, say $i \in N$, observes s. She meets another agent (her successor), say $j \in N$, observes j's type, and decides whether to pass s on to j or not.
    - Upon receiving s, agent j meets yet another agent, say $k \in N$, observes k's type, and decides whether to pass s on to k or not, and so on.
    - The chain $C = \{i, j, k, \ldots\}$ continues until (i) an agent fails to meet a successor, which happens with exogenous probability $\epsilon \in (0, 1)$ to each agent, or (ii) an agent decides not to pass s on to her successor. If an agent decides not to pass s on to her successor, then the successor is left outside the chain.
    - Meetings are assortative, with $r \in [0, 1]$ being the measure of assortativity: an agent meets a successor of the same type with probability r and a successor of the other type with probability $1 - r$.[7]

[7] Thus, $r = 1$ describes a perfectly assortative population, $r = 0.5$ a non-assortative population, and $r = 0$ a perfectly disassortative population. In the rest of the paper, assortativity is referred to as lower or higher depending on the distance of r from 0.

- Each agent can appear only once in the chain.

4. Once the chain breaks, each agent in the population chooses an action.

The information policy π is effectively fully specified by the vector $(\pi(1 \mid 1), \pi(1 \mid 0))$, where $\pi(1 \mid 1)$ and $\pi(1 \mid 0)$ denote the probabilities that the signal realisation is $s = 1$ when the state is $\omega = 1$ and when the state is $\omega = 0$, respectively. Without loss of generality, I assume that $\pi(1 \mid 1) \geq \pi(1 \mid 0)$ so that the realisations $s = 1$ and $s = 0$ have intuitive meanings. The signal realisation is verifiable by the agents; hence, each agent can only either pass the signal realisation on or not, but cannot transform it.

Information sets. From the perspective of each agent, her own type is denoted by $t_0 \in T$ and her own action is denoted by $a_0 \in \mathbb{R}$. The type and action of an agent's successor are denoted by t_1 and a_1. The type and action of the successor of the agent's successor are denoted by t_2 and a_2, and so on.

Each agent knows her own type, t_0. An agent's own information set, I_0, describes the agent's additional information beyond her own type. Each agent in the communication chain observes $I_0 = \{s, t_1\}$ (conditional on meeting a successor). Since the content of communication is verifiable, once an agent receives s from her predecessor, it is irrelevant whether the predecessor's type is observable. The information set is empty, i.e. $I_0 = \{\}$, for all agents outside the chain.

Beliefs. An agent's belief that $\omega = 1$ upon observing information set I_0 is denoted by $\beta(I_0)$. For agents in the communication chain, I simply write $\beta(s)$ to indicate the belief about $\omega = 1$ upon observing $I_0 = \{s, t_1\}$, as t_1 is irrelevant for the belief. Any agent outside the communication chain is assumed to keep her prior belief, i.e. $\beta(\{\}) = p$. Therefore, if an agent does not pass s on to her successor, then the successor as well as all remaining agents in the population keep their prior belief.[8]

[8]Given that not passing on is endogenous, an agent may form a belief $\beta(\{\}) \neq p$. However, since the population is infinite and the probability of exogenous breakdown of the chain is strictly positive at each meeting, the probability of observing $\{\}$ tends to 1 regardless of the signal realisation and the equilibrium strategies. Therefore, $\beta(\{\}) \to p$ in any equilibrium.

Strategies. The manipulator's strategy is his information policy, effectively given by a vector $(\pi(1 \mid 1), \pi(1 \mid 0))$. Each agent has a communication strategy (conditional on being in the communication chain) and an action strategy. I will consider equilibria such that all agents of the same type have the same strategies. Thus, the communication strategy of an agent of type $t_0 \in \{L, H\}$ is a function $\mu_{t_0} : \{s, t_1\} \to \{s, \emptyset\}$, where $\mu_{t_0}(s, t_1) = s$ denotes passing s on to the successor of type t_1 and $\mu_{t_0}(s, t_1) = \emptyset$ denotes not passing s on to her. The action strategy of an agent of type $t_0 \in \{L, H\}$ is a function $\sigma_{t_0} : \{s\} \to \mathbb{R}$ for agents in the communication chain, and $\sigma_{t_0} : \{\} \to \mathbb{R}$ for agents outside the chain.

Manipulator's payoffs. The manipulator obtains a payoff from the actions of agents in the communication chain. Fix an infinite sequence of actions $\mathbf{a} = \{a(1), a(2), a(3), ...\}$, where $a(i)$ denotes the action of the agent in ith position in a communication chain of infinite length.

The payoff to the manipulator from action $a(i)$ (where $i = 1, 2, 3, ...$) is given by a threshold function $v(a(i))$:

$$v(a(i)) = \begin{cases} 1 & \text{if } a(i) \geq \bar{a} \\ 0 & \text{if } a(i) < \bar{a} \end{cases}, \tag{1}$$

where $\overline{a}$ is an exogenous threshold. If an agent takes an action weakly above $\overline{a}$, then we say that she is *persuaded*.

The manipulator's payoff function is assumed to be separately additive for agents' actions. The payoff to the manipulator from a fixed infinite sequence of actions $\mathbf{a}$ is written as

$$V(\mathbf{a}) = \sum_{i=1}^{\infty} \delta^{i-1} v(a(i)), \tag{2}$$

where $\delta = 1 - \epsilon$ is the discount factor that reflects the fact that the communication chain breaks with probability ϵ at each meeting.

When choosing his information policy $(\pi(1 \mid 1), \pi(1 \mid 0))$, the manipulator maximises the expected value of $V(\mathbf{a})$ given the strategies of agents.

Agents' payoffs. Each agent obtains a payoff from her own action, a_0, and from the actions of her successors in the communication chain, $a_1, a_2, a_3, ...$[9] Fix an infinite sequence of actions $\mathbf{a_0} = \{a_0, a_1, a_2, ...\}$, where a_i denotes the action of the agent's ith successor in a communication chain of infinite length.

I assume that agents have the same preferences regarding their own action and regarding the actions of their successors. The payoff to an agent of type t_0 from action a_i (where $i = 0, 1, 2, ...$) in state ω is expressed as $u_{t_0}(a_i, \omega)$. I assume the following functional forms: $u_L(a_i, \omega) = -(a_i - \omega)^2$ for an L-type and $u_H(a_i, \omega) = -(a_i - (\omega + b))^2$ for an H-type agent, where $b > 0$ is the bias of H-type agents. Thus, b measures the polarisation of H-type and L-type agents' preferences.

Each agent's payoff function is assumed to be separately additive for his own action and the actions of his successors. The payoff to an agent of type t_0 from a fixed infinite sequence of actions $\mathbf{a_0} = \{a_0, a_1, a_2, ...\}$ in state ω (conditional on her meeting her successor) is

$$U_{t_0}(\mathbf{a_0}, \omega) = u_{t_0}(a_0, \omega) + \sum_{i=1}^{\infty} \delta^{i-1} u_{t_0}(a_i, \omega), \tag{3}$$

where $\delta = 1 - \epsilon$ is the discount factor.

At the time of deciding whether to pass s on and what action to take, the agent maximises the expected value of $U_{t_0}(\mathbf{a_0}, \omega)$ given her information set, I_0, and given the agents' strategies. However, note that when choosing her action, a_0, each agent in fact maximises only the value of $u_{t_0}(a_0, \omega)$ given her information set, I_0, because her action cannot affect other agents' actions.[10]

[9] In principle, each agent could also receive a payoff from the actions of the predecessors and of the agents outside the communication chain but, in this setup, she cannot affect these actions, and so—as far as payoff maximisation is concerned—these actions can be neglected in her payoff function.

[10] The setup makes some simplifying assumptions. Allowing each agent to meet multiple agents would change the formulation of the agents' payoff function but each agent's decision whether to pass on or not would still be driven by the forces highlighted in this model, in particular by the extent to which the successors' preferences are aligned with hers (which in turn is determined by assortativity and polarisation). Receiving multiple messages would not make the agents' inference problem more complex because the content of messages is verifiable. A cheap-talk setup would allow the agents to costlessly transform the signal that they pass on; however, the usefulness of such transformation for an agent would be constrained by the fact that receivers would be aware of her incentives to transform.

3 Equilibrium in the Communication Chain

In this section, I analyse the equilibrium in the communication chain for a given information policy $(\pi(1 \mid 1), \pi(1 \mid 0))$. Therefore, the probabilities that the realisation of the signal is $s = 1$ when the state is $\omega = 1$ and when the state is $\omega = 0$ are taken as given throughout this section.

The solution concept is the perfect Bayesian equilibrium (henceforth, simply referred to as an equilibrium), which is defined in the usual way. An equilibrium consists of communication strategies, action strategies and beliefs, (μ, σ, β), where $\mu = (\mu_L, \mu_H)$ and $\sigma = (\sigma_L, \sigma_H)$, such that each agent's strategy is sequentially rational given the strategies and beliefs of other agents, and beliefs are derived by Bayes' rule from the strategies whenever it is possible.

Proposition 1 describes the strategies of the agents in the communication chain in the unique equilibrium.[11] For the agents outside the communication chain, the belief is $\beta(\{\}) = p$, and hence their action strategy is trivially $\sigma_L(\{\}) = p$ and $\sigma_H(\{\}) = p + b$.

Proposition 1. *In the unique equilibrium:*
(i) the agents' beliefs are

$$\beta(s) = \begin{cases} \frac{p\pi(0|1)}{p\pi(0|1)+(1-p)\pi(0|0)} & \text{for } s = 0 \\ \frac{p\pi(1|1)}{p\pi(1|1)+(1-p)\pi(1|0)} & \text{for } s = 1 \end{cases},$$

(ii) the agents' action strategies are

$$\sigma_{t_0}(s) = \begin{cases} \beta(s) & \text{for } t_0 = L \\ \beta(s) + b & \text{for } t_0 = H \end{cases} \text{ for } s \in \{0, 1\},$$

[11] I assume that if an agent is indifferent between passing s on and not, then she chooses to pass it on.

(iii) the agents' communication strategies are

$$\mu_L(s,t_1) = \begin{cases} s & \text{for } s \in \{0,1\} \text{ and } t_1 = L \\ s & \text{for } s = 0 \text{ and } t_1 = H \\ \begin{cases} s & \text{if and only if } \beta(1) \geq \beta(1)^* \\ \varnothing & \text{if and only if } \beta(1) < \beta(1)^* \end{cases} & \text{for } s = 1 \text{ and } t_1 = H \end{cases}$$

where $\beta(1)^* = p + 2b\frac{1-\delta r}{1+\delta-2\delta r}$, *and*

$$\mu_H(s,t_1) = \begin{cases} s & \text{for } s \in \{0,1\} \text{ and } t_1 = H \\ s & \text{for } s = 1 \text{ and } t_1 = H \\ \begin{cases} s & \text{if and only if } \beta(0) \leq \beta(0)^* \\ \varnothing & \text{if and only if } \beta(0) > \beta(0)^* \end{cases} & \text{for } s = 0 \text{ and } t_1 = L \end{cases}$$

where $\beta(0)^* = p - 2b\frac{1-\delta r}{1+\delta-2\delta r}$.

The agents form their beliefs through Bayesian updating. Thus, given an information policy $(\pi(1 \mid 1), \pi(1 \mid 0))$, if an agent observes a signal realisation $s = 0$, then she forms a posterior belief

$$\beta(0) = \Pr(\omega = 1 \mid s = 0) = \frac{p\pi(0 \mid 1)}{p\pi(0 \mid 1) + (1-p)\pi(0 \mid 0)}, \tag{4}$$

and if she observes a signal realisation $s = 1$, then she forms a posterior belief

$$\beta(1) = \Pr(\omega = 1 \mid s = 1) = \frac{p\pi(1 \mid 1)}{p\pi(1 \mid 1) + (1-p)\pi(1 \mid 0)}. \tag{5}$$

The equilibrium action strategies are such that an L-type agent takes an action that equals his posterior belief, while an H-type agent takes an action that equals his posterior belief plus her bias. Therefore, the information policy—by influencing the agents' posterior beliefs—also influences their actions, with H-type agents taking higher actions than L-type agents for a given belief. More formally, the equilibrium action strategies are $\sigma_L(s) = \text{argmax}_{a_0} u_L(a_0, \omega \mid \beta(s)) = \beta(s)$ and $\sigma_H(s) = \text{argmax}_{a_0} u_H(a_0, \omega \mid \beta(s)) =$

$\beta(s) + b$.

The communication strategy is such that an L-type agent passes s on to her successor under any information policy, except when $s = 1$ and her successor is of type $t_1 = H$: in that case she passes it on if and only if $\beta(1)$ is high enough or—put differently—if and only if $\pi(1 \mid 1)$ is high enough and $\pi(1 \mid 0)$ is low enough. Conversely, an H-type agent passes s on to her successor under any information policy, except when $s = 0$ and her successor is of type $t_1 = L$: in that case she passes it on if and only if $\beta(0)$ is low enough, i.e. if and only if $\pi(0 \mid 0)$ is high enough and $\pi(0 \mid 1)$ is low enough. Therefore, an information policy that is sufficiently informative about $\omega = 1$ ($\omega = 0$) can improve the diffusion of $s = 1$ ($s = 0$) by inducing L-type agents to pass $s = 1$ on to H-type agents (H-type agents to pass $s = 0$ on to L-type agents).

The reasoning behind the communication strategy is as follows. Consider an L-type agent (an analogous logic applies to an H-type agent). She clearly has no incentive to suppress $s = 0$ from any agent. If she suppresses it, then all remaining agents take an action based on the prior belief, but if she passes it on, then she can only be better off because her successor—and potentially further successors—take lower actions, which moves them closer to her optimal action for $s = 0$. An L-type agent also has no incentive to suppress $s = 1$ from an L-type successor because their preferences are perfectly aligned. Finally, if an L-type agent receives $s = 1$ and meets an H-type agent, then her communication depends on the information policy, given by $\pi(1 \mid 1)$ and $\pi(1 \mid 0)$, which determine the belief $\beta(1)$. More precisely, she passes $s = 1$ on if and only if $\beta(1)$, i.e. if and only if $s = 1$ is sufficiently informative about $\omega = 1$. Figure 1 provides an illustration of how a change in $\beta(1)$ (and $\beta(0)$) affects the incentives of an L-type agent to pass $s = 1$ (and $s = 0$) on to an H-type successor.

For simplicity, Figure 1 shows only the immediate H-type successor, but L-type agent naturally takes into account also the actions of further successors in the chain. In Panel A, $\beta(1)$ is high, which means that $s = 1$ is relatively informative about $\omega = 1$ and so the optimal actions for $s = 1$ are relatively high for both types of agents—and far away from their optimal actions for the prior belief. Consequently, the L-type agent prefers passing $s = 1$ on to the H-type agent to not passing it on. On the other hand, in Panel B, a lower $\beta(1)$ means that $s = 1$ becomes less informative about $\omega = 1$, which

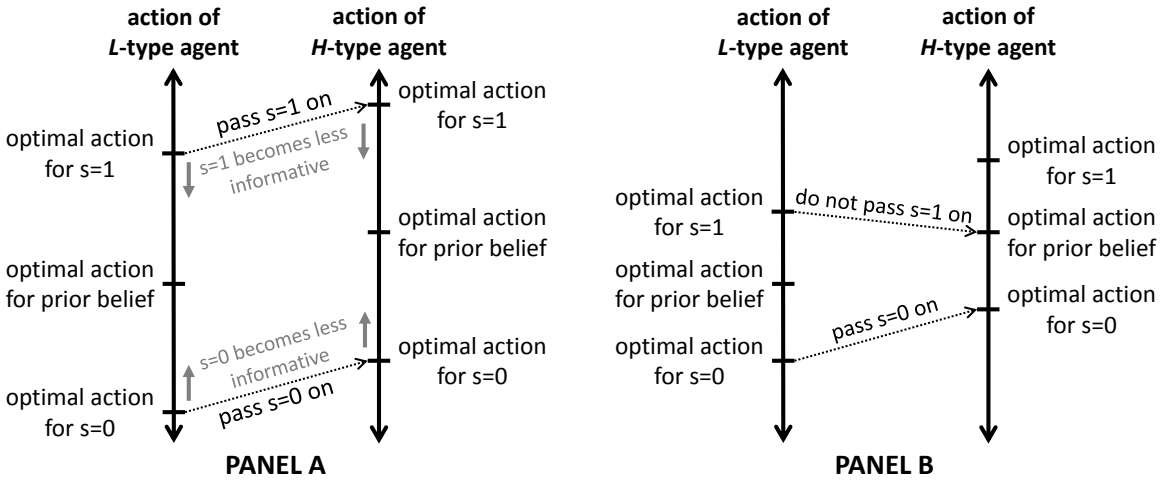

Figure 1: Graphical illustration of how a change in $\beta(1)$ and $\beta(0)$ affects the incentives of an L-type agent to pass $s = 1$ and $s = 0$ on to an H-type agent.

moves the optimal actions for $s = 1$ closer to the optimal actions for the prior belief. Then, the L-type agent is better off by not passing $s = 1$ on to the H-type agent. In addition, we can see that $\beta(0)$ increases as we move from Panel A to Panel B, which means that $s = 0$ becomes less informative about $\omega = 0$. However, this does not affect the incentive of the L-type agent to pass $s = 0$ on to the H-type agent.

Overall, the main message behind Proposition 1 is that the information policy has two effects: (1) it influences the beliefs of agents, and hence also their actions, upon receiving $s = 0$ and $s = 1$, and (2) it influences the agents' incentives to pass $s = 0$ and $s = 1$ on to their successors, and thus affects the diffusion in the communication chain. These two effects will play an important role in the analysis of the manipulator's optimal information policy in Section 4.

4 Optimal Information Policy

In this section, I analyse the manipulator's optimal information policy, which is the policy that maximises his expected payoff.

4.1 Preliminaries

I start the analysis by making the following assumption.

Assumption 1. *The prior belief p is such that, for all agents, the optimal action of an agent with a belief p is below $\bar{a}$, i.e. $argmax_{a_0} u_{t_0}(a_0, \omega \mid p) < \bar{a}$ for $t_0 \in \{L, H\}$. This assumption is satisfied if and only if $p < \bar{a} - b$.*

This assumption means that all agents are ex ante unpersuaded.[12] Therefore, the manipulator receives a payoff of 0 from an action of any agent whose posterior belief is the same as the prior. To see why this assumption is satisfied if and only if $p < \bar{a} - b$, note that the optimal actions for agents with a belief p are given by $\text{argmax}_{a_0} u_H(a_0, \omega \mid p) = p + b$ for an H-type agent and by $\text{argmax}_{a_0} u_L(a_0, \omega \mid p) = p$ for an L-type agent.

It is important to note that, in the optimal information policy, $\pi(1 \mid 1) = 1$ must hold, i.e. whenever the state of the world is $\omega = 1$, the signal realisation must be $s = 1$. In other words, whenever the state is *good*, the news report must also be *good*. The manipulator has an incentive to increase $\pi(1 \mid 1)$ as much as possible because it increases the probability that the signal realisation is $s = 1$, increases the agents' actions upon observing $s = 1$, and can only enhance the agents' incentives to pass $s = 1$ on.

On the other hand, $\pi(1 \mid 0)$ may be weakly positive in the optimal information policy. Yet, $\pi(1 \mid 0) = 1$ cannot be optimal because, combined with $\pi(1 \mid 1) = 1$, it would make the signal uninformative and so—given Assumption 1—all agents would remain unpersuaded upon observing $s = 1$. The conditional probability $\pi(1 \mid 0) \in [0, 1)$ then fully specifies any optimal information policy.

Definition 1. The *slant* of the information policy is defined as $\pi(1 \mid 0)$, i.e. the probability that the signal realisation is $s = 1$ given that the state of the world is $\omega = 0$.

In the context of media, the slant is defined here as the tendency of a media outlet to publish a *good* (from the perspective of the outlet) news report when the state of the world is *bad* (from the perspective of the outlet). Thus, the slant is modelled here

[12] If all agents were ex ante persuaded, then the manipulator's optimal information policy would be to make the signal completely uninformative. If H-type agents were ex ante persuaded but L-type agents were not, then the manipulator would prefer to target an L-type agent as the first agent in the chain. Under Assumption 1, the manipulator prefers to target an H-type agent.

as lying in the most direct form, i.e. negating the truth. Naturally, this is not the only possible form of media slant. Gentzkow, Shapiro and Stone (2015) distinguish two categories of media slant: distortion of information and filtering of information. Put briefly, the former captures manipulation by reporting outright false information, whereas the latter captures manipulation by selective reporting of information and biased summarising of information. The definition of slant used in this paper falls under the category of distortion. Other papers modelling media slant as distortion include Mullainathan and Shleifer (2005), Baron (2006), and Gentzkow and Shapiro (2006).

4.2 Two Effects: on Persuasion and on Diffusion

As already mentioned in Section 3, the information policy has two effects: (i) it influences the agents' beliefs and actions upon receiving $s = 0$ and $s = 1$, and (ii) it influences whether the agents pass $s = 0$ and $s = 1$ on to their successors. In short, the information policy influences (i) persuasion and (ii) diffusion. I now analyse these effects in the context of the slant.

Effect of the information policy on persuasion. First, I consider the effect of the slant on whether the agents are persuaded.

Proposition 2. *In the optimal information policy:*
(i) an H-type agent is persuaded by $s = 1$ if and only if the slant is $\pi(1 \mid 0) \leq \pi^H$, where

$$\pi^H = \frac{p}{1-p}\frac{1-(\bar{a}-b)}{\bar{a}-b}, \tag{6}$$

(ii) an L-type agent is persuaded by $s = 1$ if and only if the slant is $\pi(1 \mid 0) \leq \pi^L$, where

$$\pi^L = \frac{p}{1-p}\frac{1-\bar{a}}{\bar{a}}. \tag{7}$$

The thresholds π^H and π^L are increasing in prior belief p and decreasing in $\bar{a}$, and π^H is increasing in polarisation b.

Proposition 2 uses the equilibrium action strategies described in Proposition 1 and the observation that $\pi(1 \mid 1) = 1$ must hold in the optimal policy. Given (4), (5), and $\pi(1 \mid 1) = 1$, the posterior belief about $\omega = 1$ upon receiving $s = 1$ is

$$\beta(1) = \frac{p}{p + (1-p)\,\pi(1 \mid 0)}, \tag{8}$$

and the posterior belief belief about $\omega = 1$ upon receiving $s = 0$ is $\beta(0) = 0$. Then, an H-type agent takes action above $\bar{a}$ if and only if $\beta(1) + b \geq \bar{a}$, which can be rearranged to $\pi(1 \mid 0) \leq \pi^H$, where π^H is given by (6). Similarly, an L-type agent takes action above $\bar{a}$ if and only if $\beta(1) \geq \bar{a}$, which can be rearranged to $\pi(1 \mid 0) \leq \pi^L$, where π^L is described by (7). It follows easily that $\pi^H > \pi^L$.

Hence, the message behind Proposition 2 is that the manipulator must choose a low enough slant in order to be able to persuade the agents. The thresholds π^H and π^L denote the maximum levels of slant for which $s = 1$ persuades an H-type and an L-type agent, respectively. The relation $\pi^H > \pi^L$ means that H-type agents are more easily persuaded than L-type agents.

Effect of the information policy on diffusion. Second, I consider the effect of the slant on the diffusion of information. Given Assumption 1, the signal realisation $s = 0$ cannot persuade any agents, so the manipulator is concerned only about the diffusion of $s = 1$. Therefore, I focus on the effect on the diffusion of $s = 1$.

Proposition 3. *In the optimal information policy:*
(i) an H-type agent passes $s = 1$ on to an H-type agent under any slant,
(ii) an H-type agent passes $s = 1$ on to an L-type agent under any slant,
(iii) an L-type agent passes $s = 1$ on to an L-type agent under any slant,
(iv) an L-type agent passes $s = 1$ on to an H-type agent if and only if the slant is $\pi(1 \mid 0) \leq \pi^D$, where

$$\pi^D = \frac{p}{1-p} \frac{1 - \left(p + 2b\frac{1-\delta r}{1+\delta-2\delta r}\right)}{p + 2b\frac{1-\delta r}{1+\delta-2\delta r}}. \tag{9}$$

The threshold π^D is decreasing in assortativity r, $\frac{\partial \pi^D}{\partial r} < 0$, and decreasing in polarisation

b, $\frac{\partial \pi^D}{\partial b} < 0$, *with* $\frac{\partial^2 \pi^D}{\partial r \partial b} < 0$ *for sufficiently low* r *and* b, *and* $\frac{\partial^2 \pi^D}{\partial r \partial b} \geq 0$ *otherwise.*

Proposition 3 uses the equilibrium communication strategies in Proposition 1 and the observation that $\pi(1 \mid 1) = 1$ in the optimal information policy. As stated in Proposition 1, an L-type agent passes $s = 1$ on to an H-type agent if and only if $\beta(1) \geq \beta(1)^*$. Given that $\pi(1 \mid 1) = 1$, this condition becomes $\pi(1 \mid 0) \leq \pi^D$, where π^D is given by (9).

The message behind Proposition 3 is therefore that, by decreasing the slant, the manipulator can facilitate the diffusion of $s = 1$ from L-type to H-type agents, and thus improve the diffusion of $s = 1$. Figure 2 illustrates how a change in slant $\pi(1 \mid 0)$ can affect an L-type agent's incentives to pass $s = 1$ on to an H-type agent.

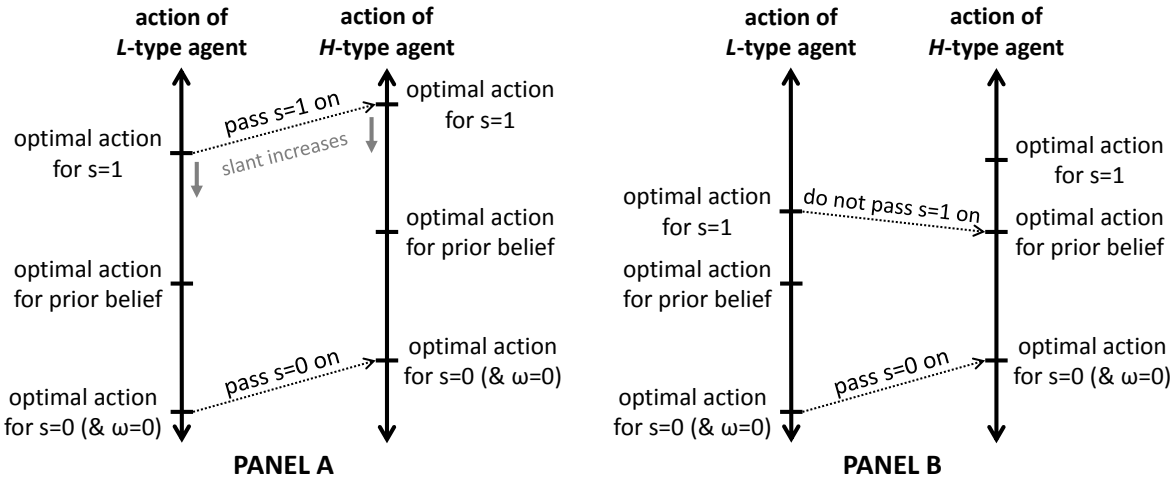

Figure 2: Graphical illustration of how a change in slant $\pi(1 \mid 0)$ can affect an L-type agent's incentives to pass $s = 1$ on to an H-type agent.

The slant is lower in Panel A than in Panel B. As the slant increases, the L-type agent's optimal action for $s = 1$ becomes relatively closer to the agents' optimal actions for the prior belief than to the H-type agent's optimal action for $s = 1$. Consequently, when the slant is high enough, the L-type agent no longer has an incentive to pass $s = 1$ on to the H-type agent. The threshold π^D denotes the maximum slant under which the L-type agent would pass $s = 1$ on to the H-type agent. The optimal actions for $s = 0$ are not affected by the slant; in fact, they coincide with the optimal actions

in state $\omega = 0$ because the agent's belief about $\omega = 1$ upon observing $s = 0$ must equal zero in the optimal information policy (which follows from $\pi(1 \mid 1) = 1$).

The threshold π^D is decreasing in polarisation and assortativity. In other words, as polarisation and assortativity increase, it becomes more difficult to facilitate diffusion of $s = 1$ from L-type to H-type agents. The intuition for the effect of polarisation is clear: as the preferences of the two types of agents become less aligned, L-type agents are less willing to share $s = 1$ with H-type agents. The intuition for the effect of assortativity is that, as assortativity increases, an L-type agent who meets an H-type agent realises that the H-type agent's successors are more likely to be H-type agents too. Thus, the preferences of the L-type agent and her successors in the chain are less likely to be aligned.

For sufficiently low assortativity and polarisation, $\frac{\partial^2 \pi^D}{\partial r \partial b} < 0$ holds, i.e. polarisation and assortativity reinforce each other's negative effect on the threshold π^D. However, when assortativity and polarisation are high, then $\frac{\partial^2 \pi^D}{\partial r \partial b} \geq 0$ holds, which means that polarisation and assortativity dampen each other's negative effect on π^D.

Relation between the two effects. The relation between the values of thresholds π^H, π^L and π^D depends on the parameters of the model. Since $\pi^H > \pi^L$ always holds, there are three possibilities: (i) $\pi^D < \pi^L$, (ii) $\pi^L \leq \pi^D < \pi^H$, and (iii) $\pi^D \geq \pi^H$. Proposition 4 describes how the relation between π^H, π^L and π^D depends on polarisation b and on assortativity r.

Proposition 4. *(a) As far as assortativity r is concerned, the relation between π^D, π^L and π^H is:*

(i) $\pi^D < \pi^L$ if and only if $r > r^{**}$,

(ii) $\pi^L \leq \pi^D < \pi^H$ if and only if $r^* < r \leq r^{**}$,

(iii) $\pi^D \geq \pi^H$ if and only if $r \leq r^*$,

where r^* and r^{**} are functions of other parameters, and $r^* < r^{**}$.

(b) As far as polarisation b is concerned, the relation between π^D, π^L and π^H is:

(i) $\pi^D < \pi^L$ if and only if $b > b^{**}$,

(ii) $\pi^L \leq \pi^D < \pi^H$ if and only if $b^* < b \leq b^{**}$,

(iii) $\pi^D \geq \pi^H$ if and only if $b \leq b^*$,

where b^* and b^{**} are functions of other parameters, and $0 < b^* < b^{**} < \overline{a} - p$.

Proposition 4 uses the thresholds π^H, π^L and π^D derived in Propositions 2 and 3. Part (a) follows from the fact that π^D is decreasing in r, while both π^L and π^H do not depend on r. Part (b) follows from the fact that π^D is decreasing in b, π^L does not depend on b, and π^H is increasing in b. It is worth noting that r^* and r^{**} take values in $[0, 1]$ only under some conditions for other parameters. On the other hand, b^* and b^{**} take values in $(0, \overline{a} - p)$ regardless of the values of other parameters.

Overall, Proposition 4 tells us that polarisation and assortativity of the population determine how difficult it is for the manipulator to induce the diffusion of $s = 1$ from L-type agents to H-type agents relative to persuading L-type and H-type agents. High polarisation and high assortativity both contribute to this diffusion being relatively more difficult to induce.

4.3 Characterisation of the Optimal Information Policy

I now characterise the optimal information policy under the following assumption.

Assumption 2. *The values of parameters of the model are such that $\pi^D < \pi^L$ holds.*

The objective of this analysis is to illuminate the role of diffusion for the optimal information policy. Thus, it makes sense to assume that $\pi^D < \pi^L$. As π^D increases above π^L, the impact of diffusion on the optimal information policy diminishes. Ultimately, when $\pi^D > \pi^H$, the diffusion of $s = 1$ from L-type to H-type agents is so easy to induce that even persuading H-type agents requires a slant such that the diffusion is induced anyway. Therefore, I focus here on $\pi^D < \pi^L$ and briefly discuss the cases $\pi^L \leq \pi^D < \pi^H$ and $\pi^D \geq \pi^H$ at the end of the section.

The relation $\pi^D < \pi^L$ has implications for persuasion and diffusion under various levels of slant. These implications are summarised in the following table.

Slant of the information policy, $\pi(1 \mid 0)$	Are H-type agents persuaded by $s = 1$?	Are L-type agents persuaded by $s = 1$?	Do L-type agents pass $s = 1$ on to H-type agents?
$[0, \pi^D]$	yes	yes	yes
$(\pi^D, \pi^L]$	yes	yes	no
$(\pi^L, \pi^H]$	yes	no	no
$(\pi^H, 1]$	no	no	no

Table 1: Persuasion and diffusion under various levels of slant in the case $\pi^D < \pi^L$.

Naturally, the slant in the optimal information policy cannot be higher than π^H, as then no agents could be persuaded. Therefore, there are three possible persuasion and diffusion patterns that can be induced in the optimal information policy, which are listed below. These patterns correspond to a spectrum of possible information policies, which I refer to as "mainstream", "intermediate", and "partisan". They differ in the expected payoff to the manipulator conditional on $s = 1$, which is given by $\mathbf{E}[V(\mathbf{a}) \mid s = 1]$. The expressions for $\mathbf{E}[V(\mathbf{a}) \mid s = 1]$ in the three different policies are denoted by V^D, V^L, and V^H, respectively.

1. *For $\pi(1 \mid 0) \in [0, \pi^D]$, both types of agents are persuaded by $s = 1$, and $s = 1$ is always passed on by agents.* This pattern corresponds to a "mainstream" information policy, which aims to persuade and spread the information across both types. The expected payoff to the manipulator conditional on $s = 1$ is

$$V^D = \sum_{i=1}^{\infty} \delta^{i-1} = 1 + \frac{\delta}{1-\delta}, \tag{10}$$

   which follows from the fact that diffusion can only stop exogenously and all agents in the communication chain are persuaded.

2. *For $\pi(1 \mid 0) \in (\pi^D, \pi^L]$, both types of agents are persuaded by $s = 1$, and $s = 1$ is passed on by agents, except from L-type agents to H-type agents.* This pattern is an "intermediate" information policy, which aims to persuade both types of

agents but is not concerned about spreading the information across both types. The expected payoff to the manipulator conditional on $s = 1$ is

$$V^L = \sum_{i=1}^{\infty} \delta^{i-1} \left(r^{i-1} + (i-1) r^{i-2} (1-r) \right) = 1 + \frac{\delta (1 - \delta r^2)}{(1 - \delta r)^2}, \tag{11}$$

which follows from the fact that the diffusion of $s = 1$ stops as soon an L-type agent meets an H-type agent, but all agents are persuaded along the way.

3. *For $\pi(1 \mid 0) \in (\pi^L, \pi^H]$, only H-type agents are persuaded by $s = 1$, and $s = 1$ is passed on by agents, except from L-type agents to H-type agents.* This pattern corresponds to a "partisan" information policy, whose objective is to persuade H-type agents only and spread the information primarily among them. The expected payoff to the manipulator conditional on $s = 1$ is

$$V^H = \sum_{i=1}^{\infty} \delta^{i-1} r^{i-1} = 1 + \frac{\delta r}{1 - \delta r}, \tag{12}$$

which follows from the fact that agents pass $s = 1$ on and are persuaded by it as long as only H-types agents appear in the communication chain.

It is straightforward to see that the mainstream information policy gives the highest expected payoff to the manipulator conditional on $s = 1$, while the partisan information policy gives the lowest one, for any $r < 1$. In a perfectly assortative population, i.e. with $r = 1$, the expected payoff to the manipulator conditional on $s = 1$ is the same in all three patterns and equals the one guaranteed by the mainstream information policy, i.e. $V^H = V^L = V^D = 1 + \frac{\delta}{1-\delta}$.

The manipulator's optimal information policy is the one that maximises his expected payoff. His expected payoff from an information policy with a slant $\pi(1 \mid 0)$ is given by $\Pr(s = 1)\mathbf{E}[V(\mathbf{a}) \mid s = 1]$, where both $\Pr(s = 1)$ and $\mathbf{E}[V(\mathbf{a}) \mid s = 1]$ depend on $\pi(1 \mid 0)$. The term $\Pr(s = 1)$ is clearly increasing in $\pi(1 \mid 0)$, as $\Pr(s = 1) = p + (1 - p)\pi(1 \mid 0)$. The term $\mathbf{E}[V(\mathbf{a}) \mid s = 1]$ is weakly decreasing in $\pi(1 \mid 0)$, as a higher slant weakly worsens the persuasion and diffusion pattern.

It is important to note that the optimal slant can only be equal to π^D, π^L, or π^H:

if the slant does not take one of these values, then the manipulator can increase the slant without affecting the persuasion and diffusion pattern, which necessarily increases his expected payoff. Thus, the choice of the manipulator is effectively between (i) π^D (which corresponds to the mainstream policy), (ii) π^L (the intermediate policy), and (iii) π^H (the partisan policy).

Proposition 5 characterises the optimal information policy in the context of two key characteristics of the environment: assortativity r and polarisation b.[13]

Proposition 5. *(a) As far as assortativity r is concerned, the optimal information policy for the manipulator is:*

(i) the mainstream information policy if and only if $r \leq \min\{r^{DL}, r^{DH}\}$,

(ii) the intermediate information policy if and only if $r > r^{DL}$ and $r \leq r^{LH}$,

(iii) the partisan information policy if and only if $r > \max\{r^{DH}, r^{LH}\}$,

where r^{DL}, r^{DH} and r^{LH} are functions of other parameters of the model, with $r^{LH} < r^{DH} < r^{DL}$, $r^{LH} = r^{DH} = r^{DL}$, and $r^{DL} < r^{DH} < r^{LH}$ being the only possible relations. For all parameter values, $\max\{r^{DH}, r^{LH}\} < 1$ holds.

(b) As far as polarisation b is concerned, the optimal information policy for the manipulator is:

(i) the mainstream information policy if and only if $b \leq \min\{b^{DL}, b^{DH}\}$,

(ii) the intermediate information policy if and only if $b > b^{DL}$ and $b \leq b^{LH}$,

(iii) the partisan information policy if and only if $b > \max\{b^{DH}, b^{LH}\}$,

where b^{DL}, b^{DH} and b^{LH} are functions of other parameters of the model, with $b^{LH} < b^{DH} < b^{DL}$, $b^{LH} = b^{DH} = b^{DL}$, and $b^{DL} < b^{DH} < b^{LH}$ being the only possible relations.

The first immediate observation from Proposition 5 is that the possibility of diffusion weakly decreases the slant. In a setup without diffusion, i.e. where the signal realisation is observed by an H-type agent but cannot diffuse further, the optimal information policy would have a slant π^H, whereas here lower slants of π^L and π^D can be optimal too.

[13] I assume that if the manipulator is indifferent between information policies, then he chooses the one with a lower slant.

The mainstream information policy, i.e. slant π^D, is optimal for the manipulator in a population with low polarisation and low assortativity. Several effects contribute to this. First, low polarisation means that L-type agents can be persuaded almost as easily as H-type agents. Second, low polarisation and low assortativity imply that it is easy for the manipulator to induce L-type agents to pass $s = 1$ on to H-type agents. Third, due to low assortativity, there is likely to be a mix of both types of agents in the chain, so inducing L-type agents to pass $s = 1$ on to H-type agents can significantly improve the extent of diffusion.

At the other extreme, in a highly polarised and highly assortative population, the optimal choice is the partisan information policy, i.e. slant π^H. High polarisation means that H-type agents are much easier to persuade than L-type agents. Furthermore, both high polarisation and high assortativity make it difficult for the manipulator to induce L-type agents to pass $s = 1$ on to H-type agents. Finally, high assortativity means that it is likely that there are only H-type agents in the communication chain, and hence a lack of transmission of $s = 1$ from L-type to H-type agents has little influence on the extent of diffusion. For any given values of other parameters, once assortativity becomes high enough, then the partisan policy must be optimal. Eventually, in a perfectly assortative population ($r = 1$), the partisan policy is optimal for all possible values of other parameters, as only H-type agents appear in the communication chain.

The intermediate information policy, i.e. slant π^L, is optimal when polarisation and assortativity are at a moderate level. More specifically, the following conditions must be satisfied: (i) polarisation must be low enough to make L-type agents comparably easy to persuade to H-type agents, (ii) polarisation and assortativity must be high enough to make it difficult to induce diffusion of $s = 1$ from L-type to H-type agents, and (iii) assortativity must be high enough to ensure that not much is lost by not inducing such diffusion but must be low enough to ensure that it makes sense to aim to persuade L-type agents. It is important to note that, under some conditions, the intermediate policy is never optimal. This happens if the relation between r^{DL}, r^{DH} and r^{LH} is $r^{LH} < r^{DH} < r^{DL}$, in which case there are no values of r that satisfy both $r > r^{DL}$ and $r \leq r^{LH}$.

For illustration, Figure 3 shows the optimal information policy in the (r, b) parameter

space for values of r and b that satisfy $\pi^D < \pi^L$, with fixed values of other parameters.[14]

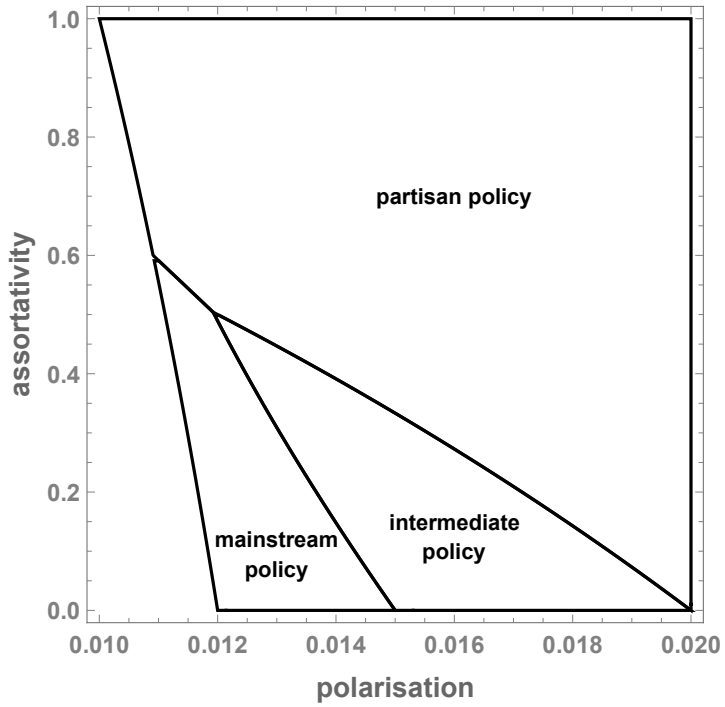

Figure 3: Regions of parameter values under which the mainstream, the intermediate and the partisan policies are the optimal information policies for the manipulator, illustrated in the (r, b) parameter space for values of r and b that satisfy $\pi^D < \pi^L$, with other parameters fixed at $\delta = 0.2$, $p = 0.1$, $\overline{a} = 0.12$.

We can now make an observation about how the expected diffusion of $s = 1$ depends on assortativity and polarisation. The expected diffusion of $s = 1$ is measured by the expected length of the communication chain conditional on $s = 1$.

Corollary 1. *(a) The expected diffusion of $s = 1$ is non-monotonic with respect to assortativity r: it is constant in r for $r < \min\{r^{DL}, r^{DH}\}$, decreases discontinuously at $r = \min\{r^{DL}, r^{DH}\}$, and is increasing in r for $r > \min\{r^{DL}, r^{DH}\}$.*

[14]The white region in the bottom-left corner corresponds to values of r and b that do not satisfy $\pi^D < \pi^L$. Other parameters are fixed at $\delta = 0.2$, $p = 0.1$, $\overline{a} = 0.12$. These values are calibrated so that all three policies (mainstream, intermediate and partisan) are optimal for some values of r and b.

(b) The expected diffusion of $s = 1$ is weakly decreasing in polarisation b: it is constant in b for $b < \min\{b^{DL}, b^{DH}\}$, decreases discontinuously at $b = \min\{b^{DL}, b^{DH}\}$, and is constant in b for $b > \min\{b^{DL}, b^{DH}\}$.

The result in Corollary 1 follows in a straightforward way from Proposition 5. Whenever the mainstream policy is chosen by the manipulator, the expected diffusion of $s = 1$ is equal to V^D, as the diffusion can only be stopped exogenously. Whenever the intermediate or partisan policy is chosen, it is equal to V^L because the diffusion stops as soon as an L-type agent meets an H-type agent. The result follows from the fact that V^D is constant in r and b, while V^L is increasing in r and constant in b, and $V^D > V^L$ always holds (unless $r = 1$, in which case $V^D = V^L$). As r approaches 1, the expected diffusion of $s = 1$ approaches V^D, since $r \to 1$ implies that there are only H-type agents in the chain.

For illustration, Figure 4 shows the expected diffusion of $s = 1$ as a function of assortativity for fixed values of other parameters.[15]

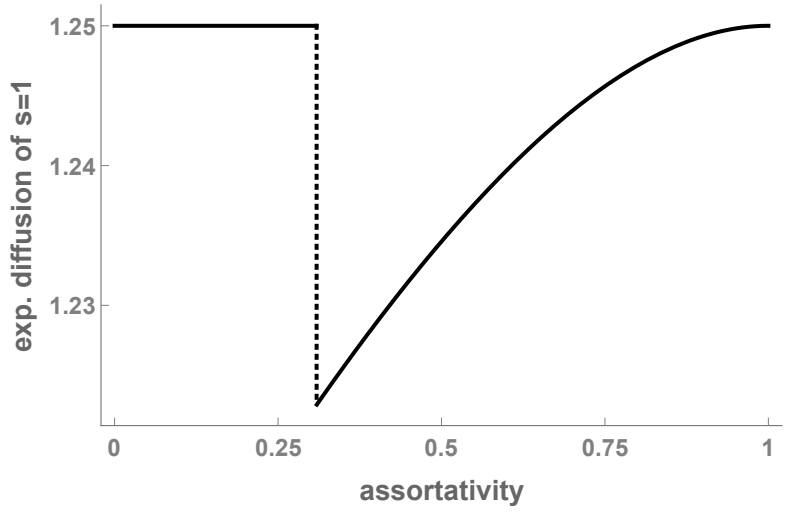

Figure 4: The non-monotonic relation between the expected diffusion of $s = 1$ and the assortativity of the population, with other parameters fixed at $\delta = 0.2$, $p = 0.1$, $\overline{a} = 0.12$, $b = 0.013$.

[15]Other parameters are fixed at $\delta = 0.2$, $p = 0.1$, $\overline{a} = 0.12$ (same as in Figure 3), and $b = 0.013$.

I close this subsection with a brief discussion of the cases $\pi^L \leq \pi^D < \pi^H$ and $\pi^D \geq \pi^H$. By and large, the insights that these two cases provide are similar to those in the case of $\pi^D < \pi^L$, with the role of diffusion being less prominent.

The available information policies are a little different than those in the case $\pi^D < \pi^L$. Under $\pi^L \leq \pi^D < \pi^H$, the slants π^L and π^H correspond respectively to the mainstream policy and the partisan policy, which are defined the same way as earlier. However, the slant π^D corresponds to a new type of intermediate policy: $s = 1$ are always passed by both types of agents but only H-type agents are persuaded by it.[16] Thus, we could say that this intermediate policy is diffusion-oriented, whereas the previous one was persuasion-oriented. Under $\pi^D \geq \pi^H$, the manipulator's choice is effectively only between the slants π^L (which corresponds to the mainstream policy) and π^H (the diffusion-oriented intermediate policy), both of which induce L-type agents to pass $s = 1$ on to H-type agents. Hence, under $\pi^D \geq \pi^H$, the possibility of improving diffusion plays no role for the optimal information policy.

Despite the difference in the available policies, the channels through which polarisation and assortativity influence the optimal policy work in a similar manner as in the case $\pi^D < \pi^L$. Consider assortativity, for example. Low assortativity favours especially the mainstream policy because there is likely to be a mix of both types of agents in the chain, so the manipulator cares about persuading and spreading information among both types of agents. High assortativity makes the partisan policy particularly attractive because it is likely that only H-type agents appear in the chain, so there is no need to persuade L-type agents and to make sure that information is passed from L-type to H-type agents. The new intermediate policy is affected by increased assortativity in two opposite ways: the required slant π^D decreases, which makes this policy less attractive to the manipulator, but on the other hand, it becomes more likely that only H-type agents appear in the chain, which makes the policy more attractive since not much is lost by not persuading L-type agents.

[16]The formulation of the expected payoff from this intermediate policy is more complicated because persuasion of n-th agent in the chain depends on whether that agent is an H-type agent; thus, for each n, one needs to consider all possible combinations of same-type and opposite-type meetings that make the n-th agent an H-type agent.

4.4 Graphical Interpretation

In this subsection, I provide a graphical interpretation of the result on the optimal information policy by using the concavification method (Kamenica and Gentzkow, 2011). Again, I consider the case of $\pi^D < \pi^L$, but the method can be applied to other cases equally well. The concavification method relies on two observations by Kamenica and Gentzkow (2011), which I discuss below in the context of my model.

The first observation by Kamenica and Gentzkow (2011) is that, in their setting, the manipulator's payoff is fully determined by the posterior induced by the signal realisation. This is also true in my setting in the following sense: the manipulator's interim (i.e. after the realisation of the signal but before the length of the communication chain is known) expected payoff is fully determined by the posterior induced by the signal realisation on the first agent in the communication chain. The reasoning behind this statement is as follows.

Let the manipulator's interim expected payoff, given posterior belief β, be denoted by $E_\beta\left[V\left(\mathbf{a}^*(\beta)\right)\right]$ with $\mathbf{a}^*(\beta)$ denoting the equilibrium actions of agents in the communication chain given that the first agent has posterior belief β. Naturally, if the posterior belief is $\beta < p$, then the interim expected payoff is 0. Given that $\pi\left(1 \mid 1\right) = 1$ must hold in the optimal information policy, the only possible $\beta < p$ is for $s = 0$ and equals zero, i.e. $\beta(0) = 0$. The posterior belief $\beta \geq p$ can be achieved only for $s = 1$. Given that $\pi\left(1 \mid 1\right) = 1$, the posterior belief for $s = 1$, denoted by $\beta(1)$, is determined by the slant, $\pi\left(1 \mid 0\right)$. For the case $\pi^D < \pi^L$, there are thresholds β^D, β^L, and β^H such that the posterior belief $\beta(1)$ is:

(i) $\beta(1) \geq \beta^D$ if and only if $\pi\left(1 \mid 0\right) \leq \pi^D$;

(ii) $\beta^L \leq \beta(1) < \beta^D$ if and only if $\pi^D < \pi\left(1 \mid 0\right) \leq \pi^L$;

(iii) $\beta^H \leq \beta(1) < \beta^L$ if and only if $\pi^L < \pi\left(1 \mid 0\right) \leq \pi^H$; and

(iv) $p \leq \beta(1) < \beta^H$ if and only if $\pi\left(1 \mid 0\right) > \pi^H$.

Then, the posterior belief $\beta(1)$ determines the persuasion and diffusion pattern and, in effect, the interim expected payoff, which is: (i) V^D for $\beta(1) \geq \beta^D$, (ii) V^L for $\beta^L \leq \beta(1) < \beta^D$, (iii) V^H for $\beta^H \leq \beta(1) < \beta^L$, and (iv) 0 for $p \leq \beta(1) < \beta^H$. Hence, overall, the manipulator's interim expected payoff is indeed fully determined by

the posterior belief that is induced by the signal realisation on the first agent in the communication chain. We can write $\hat{V}(\beta) = E_\beta \left[V \left(\mathbf{a}^*(\beta) \right) \right]$, i.e. if the posterior belief of the first agent in the communication chain is β, then the interim expected payoff to the manipulator is $\hat{V}(\beta)$.

The second observation by Kamenica and Gentzkow (2011) is that for any distribution of posteriors τ such that the expected posterior under this distribution equals the prior, i.e. $E_\tau \left[\beta(s) \right] = p$, there exists a signal π which, given the prior p, induces the distribution of posteriors τ. This allows us to express the manipulator's problem as

$$\max_{\tau \text{ s.t. } E_\tau[\beta]=p} E_\tau \left[\hat{V}(\beta) \right], \tag{13}$$

i.e. we can simply look for the optimal distribution of posteriors such that $E_\tau \left[\beta \right] = p$. The solution to this problem of the manipulator can be easily found using the concavification of $\hat{V}$, i.e. the smallest concave function that is everywhere weakly greater than $\hat{V}$. Let $\mathbf{V}$ denote the concavification of $\hat{V}$. The ex-ante expected payoff from the manipulator's optimal signal—which I refer to as the value of the optimal signal—is then simply $\mathbf{V}(p)$, i.e. the value of the concavification at prior belief p (Kamenica and Gentzkow, 2011).

In order to find the optimal signal, we need to obtain the concavification of $\hat{V}$. The function $\hat{V}$ takes a value of (i) 0 for $0 \leq \beta < \beta^H$, (ii) V^H for $\beta^H \leq \beta < \beta^L$, (iii) V^L for $\beta^L \leq \beta < \beta^D$, and (iv) V^D for $\beta \geq \beta^D$. The exact shape of $\hat{V}$ thus depends on the values of β^H, β^L, β^D, V^H, V^L , and V^D. That shape determines the concavification $\mathbf{V}$ and hence the optimal signal.

Figures 5, 6, and 7 illustrate the manipulator's interim expected payoff $\hat{V}$ and its concavification $\mathbf{V}$ as functions of β for three different combinations of parameter values.

In Figure 5, high polarisation b keeps β^H far away to the left from β^L, as well as β^D far away to the right from β^L. High assortativity r keeps V^H high, close to V^L and V^D. The value of the optimal signal is identified by finding the value of the concavification $\mathbf{V}$ at prior belief p. Once we find the value of the optimal signal, $\mathbf{V}(p)$, it is easy to identify the posterior beliefs induced by the optimal signal. The value $\mathbf{V}(p)$ is a linear combination of $\hat{V}(0)$ and $\hat{V}(\beta^H)$. Hence, the posterior beliefs induced by the optimal

signal are 0 and β^H. Having identified the induced posterior beliefs, it is straightforward to derive the unique signal that induces these beliefs: it is given by $\pi(1 \mid 1) = 1$ and $\pi(1 \mid 0) = \pi^H$. We conclude that the optimal slant is $\pi(1 \mid 0) = \pi^H$.

In Figure 6, moderate b moves β^H to the right, closer to β^L. Moderate b and moderate r keep β^D relatively far away to the right from β^L. Moderate r keeps V^L high, close to V^D, while making V^H relatively low. The optimal signal then induces posterior beliefs 0 and β^L. Therefore, the optimal signal is given by $\pi(1 \mid 1) = 1$ and $\pi(1 \mid 0) = \pi^L$. The optimal slant is thus $\pi(1 \mid 0) = \pi^L$.

In Figure 7, low b and low r shift β^D to the left, closer to β^L. Low b also shifts β^H to the right, closer to β^L. Low r means that both V^L and V^H are low relative to V^D. The optimal signal induces posterior beliefs 0 and β^D. Therefore, the optimal signal is given by $\pi(1 \mid 1) = 1$ and $\pi(1 \mid 0) = \pi^D$. Hence, the optimal slant is $\pi(1 \mid 0) = \pi^D$.

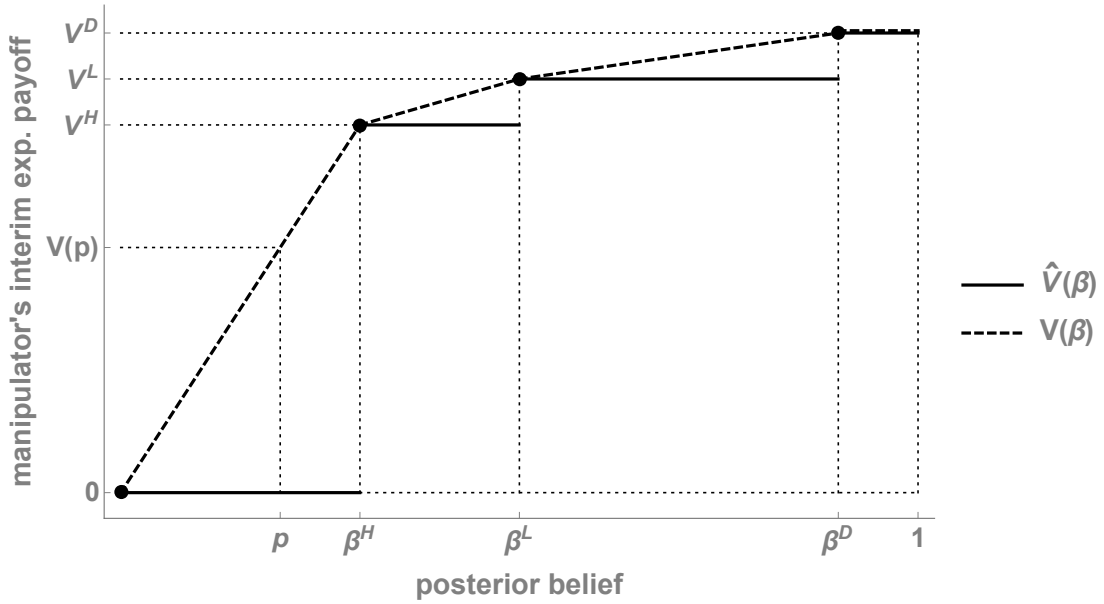

Figure 5: An illustration of the manipulator's interim expected payoff $\hat{V}$ and its concavification $\mathbf{V}$ such that the optimal slant is $\pi(1 \mid 0) = \pi^H$.

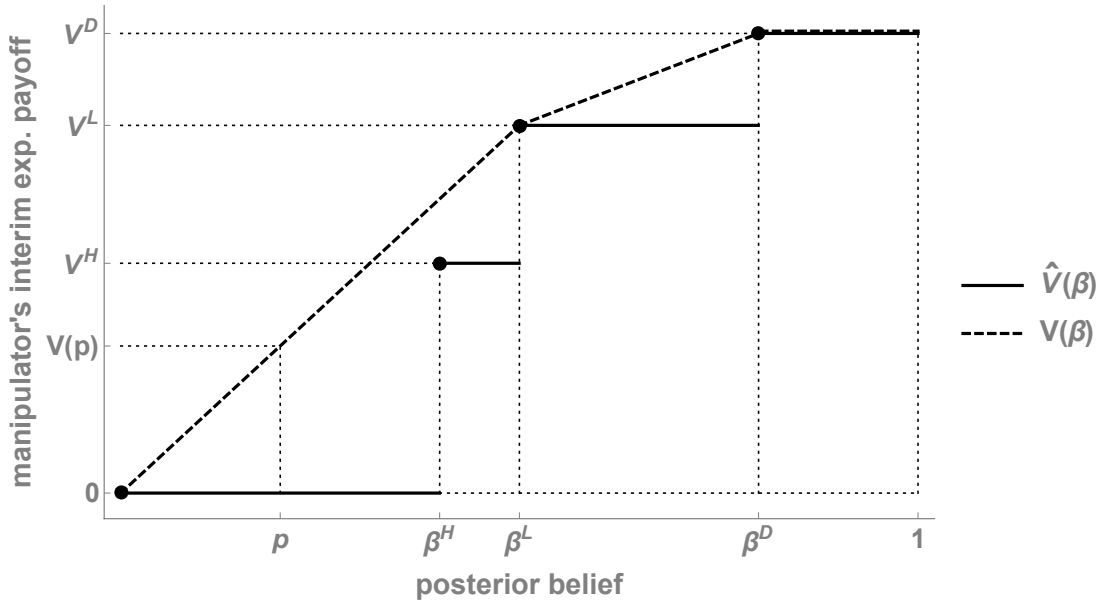

Figure 6: An illustration of the manipulator's interim expected payoff $\hat{V}$ and its concavification $\mathbf{V}$ such that the optimal slant is $\pi(1 \mid 0) = \pi^L$.

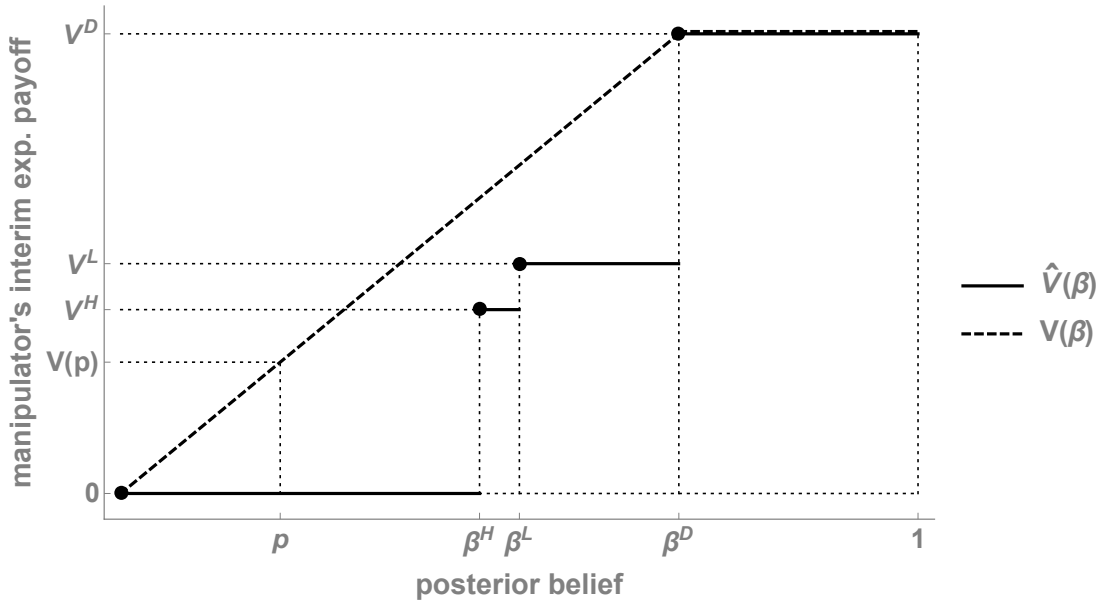

Figure 7: An illustration of the manipulator's interim expected payoff $\hat{V}$ and its concavification $\mathbf{V}$ such that the optimal slant is $\pi(1 \mid 0) = \pi^D$.

On a final note, it is instructive to compare these three figures with the concavification in a setup without diffusion, i.e. where the signal realisation is received only by one H-type agent. In such a setup, the manipulator's payoff would simply be a step function of β that takes values of 0 for $0 \leq \beta < \beta^H$ and 1 for $\beta^H \leq \beta \leq 1$. The posterior beliefs induced by the optimal signal would then be 0 and β^H, and so the optimal slant would be $\pi(1 \mid 0) = \pi^H$.

5 Extensions

This section considers two extensions: (i) diffusion in a communication chain where agents do not observe the types of their successors, and (ii) misestimation of the slant by the agents, with H-type agents underestimating it and L-type agents overestimating it. I also discusses the implications of these extensions for the optimal information policy of the manipulator. As discussed earlier, the first extension is motivated by the growing role of social media, where people often share news with their social network rather than with specific individuals, while the second is motivated by the tendency of people to trust media which match their political views and distrust those which do not.

5.1 Diffusion with Unobservable Types of Successors

The main model assumes that each agent observes the type of her successor in the communication chain. Here, I analyse a modified model, where each agent does not observe her successor's type in the chain.

Formally, the modification of the model is that the information set of each agent in the communication chain (conditional on meeting a successor) is $I_0 = \{s\}$, rather than $I_0 = \{s, t_1\}$. It follows then that the communication strategy of an agent of type $t_0 \in \{L, H\}$ is a function $\mu_{t_0} : \{s\} \to \{s, \emptyset\}$, where $\mu_{t_o}(s) = s$ denotes passing s on and $\mu_{t_o}(s) = \emptyset$ denotes not passing s on. Otherwise, the model is unchanged. In particular, the value of the assortativity parameter, r, is still common knowledge. Therefore, each agent knows that, if she passes s on, it will be received by an agent of the same type

with probability r and by an agent of the opposite type with probability $1-r$.[17]

In this modified setup, the game has a unique equilibrium, which is described in the following proposition.

Proposition 6. *In the unique equilibrium, the agents' beliefs and action strategies are the same as in Proposition 1, and the agents' communication strategies are*

$$\mu_L(s,t_1)=\begin{cases} s & \text{for } s=0 \\ \begin{cases} s & \text{if and only if } \beta(1)\geq\beta(1)^*_{\mathrm{UT}} \\ \varnothing & \text{if and only if } \beta(1)<\beta(1)^*_{\mathrm{UT}} \end{cases} & \text{for } s=1 \end{cases}$$

where $\beta(1)^*_{\mathrm{UT}}=p+2b\frac{1-r}{1+\delta-2r\delta}$, *and*

$$\mu_H(s,t_1)=\begin{cases} s & \text{for } s=1 \\ \begin{cases} s & \text{if and only if } \beta(0)\leq\beta(0)^*_{\mathrm{UT}} \\ \varnothing & \text{if and only if } \beta(0)>\beta(0)^*_{\mathrm{UT}} \end{cases} & \text{for } s=0 \end{cases}$$

where $\beta(0)^*_{\mathrm{UT}}=p-2b\frac{1-r}{1+\delta-2r\delta}$.

Thus, the equilibrium communication strategy is such that an L-type agent always passes $s=0$ on, and passes $s=1$ on if and only if $\beta(1)$ is high enough, i.e. if and only if $s=1$ is sufficiently informative about $\omega=1$. Conversely, an H-type agent always passes $s=1$ on, but passes $s=0$ on if and only if $\beta(0)$ is low enough, i.e. if and only if $s=0$ is sufficiently informative about $\omega=0$. The reasoning behind this is similar to the one in the main setup: as the information policy becomes more informative, the preferences of the two types of agents regarding actions become relatively more aligned, and so an L-type agent (an H-type agent) prefers to pass $s=1$ ($s=0$) on rather to suppress it, even if it is received with some probability by an H-type agent (L-type agent).

Like in the main model, the information policy has two effects: (i) on persuasion and (ii) on diffusion. Given Assumption 1, $\pi(1\mid 1)=1$ must hold in the optimal

[17] I use subscript "UT" (for "Unobservable Types") to distinguish the notation in this extension from the one in the main setup.

information policy, and hence I can analyse these effects in the context of the slant, $\pi(1 \mid 0)$. The effect on persuasion is the same as in the main model, i.e. as described in Proposition 2. However, the effect on diffusion is now somewhat different. It is described in the following proposition.

Proposition 7. *In the optimal information policy:*
(i) an H-type agent passes $s = 1$ on under any slant,
(ii) an L-type agent passes $s = 1$ on if and only if the slant is $\pi(1 \mid 0) \leq \pi^D_{\mathrm{UT}}$, where

$$\pi^D_{\mathrm{UT}} = \frac{p}{1-p} \frac{1 - \left(p + 2b\frac{1-r}{1+\delta-2r\delta}\right)}{p + 2b\frac{1-r}{1+\delta-2r\delta}}. \tag{14}$$

The threshold π^D_{UT} is increasing in assortativity r, $\frac{\partial \pi^D_{\mathrm{UT}}}{\partial r} > 0$, and decreasing in polarisation b, $\frac{\partial \pi^D_{\mathrm{UT}}}{\partial b} < 0$, with $\frac{\partial^2 \pi^D_{\mathrm{UT}}}{\partial r \partial b} < 0$ for sufficiently low r and sufficiently high b, and $\frac{\partial^2 \pi^D_{\mathrm{UT}}}{\partial r \partial b} \geq 0$ otherwise.

Proposition 7 uses the equilibrium communication strategies in Proposition 6 and the observation that $\pi(1 \mid 1) = 1$ must hold in the optimal information policy. As stated in Proposition 6, an L-type agent passes $s = 1$ on if and only if $\beta(1) > p + 2b\frac{1-r}{1+\delta-2r\delta}$. Given that $\pi(1 \mid 1) = 1$, this condition is equivalent to $\pi(1 \mid 0) \leq \pi^D_{\mathrm{UT}}$, where π^D_{UT} is given by (14).

The main difference between the setups with unobservable and observable types of successors is that the threshold π^D_{UT} is increasing in assortativity r, whereas π^D is decreasing in r. In other words, as assortativity increases, inducing maximal diffusion becomes easier under unobservable types of successors, but more difficult under observable types. The intuition is that, under unobservable types, as r increases, it becomes more likely that an L-type agent's successor is of the same type, which means that she has a greater incentive to pass $s = 1$ on. Under observable types, as r increases, an L-type agent who meets an H-type agent realises that the H-type agent's successors are more likely to be H-type agents too, so the incentive to pass $s = 1$ on to the H-type successor diminishes.

What this implies for the optimal information policy is that, under unobservable types, as assortativity increases, the manipulator becomes less constrained by diffusion.

Eventually, when r is high enough, the relation $\pi^D_{\mathrm{UT}} \geq \pi^H$ holds, which follows from the fact that π^H does not depend on r and π^D_{UT} is increasing in r. Then, the manipulator is not constrained by diffusion at all: even persuading H-type agents requires a slant such that the diffusion of $s = 1$ by L-type agents is induced anyway. On the other hand, under observable types, as assortativity increases, the manipulator becomes more constrained by diffusion. When r is high enough, the relation $\pi^D < \pi^L$ holds, and thus the slant that is needed to induce L-type agents to pass $s = 1$ on to H-type agents is lower than the slant that is needed to persuade L-type agents.

5.2 Misestimation of the Slant by the Agents

In the main model, the agents observe the information policy—effectively described by the slant—chosen by the manipulator. Put differently, they correctly estimate the slant. Here, I assume that the agents misestimate the slant: L-type agents overestimate it, while H-type agents underestimate it.

Formally, suppose that if the manipulator chooses a slant $\pi(1 \mid 0)$, then an L-type and an H-type agent's beliefs upon observing $s = 1$ are respectively $\beta_L(1) = \beta(1) - e$ and $\beta_H(1) = \beta(1) + e$, where $e \geq 0$ measures the extent to which an L-type agent overestimates and H-type agent underestimates the slant. If $e = 0$, then both types of agents form a belief $\beta_L(1) = \beta_H(1) = \beta(1)$, which means that they update their beliefs using a correct estimate of the slant. I assume that the value of e is such that $\beta(1) - e \geq p$ and $\beta(1) + e \leq 1$ hold, i.e. even if they misestimate the slant, an L-type agent's belief upon observing $s = 1$ is not lower than the prior belief and an H-type agent's belief is not higher than 1.

Proposition 8 describes how misestimation of the slant changes the effects of the slant on persuasion and diffusion, both under observable and unobservable types of successors.

Proposition 8. *Suppose that agents misestimate the slant, with $e \geq 0$ measuring how much an L-type agent overestimates it and an H-type agent underestimates it. In the optimal information policy:*

(i) an H-type agent is persuaded by $s = 1$ if and only if the slant is $\pi(1 \mid 0) \leq \pi^H$, where π^H is increasing in e,
(ii) an L-type agent is persuaded by $s = 1$ if and only if the slant is $\pi(1 \mid 0) \leq \pi^L$, where π^L is decreasing in e,
(iii) under observable types of successors, both types of agents pass $s = 1$ on to their successor regardless of the successor's type if and only if the slant is $\pi(1 \mid 0) \leq \pi^D$, where π^D is decreasing in e; otherwise, $s = 1$ is passed on by agents except when an L-type agent meets an H-type agent,
(iv) under unobservable types of successors, both types of agents pass $s = 1$ on to their successor if and only if the slant is $\pi(1 \mid 0) \leq \pi^D_{\mathrm{UT}}$, where π^D_{UT} is decreasing in e; otherwise, $s = 1$ is passed on only by H-type agents.

Thus, as H-type agents underestimate the slant more and more while L-type agents overestimate it more and more, it becomes easier for the manipulator to persuade H-type agents, but more difficult to persuade L-type agents. Furthermore, it becomes more difficult to induce maximal diffusion—both under observable and unobservable types of successors.

Figure 8 provides a graphical illustration of how the incentives of an L-type agent to pass $s = 1$ to an H-type agent are shaped by the misestimation of the slant. In Panel A, the misestimation is lower than in Panel B. As the overestimation by the L-type agent and the underestimation by the H-type agent increase, the L-type agent's optimal action for $s = 1$ decreases while the H-type agent's optimal action for $s = 1$ increases. In effect, the L-type agent's optimal action for $s = 1$ moves closer to the agents' optimal actions for the prior belief and hence not passing $s = 1$ on to an H-type agent becomes more attractive than passing it on.

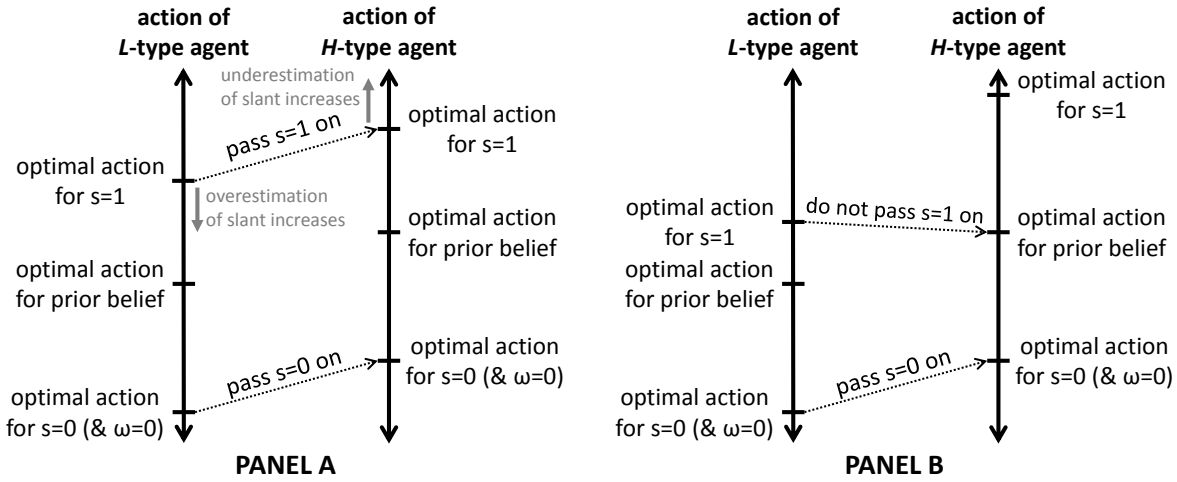

Figure 8: Graphical illustration of how an increase in the overestimation of the slant by L-type agents and its underestimation by H-type agents affects the incentives of an L-type agent to pass $s = 1$ to an H-type agent.

Misestimation of the slant (with L-type agents overestimating it and H-type agents underestimating it) has clear implications for the optimal information policy. First, both under observable and unobservable types of successors, it contributes to making the partisan information policy possible because, by decreasing π^D (π^D_{UT}) and increasing π^H, it can make the relation between π^D (π^D_{UT}) and π^H be $\pi^D < \pi^H$ ($\pi^D_{\mathrm{UT}} < \pi^H$). Then, a slant $\pi(1 \mid 0) = \pi^H$ corresponds to a partisan policy. Second, by decreasing both π^D (π^D_{UT}) and π^L while increasing π^H, it makes the partisan information policy more profitable and other policies less profitable. Therefore, overall, the misestimation of the slant increases the chances that the manipulator chooses the partisan information policy.

6 Conclusion

Partisan slant is a common feature of today's media landscape and it is widely documented that it can influence people's beliefs and behaviour. It is therefore important to study what drives media outlets to have a partisan slant. The existing theoretical

work has given little attention to the fact that news reported by media outlets do not only reach their direct audience, but can also spread to the wider public.

This paper aims to fill this gap by developing a model of media slant where manipulation of information by an outlet is followed by diffusion of the information by word of mouth. It features a manipulator who designs an information policy, which is a mapping from facts to news reports. The reported news then spread via a communication chain in a population of agents with heterogeneous preferences. At the methodological level, the model combines Bayesian persuasion with diffusion via a communication chain. The model is stylised but its simplicity allows us to see clearly the mechanisms through which diffusion can influence media slant. The key to the results is that the slant of the information policy has an effect not only on whether the agents find the news credible, but also on the agents' incentives to pass them on to others. The interplay between these two effects gives a spectrum of possible information policies, ranging from a partisan policy to a mainstream policy. The analysis elucidates how two characteristics of the environment, i.e. polarisation and assortativity of the population, influence the choice of policy by the manipulator.

The model offers plenty of scope for further analysis. One simplifying assumption is that there are only two types of agents. In future work, one could introduce a spectrum of types of agents, with assortativity varying along the spectrum. The media outlet's strategy could then be two-dimensional: it could consist of choosing not only the slant but also the type of its audience, i.e. the type of the first agent in the chain. Another simplifying assumption is that there is a single media outlet, while in the real world there are usually multiple media outlets that compete with each other. It would be interesting to analyse whether—in an environment where diffusion by word of mouth is possible—competition between media outlets leads to lower or higher media slant. Finally, a promising avenue is to analyse the role of word of mouth in demand-driven slant. For example, a media outlet may care about the diffusion of its news because its advertising revenue depends on the number of people who enter its website (e.g., after receiving a link from a friend) to read the news. At the same time, people may prefer to receive and pass on confirmatory news, i.e. news which match their preferences or prior beliefs. Then, the outlet may prefer to report its news with a partisan slant even

in the absence of a direct preference to influence people's beliefs.

References

[1] Alonso, R., and O. Câmara (2016): "Persuading voters," *American Economic Review*, 106 (11), 3590-3605.

[2] Ambrus, A., E. Azevedo and Y. Kamada (2013): "Hierarchical cheap talk," *Theoretical Economics*, 8, 233-261.

[3] Anderlini, L., D. Gerardi and R. Lagunoff (2012): "Communication and learning," *Review of Economic Studies*, 79, 419-450.

[4] Anderson, S.P., and J. McLaren (2012): "Media mergers and media bias with rational consumers," *Journal of European Economic Association*, 10 (4), 831–859.

[5] Ansolabehere, S., R.R. Lessem and J.M. Snyder Jr. (2006): "The orientation of newspaper endorsements in U.S. elections, 1940–2002," *Quarterly Journal of Political Science*, 1 (4), 393–404.

[6] Baron, D.P. (2006): "Persistent media bias," *Journal of Public Economics*, 90 (1–2), 1–36.

[7] Bloch, F., G. Demange and R. Kranton (2017): Rumors and social networks," *International Economic Review*, forthcoming.

[8] Chiang, C.-F., and B. Knight (2011): "Media bias and influence: evidence from newspaper endorsements," *Review of Economic Studies*, 78 (3), 795–820.

[9] DellaVigna, S., and M. Gentzkow (2010): "Persuasion: empirical evidence," in Arrow, K.J., and T. Bresnahan (eds.) *Annual Review of Economics*, Volume 2.

[10] DellaVigna, S., and E. Kaplan (2007): "The Fox News effect: media bias and voting," *Quarterly Journal of Economics*, 122, 1187–1234.

[11] Durante, R., and B. Knight (2012): "Partisan control, media bias, and viewers' responses: evidence from Berlusconi's Italy," *Journal of European Economic Association*, 10 (3), 451–481.

[12] Enikopolov, R., M. Petrova and E. Zhuravskaya (2011): "Media and political persuasion: evidence from Russia," *American Economic Review*, 101 (7), 3253-3285.

[13] Gehlbach, S., and K. Sonin (2014): "Government control of the media," *Journal of Public Economics*, 118, 163–171.

[14] Gentzkow, M., and J.M. Shapiro (2006): "Media bias and reputation," *Journal of Political Economy*, 114, 280–316.

[15] ——— (2010): "What drives media slant? Evidence from U.S. daily newspapers," *Econometrica*, 78 (1), 35–71.

[16] Gentzkow, M., J.M. Shapiro and M. Sinkinson (2011): "The effect of newspaper entry and exit on electoral politics," *American Economic Review*, 101 (7), 2980–3018.

[17] Gentzkow, M., J.M. Shapiro and D.F. Stone (2015): "Media bias in the marketplace: theory," in Anderson, S., J. Waldfogel and D. Strömberg (eds.) *Handbook of Media Economics*, Volume 1 (Elsevier, Amsterdam).

[18] Gerber, A., D.S. Karlan and D. Bergan (2009): "Does the media matter? A field experiment measuring the effect of newspapers on voting behavior and political opinions," *American Economic Journal: Applied Economics*, 1 (2), 35–52.

[19] Groseclose, T., and J. Milyo (2005): "A measure of media bias," *Quarterly Journal of Economics*, 120, 1191–1237.

[20] Kamenica, E., and M. Gentzkow (2011): "Bayesian persuasion," *American Economic Review*, 101(6), 2590-2615.

[21] Klofstad, C.A., S.D. McClurg and M. Rolfe (2009): "Measurement of political discussion networks: a comparison of two "name generator" procedures," *Public Opinion Quarterly*, 73 (3), 462-483.

[22] Laclau, M., and L. Renou (2017): "Public persuasion," Working Paper; https://sites.google.com/site/marielaclauvigeral/research

[23] Larcinese, V., R. Puglisi and J.M. Snyder Jr. (2011): "Partisan bias in economic news: evidence on the agendasetting behavior of U.S. newspapers," *Journal of Public Economics*, 95 (9–10), 1178–1189.

[24] Martin, G.J., and A. Yurukoglu (2017): "Bias in cable news: persuasion and polarization," *American Economic Review*, 107 (9), 2565-2599.

[25] McPherson, M., L. Smith-Lovin and J.M. Cook (2001): "Birds of a feather: homophily in social networks," *Annual Review of Sociology*, 27, 415-444.

[26] Mullainathan, S., and A. Shleifer (2005): "The market for news," *American Economic Review*, 95 (4), 1031–1053.

[27] Prior, M. (2013): "Media and political polarisation," *Annual Review of Political Science*, 16, 101-127.

[28] Puglisi, R., and J.M. Snyder Jr. (2011): "Newspaper coverage of political scandals," *Journal of Politics*, 73 (3), 1–20.

[29] ——— (2015): "Empirical studies of media bias," in Anderson, S., J. Waldfogel and D. Strömberg (eds.) *Handbook of Media Economics*, Volume 1 (Elsevier, Amsterdam).

[30] Stone, D.F. (2011): "Ideological media bias," *Journal of Economic Behavior & Organization*, 78 (3), 256–271.

[31] Taneva, I. (2016): "Information Design," Working Paper; https://sites.google.com/site/itaneva13/research

Appendix

Proof of Proposition 1

The proofs for the equilibrium beliefs and action strategy follow from the main text. For the equilibrium communication strategy, consider first $s = 1$. From the payoff functions it follows that

$$u_H\left(\beta(1)+b,\omega_i \mid \beta(1)\right) > u_H\left(p+b,\omega_i \mid \beta(1)\right), \tag{15}$$

$$u_H\left(\beta(1),\omega_i \mid \beta(1)\right) > u_H\left(p,\omega_i \mid \beta(1)\right). \tag{16}$$

Given (15) and (16), for any $\mu_L(1,t_1)$, a sequentially rational strategy must have $\mu_H(1,H)=1$ and $\mu_H(1,L)=1$.

From the payoff functions it also follows that

$$u_L\left(\beta(1),\omega_i \mid \beta(1)\right) > u_L\left(p,\omega_i \mid \beta(1)\right), \tag{17}$$

Given (17), and given that in equilibrium $\mu_L(1,H)$ is sequentially rational, a sequentially rational strategy must have $\mu_L(1,L)=1$.

It remains to consider $\mu_L(1,H)$ given that $\mu_H(1,H)=1$, $\mu_H(1,L)=1$, and $\mu_L(1,L)=1$. Denote by $\widetilde{U}_L\left(\mathbf{a_0}(1) \mid \{1,t_1\}\right)$ the expected payoff to an L-type agent—who has received $s=1$—from passing $s=1$ on to a t_1-type agent, and denote by $\widetilde{U}_U\left(\mathbf{a_0}(\emptyset) \mid \{1,t_1\}\right)$ the expected payoff to an L-type agent—who has received $s=1$—from not passing $s=1$ on to t_1-type agent. If $\widetilde{U}_L\left(\mathbf{a_0}(1) \mid \{1,H\}\right) \geq \widetilde{U}_L\left(\mathbf{a_0}(\emptyset) \mid \{1,H\}\right)$, then a sequentially rational strategy must have $\mu_L(1,H)=1$; otherwise it must have $\mu_L(1,H)=\emptyset$.

We derive $\widetilde{U}_L\left(\mathbf{a_0}(1) \mid \{1,H\}\right)$ by solving the simultaneous equations:

$$\begin{aligned}
\widetilde{U}_L\left(\mathbf{a_0}(1) \mid \{1,H\}\right) &= -b^2 + r\delta\widetilde{U}_L\left(\mathbf{a_0}(1) \mid \{1,H\}\right) + \\
&\quad +(1-r)\delta\widetilde{U}_L\left(\mathbf{a_0}(1) \mid \{1,L\}\right) \qquad (18)\\
\widetilde{U}_L\left(\mathbf{a_0}(1) \mid \{1,L\}\right) &= r\delta\widetilde{U}_L\left(\mathbf{a_0}(1) \mid \{1,L\}\right) + \\
&\quad +(1-r)\delta\widetilde{U}_L\left(\mathbf{a_0}(1) \mid \{1,H\}\right). \qquad (19)
\end{aligned}$$

We obtain

$$\widetilde{U}_L\left(\mathbf{a_0}(1) \mid \{1,H\}\right) = \left(-b^2\right) \frac{1-r\delta}{(1-r\delta)^2-(1-r)^2\delta^2}. \tag{20}$$

Similarly, we derive $\widetilde{U}_L\left(\mathbf{a_0}(\emptyset) \mid \{1,H\}\right)$ by solving the simultaneous equations:

$$\begin{aligned}
\widetilde{U}_L\left(\mathbf{a_0}(\emptyset) \mid \{1,H\}\right) &= -\left(\beta(1)-p-b\right)^2 + r\delta\widetilde{U}_L\left(\mathbf{a_0}(\emptyset) \mid \{1,H\}\right) + \\
&\quad +(1-r)\delta\widetilde{U}_L\left(\mathbf{a_0}(\emptyset) \mid \{1,L\}\right) \qquad (21)\\
\widetilde{U}_L\left(\mathbf{a_0}(\emptyset) \mid \{1,L\}\right) &= -\left(\beta(1)-p\right)^2 r\delta\widetilde{U}_L\left(\mathbf{a_0}(\emptyset) \mid \{1,L\}\right) + \\
&\quad +(1-r)\delta\widetilde{U}_L\left(\mathbf{a_0}(\emptyset) \mid \{1,H\}\right). \qquad (22)
\end{aligned}$$

We obtain

$$\begin{aligned}
\widetilde{U}_L\left(\mathbf{a_0}(\emptyset) \mid \{1,H\}\right) &= \left(-\left(\beta(1)-p-b\right)^2 - \frac{(1-r)\delta}{1-r\delta}\left(\beta(1)-p\right)^2\right) \times \\
&\quad \times \frac{1-r\delta}{(1-r\delta)^2-(1-r)^2\delta^2}. \qquad (23)
\end{aligned}$$

Then, $\widetilde{U}_L\left(\mathbf{a_0}(1) \mid \{1,H\}\right) \geq \widetilde{U}_L\left(\mathbf{a_0}(\emptyset) \mid \{1,H\}\right)$ if and only if

$$\left(\beta(1)-p-b\right)^2 + \frac{(1-r)\delta}{1-r\delta}\left(\beta(1)-p\right)^2 - b^2 \geq 0, \tag{24}$$

which can be rearranged to

$$\beta(1) \geq p + 2b\frac{1-\delta r}{1+\delta-2\delta r}. \tag{25}$$

We repeat analogous steps to derive the equilibrium communication strategy for $s = 0$. First, we show that $\mu_L(0,L) = 0$, $\mu_L(0,H) = 0$ and $\mu_H(0,H) = 0$ must hold in equilibrium. Then, we consider $\mu_H(0,L)$ and show that $\widetilde{U}_H\left(\mathbf{a_0}(0) \mid \{0,L\}\right) \geq \widetilde{U}_H\left(\mathbf{a_0}(\emptyset) \mid \{0,L\}\right)$ if and only if

$$\left(p-\beta(0)-b\right)^2 + \frac{(1-r)\delta}{1-r\delta}\left(p-\beta(0)\right)^2 - b^2 \geq 0, \tag{26}$$

which can be rearranged to

$$\beta(0) \leq p - 2b\frac{1-\delta r}{1+\delta-2\delta r}. \tag{27}$$

Proof of Proposition 2

The proof follows from the main text.

Proof of Proposition 3

The proof follows from the main text. To see that $\frac{\partial \pi^D}{\partial r} < 0$ and $\frac{\partial \pi^D}{\partial b} < 0$, note that the term $2b\frac{1-\delta r}{1+\delta-2\delta r}$ is increasing in r and b. In terms of b, one can show that $\frac{\partial^2 \pi^D}{\partial r \partial b} \geq 0$ if and only if $b \geq \frac{p(1+\delta-2\delta r)}{2(1-\delta r)}$, and $\frac{\partial^2 \pi^D}{\partial r \partial b} < 0$ otherwise. In terms of r, one can show that $\frac{\partial^2 \pi^D}{\partial r \partial b} \geq 0$ if and only if $r = 1$ for $b = \frac{p}{2}$, $\frac{p(1+\delta)-2b}{2\delta(p-b)} \leq r < 1$ for $\frac{p}{2} < b \leq \frac{1}{2}p\,(1+\delta)$, and $0 \leq r \leq 1$ for $b > \frac{1}{2}p\,(1+\delta)$, and $\frac{\partial^2 \pi^D}{\partial r \partial b} < 0$ otherwise.

Proof of Proposition 4

(a) The derivatives with respect to r are $\frac{\partial \pi^D}{\partial r} < 0$, $\frac{\partial \pi^L}{\partial r} = 0$, and $\frac{\partial \pi^H}{\partial r} = 0$. Since $\pi^H > \pi^L$ holds, this implies that there exists r^{**} such that $\pi^D < \pi^L$ if and only if $r > r^{**}$ (and $\pi^D \geq \pi^L$ otherwise) and r^* such that $\pi^D < \pi^H$ if and only if $r > r^*$ (and $\pi^D \geq \pi^H$ otherwise), where $r^{**} > r^*$. The result in the proposition follows.
(b) The derivatives with respect to b are $\frac{\partial \pi^D}{\partial b} < 0$, $\frac{\partial \pi^L}{\partial b} = 0$, and $\frac{\partial \pi^H}{\partial b} > 0$. Since $\pi^H > \pi^L$ holds, this implies that there exists b^{**} such that $\pi^D < \pi^L$ if and only if $b > b^{**}$ (and $\pi^D \geq \pi^L$ otherwise) and r^* such that $\pi^D < \pi^H$ if and only if $b > b^*$ (and $\pi^D \geq \pi^H$ otherwise), where $b^{**} > b^*$. The result in the proposition follows. The values of b^* and b^{**} are

$$b^* = (\overline{a} - p)\left(\frac{1+\delta-2\delta r}{3+\delta-4\delta r}\right), \tag{28}$$

$$b^{**} = (\overline{a} - p)\left(\frac{1+\delta-2\delta r}{2-2\delta r}\right). \tag{29}$$

It is then straightforward to show that $b^* \in (0, \overline{a} - p)$ and $b^{**} \in (0, \overline{a} - p)$.

Proof of Proposition 5

Let $\tilde{V}\left(\pi\left(1 \mid 0\right)\right)$ denote the manipulator's expected payoff from an information policy with a slant $\pi\left(1 \mid 0\right)$. The expected payoffs from slants π^D, π^L, and π^H are

$$\begin{aligned}
\tilde{V}(\pi^D) &= \left(p + (1-p)\,\pi^D\right) V^D, && (30)\\
\tilde{V}(\pi^L) &= \left(p + (1-p)\,\pi^L\right) V^L, && (31)\\
\tilde{V}(\pi^H) &= \left(p + (1-p)\,\pi^H\right) V^H. && (32)
\end{aligned}$$

(a) The derivatives with respect to r are

$$\begin{aligned}
\frac{\partial \tilde{V}(\pi^D)}{\partial r} &< 0, && (33)\\
\frac{\partial \tilde{V}(\pi^L)}{\partial r} &> 0, && (34)\\
\frac{\partial \tilde{V}(\pi^H)}{\partial r} &> 0. && (35)
\end{aligned}$$

It follows from (33) and (34) that there exists r^{DL}, expressed as a function of other parameters, such that $\tilde{V}(\pi^D) \geq \tilde{V}(\pi^L)$ if and only if $r \leq r^{DL}$. Similarly, it follows from (33) and (35) that there exists r^{DH}, expressed as a function of other parameters, such that $\tilde{V}(\pi^D) \geq \tilde{V}(\pi^H)$ if and only if $r \leq r^{DH}$. Finally, we can check that $\tilde{V}(\pi^L) \geq \tilde{V}(\pi^H)$ if and only if $r \leq r^{LH}$, where $r^{LH} = \frac{\delta\bar{a} - (1+\delta)b}{\delta(\bar{a}-2b)}$ and $\bar{a} > 2b$.

Note that $r > r^{LH}$ implies $r^{DL} > r^{DH}$, and it follows that $r^{LH} < r^{DH} < r^{DL}$ as there is a contradiction otherwise. Suppose for example that $r^{DH} < r^{DL} < r^{LH}$. Then, for $r \in (r^{DH}, r^{DL})$, $r > r^{DH}$ implies $\tilde{V}(\pi^H) > \tilde{V}(\pi^D)$, $r < r^{DL}$ implies $\tilde{V}(\pi^D) > \tilde{V}(\pi^L)$, and $r < r^{LH}$ implies $\tilde{V}(\pi^L) > \tilde{V}(\pi^H)$, and hence we reach a contradiction. Similarly, $r < r^{LH}$ implies $r^{DL} < r^{DH}$, and it follows that $r^{DL} < r^{DH} < r^{LH}$ as there is a contradiction otherwise. Finally, if $r = r^{LH}$, then we must have $r^{LH} = r^{DH} = r^{DL}$, $r^{LH} < r^{DH} < r^{DL}$, or $r^{DL} < r^{DH} < r^{LH}$, as there is a contradiction otherwise.

The above implies that $\tilde{V}(\pi^D) \geq \max\{\tilde{V}(\pi^L), \tilde{V}(\pi^H)\}$ holds if and only if $r \leq \min\{r^{DL}, r^{DH}\}$; $\tilde{V}(\pi^L) > \tilde{V}(\pi^D)$ and $\tilde{V}(\pi^L) \geq \tilde{V}(\pi^H)$ hold if and only if $r > r^{DL}$ and $r \leq r^{LH}$; and $\tilde{V}(\pi^H) > \max\{\tilde{V}(\pi^D), \tilde{V}(\pi^L)\}$ holds if and only if $r > \max\{r^{DH}, r^{LH}\}$.

To see that $\max\{r^{DH}, r^{LH}\} < 1$ holds, note that $\tilde{V}(\pi^H) > \tilde{V}(\pi^D)$ and $\tilde{V}(\pi^H) > \tilde{V}(\pi^L)$ for $r = 1$ for all values of other parameters and that $\tilde{V}(\pi^D)$, $\tilde{V}(\pi^H)$, and $\tilde{V}(\pi^L)$ are continuous.

(b) The derivatives with respect to b are

$$\frac{\partial \tilde{V}(\pi^D)}{\partial b} < 0, \tag{36}$$

$$\frac{\partial \tilde{V}(\pi^L)}{\partial b} = 0, \tag{37}$$

$$\frac{\partial \tilde{V}(\pi^H)}{\partial b} > 0. \tag{38}$$

It follows from (36) and (37) that there exists b^{DL}, expressed as a function of other parameters, such that $\tilde{V}(\pi^D) \geq \tilde{V}(\pi^L)$ if and only if $b \leq b^{DL}$. Similarly, from (36) and (38) it follows that there exists b^{DH}, expressed as a function of other parameters, such that $\tilde{V}(\pi^D) \geq \tilde{V}(\pi^H)$ if and only if $b \leq b^{DH}$, and from (37) and (38) it follows that there exists b^{LH}, expressed as a function of other parameters, such that $\tilde{V}(\pi^L) \geq \tilde{V}(\pi^H)$ if and only if $b \leq b^{LH}$.

Following a similar argument as in (a), the only possible relations are $b^{LH} < b^{DH} < b^{DL}$, $b^{LH} = b^{DH} = b^{DL}$, and $b^{DL} < b^{DH} < b^{LH}$.

The above implies that $\tilde{V}(\pi^D) \geq \max\{\tilde{V}(\pi^L), \tilde{V}(\pi^H)\}$ holds if and only if $b \geq \min\{b^{DL}, b^{DH}\}$; $\tilde{V}(\pi^L) > \tilde{V}(\pi^D)$ and $\tilde{V}(\pi^L) \geq \tilde{V}(\pi^H)$ hold if and only if $b > b^{DL}$ and $b \leq b^{LH}$; and $\tilde{V}(\pi^H) > \max\{\tilde{V}(\pi^D), \tilde{V}(\pi^L)\}$ holds if and only if $b > \max\{b^{DH}, b^{LH}\}$.

Proof of Corollary 1

The proof follows from the main text.

Proof of Proposition 6

Let $\widetilde{U}_{t_0,\widehat{t}}(\mathbf{a_0}(s) \mid s)$ denote the expected payoff to a t_0-type agent from a sequence of actions of agents in which the 0-th agent is of type $\widehat{t}$ and the signal s is passed on, given that the signal realisation is s. Let $\widetilde{U}_{t_0,\widehat{t}}(\mathbf{a_0}(\emptyset) \mid s)$ denote the expected payoff to a t_0-type agent from a sequence of actions of agents in which the 0-th agent is

of type $\widehat{t}$ and the signal s is not passed on, given that the signal realisation is s. If $\widetilde{U}_{t_0,\widehat{t}}(\mathbf{a_0}(s) \mid s) \geq \widetilde{U}_{t_0,\widehat{t}}(\mathbf{a_0}(\varnothing) \mid s)$ for $\widehat{t} = t_0$, then a sequentially rational strategy must have $\mu_{t_0}(s) = s$; otherwise it must have $\mu_{t_0}(s) = \varnothing$.

It is straightforward to show that $\widetilde{U}_{H,H}(\mathbf{a_0}(1) \mid 1) > \widetilde{U}_{H,H}(\mathbf{a_0}(\varnothing) \mid 1)$, which implies that a sequentially rational strategy must have $\mu_H(1) = 1$.

Let us consider $\mu_L(1)$ given that $\mu_H(1) = 1$. We derive $\widetilde{U}_{L,L}(\mathbf{a_0}(1) \mid 1)$ by solving the simultaneous equations:

$$\begin{aligned}
\widetilde{U}_{L,L}(\mathbf{a_0}(1) \mid 1) &= (1-r)\left(-b^2\right) + r\delta\widetilde{U}_{L,L}(\mathbf{a_0}(1) \mid 1) + \\
&\quad + (1-r)\,\delta\widetilde{U}_{L,H}(\mathbf{a_0}(1) \mid 1), \qquad (39)\\
\widetilde{U}_{L,H}(\mathbf{a_0}(1) \mid 1) &= r\left(-b^2\right) + r\delta\widetilde{U}_{L,H}(\mathbf{a_0}(1) \mid 1) + \\
&\quad + (1-r)\,\delta\widetilde{U}_{L,L}(\mathbf{a_0}(1) \mid 1). \qquad (40)
\end{aligned}$$

We obtain

$$\widetilde{U}_{L,L}(\mathbf{a_0}(1) \mid 1) = \left((1-r)\left(-b^2\right) + \frac{(1-r)\,\delta}{1-r\delta} r\left(-b^2\right)\right)\frac{1-r\delta}{(1-r\delta)^2 - (1-r)^2\delta^2}. \qquad (41)$$

Similarly, we derive $\widetilde{U}_{L,L}(\mathbf{a_0}(\varnothing) \mid 1)$ by solving the simultaneous equations:

$$\begin{aligned}
\widetilde{U}_{L,L}(\mathbf{a_0}(\varnothing) \mid 1) &= r\left(-(\beta(1)-p)^2\right) + (1-r)\left(-(\beta(1)-p-b)^2\right) + \\
&\quad + r\delta\widetilde{U}_{L,L}(\mathbf{a_0}(\varnothing) \mid 1) + (1-r)\,\delta\widetilde{U}_{L,H}(\mathbf{a_0}(\varnothing) \mid 1), \qquad (42)\\
\widetilde{U}_{L,H}(\mathbf{a_0}(\varnothing) \mid 1) &= r\left(-(\beta(1)-p-b)^2\right) + (1-r)\left(-(\beta(1)-p)^2\right) + \\
&\quad + r\delta\widetilde{U}_{L,H}(\mathbf{a_0}(\varnothing) \mid 1) + (1-r)\,\delta\widetilde{U}_{L,L}(\mathbf{a_0}(\varnothing) \mid 1). \qquad (43)
\end{aligned}$$

We obtain

$$\begin{aligned}
\widetilde{U}_{L,L}(\mathbf{a_0}(\varnothing) \mid 1) &= [r\left(-(\beta(1)-p)^2\right) + (1-r)\left(-(\beta(1)-p-b)^2\right) + \\
&\quad + \frac{(1-r)\,\delta}{1-r\delta}\left(r\left(-(\beta(1)-p-b)^2\right) + (1-r)\left(-(\beta(1)-p)^2\right)\right)] \times \\
&\quad \times \frac{1-r\delta}{(1-r\delta)^2 - (1-r)^2\delta^2}. \qquad (44)
\end{aligned}$$

Then, $\widetilde{U}_{L,L}\left(\mathbf{a_0}(1) \mid 1\right) \geq \widetilde{U}_{L,L}\left(\mathbf{a_0}(\emptyset) \mid 1\right)$ if and only if

$$\beta(1) \geq p + 2b\frac{1-r}{1+\delta-2r\delta}. \tag{45}$$

We repeat analogous steps to derive the equilibrium communication strategy for $s = 0$. It is straightforward to show that $\widetilde{U}_{L,L}\left(\mathbf{a_0}(0) \mid 0\right) > \widetilde{U}_{L,L}\left(\mathbf{a_0}(\emptyset) \mid 0\right)$, which implies that a sequentially rational strategy must have $\mu_L(0) = 0$. Then, we consider $\mu_H(0)$ given that $\mu_L(0) = 0$, and show that $\widetilde{U}_{H,H}\left(\mathbf{a_0}(0) \mid 0\right) \geq \widetilde{U}_{H,H}\left(\mathbf{a_0}(\emptyset) \mid 0\right)$ if and only if

$$\beta(0) \leq p - 2b\frac{1-r}{1+\delta-2r\delta}. \tag{46}$$

Proof of Proposition 7

The proof follows from the main text. To see that $\frac{\partial \pi^D_{\text{UT}}}{\partial r} > 0$ and $\frac{\partial \pi^D}{\partial b} < 0$, note that the term $2b\frac{1-r}{1+\delta-2\delta r}$ is decreasing in r and increasing in b. In terms of b, one can show that $\frac{\partial^2 \pi^D}{\partial r \partial b} \geq 0$ if and only if $b \leq \frac{p(1+\delta-2\delta r)}{2(1-r)}$ for $0 \leq r < 1$ and $b > 0$ for $r = 1$, and $\frac{\partial^2 \pi^D_{\text{UT}}}{\partial r \partial b} < 0$ otherwise. In terms of r, one can show that $\frac{\partial^2 \pi^D}{\partial r \partial b} \geq 0$ if and only if $0 \leq r \leq 1$ for $b \leq \frac{1}{2}p\left(1+\delta\right)$ and $\frac{2b-p(1+\delta)}{2(b-\delta p)} \leq r \leq 1$ for $b > \frac{1}{2}p\left(1+\delta\right)$, and $\frac{\partial^2 \pi^D_{\text{UT}}}{\partial r \partial b} < 0$ otherwise.

Proof of Proposition 8

(i) An H-type agent takes action above $\bar{a}$ if and only if $\beta_H(1) + b \geq \bar{a}$, where $\beta_H(1) = \beta(1) + e$. We substitute $\frac{p}{p+(1-p)\pi(1|0)}$ for $\beta(1)$ to obtain that $\beta_H(1) + b \geq \bar{a}$ is equivalent to $\pi\left(1 \mid 0\right) \leq \pi^H$, where

$$\pi^H = \frac{p}{1-p}\frac{1-(\bar{a}-b-e)}{\bar{a}-b-e}. \tag{47}$$

The derivative with respect to e is $\frac{\partial \pi^H}{\partial e} > 0$.

(ii) An L-type agent takes action above $\bar{a}$ if and only if $\beta_L(1) \geq \bar{a}$, where $\beta_L(1) = \beta(1) - e$. We substitute $\frac{p}{p+(1-p)\pi(1|0)}$ for $\beta(1)$ to obtain that $\beta_L(1) \geq \bar{a}$ is equivalent to $\pi(1 \mid 0) \leq \pi^L$, where

$$\pi^L = \frac{p}{1-p}\frac{1-(\bar{a}+e)}{\bar{a}+e}. \tag{48}$$

The derivative with respect to e is $\frac{\partial \pi^L}{\partial e} < 0$.
(iii) We derive

$$\begin{aligned}
\widetilde{U}_L\left(\mathbf{a_0}(1) \mid \{1, H\}\right) &= \left(-\left(\beta_H(1) + b - \beta_L(1)\right)^2\right) \times \\
&\quad \times \frac{1 - r\delta}{(1 - r\delta)^2 - (1 - r)^2\delta^2}, \qquad (49) \\
\widetilde{U}_L\left(\mathbf{a_0}(\emptyset) \mid \{1, H\}\right) &= \left(-\left(\beta_L(1) - p - b\right)^2 - \frac{(1 - r)\delta}{1 - r\delta}\left(\beta_L(1) - p\right)^2\right) \times \\
&\quad \times \frac{1 - r\delta}{(1 - r\delta)^2 - (1 - r)^2\delta^2}. \qquad (50)
\end{aligned}$$

Then, $\widetilde{U}_L\left(\mathbf{a_0}(1) \mid \{1, H\}\right) \geq \widetilde{U}_L\left(\mathbf{a_0}(\emptyset) \mid \{1, H\}\right)$ if and only if

$$\left(\beta_L(1) - p - b\right)^2 + \frac{(1 - r)\delta}{1 - r\delta}\left(\beta_L(1) - p\right)^2 - \left(\beta_H(1) + b - \beta_L(1)\right)^2 \geq 0, \qquad (51)$$

which—given that $\beta_L(1) = \beta(1) - e$ and $\beta_H(1) = \beta(1) + e$—can be rearranged to

$$\left(\beta(1) - e - p - b\right)^2 + \frac{(1 - r)\delta}{1 - r\delta}\left(\beta(1) - e - p\right)^2 - b^2 \geq 0, \qquad (52)$$

which holds if and only if

$$\beta(1) \geq p + 2b\frac{1 - \delta r}{1 + \delta - 2\delta r} + e. \qquad (53)$$

We substitute $\frac{p}{p + (1-p)\pi(1|0)}$ for $\beta(1)$ to obtain that (53) is equivalent to $\pi\left(1 \mid 0\right) \leq \pi^D$, where

$$\pi^D = \frac{p}{1 - p}\frac{1 - \left(p + 2b\frac{1-\delta r}{1+\delta-2\delta r} + e\right)}{p + 2b\frac{1-\delta r}{1+\delta-2\delta r} + e}. \qquad (54)$$

The derivative with respect to e is $\frac{\partial \pi^D}{\partial e} < 0$.

(iv) Using the notation from Proof of Proposition 6, we derive

$$\begin{aligned}\widetilde{U}_{L,L}\left(\mathbf{a_0}(1)\mid 1\right) &= \left(1-r+\frac{(1-r)\,\delta}{1-r\delta}r\right)\left(-\left(\beta_H(1)+b-\beta_L(1)\right)^2\right)\times \\ &\quad\times\frac{1-r\delta}{(1-r\delta)^2-(1-r)^2\delta^2}, \qquad (55)\\ \widetilde{U}_{L,L}\left(\mathbf{a_0}(\emptyset)\mid 1\right) &= [r\left(-\left(\beta_L(1)-p\right)^2\right)+(1-r)\left(-\left(\beta_L(1)-p-b\right)^2\right)+ \\ &\quad+\frac{(1-r)\,\delta}{1-r\delta}\left(r\left(-\left(\beta_L(1)-p-b\right)^2\right)+(1-r)\left(-\left(\beta_L(1)-p\right)^2\right)\right)]\times \\ &\quad\times\frac{1-r\delta}{(1-r\delta)^2-(1-r)^2\delta^2}. \qquad (56)\end{aligned}$$

Then, $\widetilde{U}_{L,L}\left(\mathbf{a_0}(1)\mid 1\right)\geq\widetilde{U}_{L,L}\left(\mathbf{a_0}(\emptyset)\mid 1\right)$ if and only if

$$\begin{aligned}(1-r)\left(\left(\beta_L(1)-p-b\right)^2-\left(\beta_H(1)+b-\beta_L(1)\right)^2\right)+ & \\ +\left(r+\delta-2\delta r\right)\left(\beta_L(1)-p\right)^2 &\geq 0 \qquad (57)\end{aligned}$$

which—given that $\beta_L(1)=\beta(1)-e$ and $\beta_H(1)=\beta(1)+e$—can be rearranged to

$$(1-r)\left(\left(\beta(1)-e-p-b\right)^2-b^2\right)+\left(r+\delta-2\delta r\right)\left(\beta(1)-e-p\right)^2\geq 0, \qquad (58)$$

which holds if and only if

$$\beta(1)\geq p+2b\frac{1-r}{1+\delta-2\delta r}+e. \qquad (59)$$

We substitute $\frac{p}{p+(1-p)\pi(1|0)}$ for $\beta(1)$ to obtain that (59) is equivalent to $\pi\left(1\mid 0\right)\leq\pi^D_{\mathrm{UT}}$, where

$$\pi^D_{\mathrm{UT}}=\frac{p}{1-p}\frac{1-\left(p+2b\frac{1-r}{1+\delta-2\delta r}+e\right)}{p+2b\frac{1-r}{1+\delta-2\delta r}+e}. \qquad (60)$$

The derivative with respect to e is $\frac{\partial\pi^D_{\mathrm{UT}}}{\partial e}<0$.