Essays in Information Economics

Bartosz Redlicki
Trinity Hall
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy in Economics

October 2017
“Essays in Information Economics”

Bartosz Redlicki

Summary

This thesis consists of three essays in the field of information economics. The first essay studies manipulation of information by partisan media. The recent increase in partisan media has generated interest in what drives media outlets to become more partisan. I develop a model to study the role of diffusion of information by word of mouth. In the model, a media outlet designs an information policy, which specifies the level of partisan slant in the outlet’s news reports. The news spread via a communication chain in a population of agents with heterogeneous preferences. The 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.

The second essay, co-authored by Jakub Redlicki, investigates falsification of scientific evidence by interest groups. We analyse a game between a biased sender (an interest group) and a decision maker (a policy maker) where the former can falsify scientific evidence at a cost. The sender observes scientific evidence and knows that it will also be observed by the decision maker unless he falsifies it. If he falsifies, then there is a chance that the decision maker observes the falsified evidence rather than the true scientific evidence. First, we investigate the decision maker’s incentives to privately acquire independent evidence, which not only provides additional information to her but can also strengthen or weaken the sender’s falsification effort. Second, we analyse the decision maker’s incentives to acquire information from the sender.

The third essay analyses competition between interest groups for access to a policy maker. I study a model of lobbying in which two privately-informed experts (e.g., interest groups) with opposite goals compete for the opportunity to communicate with a policy maker. The main objective is to analyse the benefits which competition for access brings to the policy maker as opposed to hiring an expert in advance. I show that competition for access is advantageous in that it provides the policy maker with some information about the expert who did not gain access and gives the experts an incentive to invest in their communication skills. On the other hand, hiring an expert in advance allows the policy maker to use a monetary reward to incentivise the expert to invest more in his communication skills.
Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed limit of 60,000 words including tables, footnotes, bibliography, and appendices.
Acknowledgements

Many people supported me during the process of preparing this thesis. First of all, I would like to express my immense gratitude to my supervisor, Sönje Reiche, for her guidance, enthusiasm, patience, and valuable feedback over the course of my work on this thesis. Second, I am grateful to my advisor, Edoardo Gallo, for his very helpful comments and suggestions.

One chapter of this thesis is co-authored by Jakub Redlicki—it is a fortunate coincidence that I found a co-author who has so much in common with me. For their generous feedback, I would like to thank the members of the microeconomic theory group at the Faculty of Economics: Matthew Elliott, Aytek Erdil, Robert Evans, Sanjeev Goyal, Hamid Sabourian, and Flavio Toxvaerd. Further, I have benefited from conversations with visitors of the Faculty of Economics and the Cambridge-INET Institute, as well as with the people I met during my visit at the Toulouse School of Economics, at summer schools at the Hebrew University of Jerusalem and the University of Tokyo, and at various conferences. I would also like to take this opportunity to thank Outi Aarnio and Martin Slater, my economics tutors during my undergraduate studies at St Edmund Hall at the University of Oxford.

I gratefully acknowledge the financial support from the Economic and Social Research Council, the Faculty of Economics, and Trinity Hall. I also wish to thank the administrative staff at the Faculty of Economics, especially Silvana Dean and Louise Cross, for their help.

For distracting me from this thesis in the DAE office, on a volleyball court, a tennis court or elsewhere, I thank my friends, mostly from the PhD programme, the Cambridge University Polish Society, and the Cambridge University Volleyball Club, including Anil Ari, Krzysztof Bar, Paweł Gola, Axel Gottfries, Hanns Hagen Götzke, Claudia Herresthal, Alice Kuegler, Maciej Lisik, Karol Łukasik, David Minarsch, Monica Petrescu, Margit Reischer, Mikhail Safronov, Piotr Wieprzowski, Tomasz Zawisza, and others.

Finally, and most importantly, I would like to thank my parents and my grandparents for the opportunities they gave me and for their love, encouragement, and unwavering support. I dedicate this thesis to them.
Abstract

This thesis consists of three essays in the field of information economics. The first essay studies manipulation of information by partisan media. The recent increase in partisan media has generated interest in what drives media outlets to become more partisan. I develop a model to study the role of diffusion of information by word of mouth. In the model, a media outlet designs an information policy, which specifies the level of partisan slant in the outlet’s news reports. The news spread via a communication chain in a population of agents with heterogeneous preferences. The 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 impacts of social media and the fact that people with different political views tend to trust different media outlets.

The second essay, which is co-authored by Jakub Redlicki and is a part of his DPhil thesis at the University of Oxford, investigates falsification of scientific evidence by interest groups. We analyse a game between a biased sender (an interest group) and a decision maker (a policy maker) where the former can falsify scientific evidence at a cost. The sender observes scientific evidence and knows that it will also be observed by the decision maker unless he falsifies it. If he falsifies, then there is a chance that the decision maker observes the falsified evidence rather than the true scientific evidence. First, we investigate the decision maker’s incentives to privately acquire independent evidence, which not only provides additional information to her but can also strengthen or weaken the sender’s falsification effort. We characterise the circumstances under which the benefit from the additional information is boosted, unaffected, dampened, and fully offset by the adjustment in the sender’s falsification strategy. Second, we analyse the decision maker’s incentives to acquire information from the sender. We show that she may be better off by committing to pay less than full attention to the sender as this can discourage falsification.

The third essay analyses competition between interest groups for access to a policy maker. I study a model of lobbying in which two privately-informed experts (e.g., interest groups) with opposite goals compete for the opportunity to communicate with a policy maker. The main objective is to analyse the benefits which competition for
access brings to the policy maker as opposed to hiring an expert in advance. I show that competition for access is advantageous in that it provides the policy maker with some information about the expert who did not gain access and gives the experts an incentive to invest in their communication skills. On the other hand, hiring an expert in advance allows the policy maker to use a monetary reward to incentivise the expert to invest more in his communication skills. If the experts’ payoffs are highly dependent on the implemented policy, then the policy maker is better off having the experts compete for access; otherwise, she is better off hiring an expert in advance.
## Contents

1 Spreading Lies

1.1 Introduction ................................................. 1
1.2 Model ......................................................... 8
1.3 Equilibrium in the Communication Chain ................... 12
1.4 Optimal Information Policy ................................. 15
   1.4.1 Preliminaries ........................................ 15
   1.4.2 Two Effects: on Persuasion and on Diffusion ........ 16
   1.4.3 Characterisation of the Optimal Information Policy .... 20
   1.4.4 Graphical Interpretation ............................ 27
1.5 Extensions .................................................... 31
   1.5.1 Diffusion with Unobservable Types of Successors .... 31
   1.5.2 Misestimation of the Slant by the Agents ............ 33
1.6 Conclusion .................................................... 35
Appendix to Chapter 1 ......................................... 37
   A.1 Proofs ...................................................... 37

2 Persuasion by Falsification of Evidence, and Measures Against It ..... 45

2.1 Introduction ................................................... 45
2.2 Baseline Model ............................................... 52
   2.2.1 Equilibrium ........................................... 53
   2.2.2 Decision Maker’s Welfare ............................ 55
   2.2.3 Other Models of Strategic Communication ............ 55
2.3 Acquisition of Private Independent Evidence ................. 57
   2.3.1 Impact of Private Independent Evidence on Equilibrium ........................................................................ 58
   2.3.2 How Is Falsification Affected? .......................... 61
   2.3.3 How Is the Decision Maker’s Welfare Affected? .... 63
Chapter 1

Spreading Lies

1.1 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 1.2 presents the model. Section 1.3 characterises the equilibrium in the communication chain for a given information policy. Section 1.4 analyses the optimal information policy for the manipulator. Section 1.5 considers the extensions of the model. Section 1.6 concludes.

**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.

\(^1\)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).
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.

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.” 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.

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

---

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

3https://today.yougov.com/news/2017/06/20/Americans-agree-media-is-biased/
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%). 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 it, while 56% of consistent liberals trust it.

**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 Zhu.

---

4Pew Research Center, October 2014, “Political Polarization and Media Habits”.
6Pew Research Center, October 2014, “Political Polarization and Media Habits”.
ravskaya, 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 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.

1.2 Model

Players. There is a manipulator, \( M \), and an infinite population of agents, \( N = \{1, 2, 3, \ldots \} \). 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 \( \pi \), 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 \( \pi \).

2. The state of the world \( \omega \in \{0, 1\} \) is realised, with the signal realisation \( s \) deter-
ned according to \( \pi \).

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 $1 - r$.

• 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 $\pi$ 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.

\footnote{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.}
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.

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 \( a = \{a(1), a(2), a(3), \ldots\} \), where \( a(i) \) denotes the action of the agent in \( i \)th position in a communication chain of infinite length.

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

\[
v(a(i)) = \begin{cases} 1 & \text{if } a(i) \geq \bar{a} \\ 0 & \text{if } a(i) < \bar{a} \end{cases}, \tag{1.1}
\]

where \( \bar{a} \) is an exogenous threshold. If an agent takes an action weakly above \( \bar{a} \), 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 \( a \) is written as

\[
V(a) = \sum_{i=1}^{\infty} \delta^{i-1} v(a(i)), \tag{1.2}
\]

\(^8\)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.
where $\delta = 1 - \epsilon$ is the discount factor that reflects the fact that the communication chain breaks with probability $\epsilon$ at each meeting.

When choosing his information policy $(\pi(1 \mid 1), \pi(1 \mid 0))$, the manipulator maximises the expected value of $V(a)$ 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, $a_1, a_2, a_3, \ldots$ Fix an infinite sequence of actions $a_0 = \{a_0, a_1, a_2, \ldots\}$, where $a_i$ denotes the action of the agent’s $i$th 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 $i = 0, 1, 2, \ldots$) in state $\omega$ 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 $a_0 = \{a_0, a_1, a_2, \ldots\}$ in state $\omega$ (conditional on her meeting her successor) is

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

(1.3)

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 $U_{t_0}(a_0, \omega)$ 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.\(^\text{10}\)

\(^\text{9}\)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.

\(^\text{10}\)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.
1.3 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, \((\mu, \sigma, \beta)\), 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 1.1 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 \(\beta(\emptyset) = p\), and hence their action strategy is trivially \(\sigma_L(\emptyset) = p\) and \(\sigma_H(\emptyset) = p + b\).

**Proposition 1.1.** In the unique equilibrium:

(i) the agents’ beliefs are

\[
\beta(s) = \begin{cases} 
\frac{p\pi(0\mid1)}{p\pi(0\mid1)+(1-p)\pi(0\mid0)} & \text{for } s = 0 \\
\frac{p\pi(1\mid1)}{p\pi(1\mid1)+(1-p)\pi(1\mid0)} & \text{for } s = 1
\end{cases},
\]

(ii) the agents’ action strategies are

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

(iii) the agents’ communication strategies are

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

\(^{11}\)I assume that if an agent is indifferent between passing \(s\) on and not, then she chooses to pass it on.
where $\beta(1)^* = p + 2b \frac{1-\delta r}{1+\delta-2br}$, and

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

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

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

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

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

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

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 $\sigma_L(s) = \arg\max_{a_0} u_L(a_0, \omega \mid \beta(s)) = \beta(s)$ and $\sigma_H(s) = \arg\max_{a_0} u_H(a_0, \omega \mid \beta(s)) = \beta(s) + b$.

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 1.1 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.

![Figure 1.1: 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.](image)

For simplicity, Figure 1.1 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
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 1.1 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 1.4.

1.4 Optimal Information Policy

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

1.4.1 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. \( \arg\max_{a_0} u_{t_0}(a_0, \omega \mid p) < \bar{a} \) 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. 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 \( \arg\max_{a_0} u_H(a_0, \omega \mid p) = p + b \) for an \( H \)-type agent and by \( \arg\max_{a_0} u_L(a_0, \omega \mid p) = p \) 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 \( good \).

---

\(^{12}\)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.
The manipulator has an incentive to increase $\pi(1 | 1)$ 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, $\pi(1 | 0)$ may be weakly positive in the optimal information policy. Yet, $\pi(1 | 0) = 1$ cannot be optimal because, combined with $\pi(1 | 1) = 1$, it would make the signal uninformative and so—given Assumption 1—all agents would remain unpersuaded upon observing $s = 1$. The conditional probability $\pi(1 | 0) \in [0, 1)$ then fully specifies any optimal information policy.

**Definition 1.1.** The slant of the information policy is defined as $\pi(1 | 0)$, 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 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).

**1.4.2 Two Effects: on Persuasion and on Diffusion**

As already mentioned in Section 1.3, 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 1.2. 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
\[ \pi^H = \frac{p}{1 - p} \frac{1 - (\bar{a} - b)}{\bar{a} - b}, \] (1.6)
(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
\[ \pi^L = \frac{p}{1 - p} \frac{1 - \bar{a}}{\bar{a}}. \] (1.7)
The thresholds $\pi^H$ and $\pi^L$ are increasing in prior belief $p$ and decreasing in $\bar{a}$, and $\pi^H$ is increasing in polarisation $b$.

Proposition 1.2 uses the equilibrium action strategies described in Proposition 1.1 and the observation that $\pi(1 \mid 1) = 1$ must hold in the optimal policy. Given (1.4), (1.5), and $\pi(1 \mid 1) = 1$, the posterior belief about $\omega = 1$ upon receiving $s = 1$ is
\[ \beta(1) = \frac{p}{p + (1 - p) \pi(1 \mid 0)}, \] (1.8)
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 $\pi^H$ is given by (1.6). 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 $\pi^L$ is described by (1.7). It follows easily that $\pi^H > \pi^L$.

Hence, the message behind Proposition 1.2 is that the manipulator must choose a low enough slant in order to be able to persuade the agents. The thresholds $\pi^H$ and $\pi^L$ 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 1.3. 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
\[ \pi^D = \frac{p}{1-p} \frac{1 - (p + 2b \frac{1 - \delta r}{1 + \delta - 2\delta r})}{p + 2b \frac{1 - \delta r}{1 + \delta - 2\delta r}}. \] (1.9)

The threshold $\pi^D$ is decreasing in assortativity $r$, $\frac{\partial \pi^D}{\partial r} < 0$, and decreasing in polarisation $b$, $\frac{\partial^2 \pi^D}{\partial r \partial b} < 0$ for sufficiently low $r$ and $b$, and $\frac{\partial^2 \pi^D}{\partial r \partial b} \geq 0$ otherwise.

Proposition 1.3 uses the equilibrium communication strategies in Proposition 1.1 and the observation that $\pi(1 \mid 1) = 1$ in the optimal information policy. As stated in Proposition 1.1, 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 $\pi^D$ is given by (1.9).

The message behind Proposition 1.3 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 1.2 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 1.2: 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 \( \pi^D \) 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 \( \pi^D \) 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 \( \pi^D \). 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 \( \pi^D \).

**Relation between the two effects.** The relation between the values of thresholds \( \pi^H, \pi^L \) and \( \pi^D \) 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 1.4 describes how the relation between \( \pi^H, \pi^L \) and \( \pi^D \) depends on polarisation \( b \) and on assortativity \( r \).

**Proposition 1.4.** (a) As far as assortativity \( r \) is concerned, the relation between \( \pi^D, \pi^L \) and \( \pi^H \) is:

(i) \( \pi^D < \pi^L \) if and only if \( r > r^\ast \ast \),

(ii) \( \pi^L \leq \pi^D < \pi^H \) if and only if \( r^\ast < r \leq r^\ast \ast \),

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

where \( r^\ast \) and \( r^\ast \ast \) are functions of other parameters, and \( r^\ast < r^\ast \ast \).

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

(i) \( \pi^D < \pi^L \) if and only if \( b > b^\ast \ast \),

(ii) \( \pi^L \leq \pi^D < \pi^H \) if and only if \( b^\ast < b \leq b^\ast \ast \),

(iii) \( \pi^D \geq \pi^H \) if and only if \( b \leq b^\ast \),
where $b^*$ and $b^{**}$ are functions of other parameters, and $0 < b^* < b^{**} < \pi - p$.

Proposition 1.4 uses the thresholds $\pi^H$, $\pi^L$ and $\pi^D$ derived in Propositions 1.2 and 1.3. Part (a) follows from the fact that $\pi^D$ is decreasing in $r$, while both $\pi^L$ and $\pi^H$ do not depend on $r$. Part (b) follows from the fact that $\pi^D$ is decreasing in $b$, $\pi^L$ does not depend on $b$, and $\pi^H$ 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 $(0,\pi - p)$ regardless of the values of other parameters.

Overall, Proposition 1.4 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.

### 1.4.3 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 $\pi^D$ increases above $\pi^L$, 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.
Naturally, the slant in the optimal information policy cannot be higher than \( \pi^H \), 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 \( \mathbb{E}[V(a) \mid s = 1] \). The expressions for \( \mathbb{E}[V(a) \mid s = 1] \) 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

\[
V^D = \sum_{i=1}^{\infty} \delta^{i-1} = 1 + \frac{\delta}{1 - \delta},
\]

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

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

Table 1.1: Persuasion and diffusion under various levels of slant in the case \( \pi^D < \pi^L \).

<table>
<thead>
<tr>
<th>Slant of the information policy, ( \pi(1 \mid 0) )</th>
<th>Are ( H )-type agents persuaded by ( s = 1 )?</th>
<th>Are ( L )-type agents persuaded by ( s = 1 )?</th>
<th>Do ( L )-type agents pass ( s = 1 ) on to ( H )-type agents?</th>
</tr>
</thead>
<tbody>
<tr>
<td>([0, \pi^D])</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>((\pi^D, \pi^L])</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>((\pi^L, \pi^H])</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>((\pi^H, 1])</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
</tbody>
</table>
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

\[
V^H = \sum_{i=1}^{\infty} \delta^{i-1} r^{i-1} = 1 + \frac{\delta r}{1 - \delta r},
\]

which follows from the fact that agents pass \( s = 1 \) on and are persuaded by it as long as only \( H \)-type 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 \( \Pr(s = 1)E[V(a) \mid s = 1] \), where both \( \Pr(s = 1) \) and \( E[V(a) \mid s = 1] \) 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 \( E[V(a) \mid s = 1] \) 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 \( \pi^D \), \( \pi^L \), or \( \pi^H \): 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) \( \pi^D \) (which corresponds to the mainstream policy), (ii) \( \pi^L \) (the intermediate policy), and (iii) \( \pi^H \) (the partisan policy).

Proposition 1.5 characterises the optimal information policy in the context of two key characteristics of the environment: assortativity \( r \) and polarisation \( b \).\(^{[13]} \)

\(^{[13]} \) I assume that if the manipulator is indifferent between information policies, then he chooses the one with a lower slant.
Proposition 1.5. (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 1.5 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 $\pi^H$, whereas here lower slants of $\pi^L$ and $\pi^D$ can be optimal too.

The mainstream information policy, i.e. slant $\pi^D$, 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 $\pi^H$. 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 $\pi^L$, 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 1.3 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.\footnote{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 $\delta = 0.2$, $p = 0.1$, $\pi = 0.12$. These values are calibrated so that all three policies (mainstream, intermediate and partisan) are optimal for some values of $r$ and $b$.}

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 1.1.** (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 $r > \min\{r_{DL}, r_{DH}\}$.

(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 1.5. 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
Figure 1.3: 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 \(\delta = 0.2\), \(p = 0.1\), \(\overline{\pi} = 0.12\).

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 1.4 shows the expected diffusion of \(s = 1\) as a function of assortativity for fixed values of other parameters\(^\text{15}\).

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 \(\pi^L\) and \(\pi^H\) correspond respectively to the mainstream policy and the partisan policy, which are defined the same way as earlier.

\(^{15}\)Other parameters are fixed at \(\delta = 0.2\), \(p = 0.1\), \(\overline{\pi} = 0.12\) (same as in Figure 1.3), and \(b = 0.013\).
Figure 1.4: The non-monotonic relation between the expected diffusion of $s = 1$ and the assortativity of the population, with other parameters fixed at $\delta = 0.2$, $p = 0.1$, $\pi = 0.12$, $b = 0.013$.

However, the slant $\pi^D$ 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. 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 $\pi^L$ (which corresponds to the mainstream policy) and $\pi^H$ (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.

---

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.
in two opposite ways: the required slant $\pi^D$ 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.

1.4.4 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 $\beta$, be denoted by $E_\beta [V(a^*(\beta))]$ with $a^*(\beta)$ denoting the equilibrium actions of agents in the communication chain given that the first agent has posterior belief $\beta$. Naturally, if the posterior belief is $\beta < p$, then the interim expected payoff is 0. Given that $\pi (1 \mid 1) = 1$ 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 $\pi (1 \mid 1) = 1$, the posterior belief for $s = 1$, denoted by $\beta(1)$, is determined by the slant, $\pi (1 \mid 0)$. For the case $\pi^D < \pi^L$, there are thresholds $\beta^D$, $\beta^L$, and $\beta^H$ such that the posterior belief $\beta(1)$ is:

(i) $\beta(1) \geq \beta^D$ if and only if $\pi (1 \mid 0) \leq \pi^D$;
(ii) $\beta^L \leq \beta(1) < \beta^D$ if and only if $\pi^D < \pi (1 \mid 0) \leq \pi^L$;
(iii) $\beta^H \leq \beta(1) < \beta^L$ if and only if $\pi^L < \pi (1 \mid 0) \leq \pi^H$; and
(iv) $p \leq \beta(1) < \beta^H$ if and only if $\pi (1 \mid 0) > \pi^H$.

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 $p \leq \beta(1) < \beta^H$. 

27
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 \( \hat{V}(\beta) = E_\beta[V(a^*(\beta))] \), i.e. if the posterior belief of the first agent in the communication chain is \( \beta \), 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 \( \tau \) such that the expected posterior under this distribution equals the prior, i.e. \( E_\tau[\beta(s)] = p \), there exists a signal \( \pi \) which, given the prior \( p \), induces the distribution of posteriors \( \tau \). This allows us to express the manipulator’s problem as

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

i.e. we can simply look for the optimal distribution of posteriors such that \( E_\tau[\beta] = p \). 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 \( \beta^H, \beta^L, \beta^D, V^H, V^L \), and \( V^D \). That shape determines the concavification \( \mathbf{V} \) and hence the optimal signal.

Figures 1.5, 1.6, and 1.7 illustrate the manipulator’s interim expected payoff \( \hat{V} \) and its concavification \( \mathbf{V} \) as functions of \( \beta \) for three different combinations of parameter values.

In Figure 1.5, high polarisation \( b \) keeps \( \beta^H \) far away to the left from \( \beta^L \), as well as \( \beta^D \) far away to the right from \( \beta^L \). 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 \( \beta^H \). Having identified the induced posterior beliefs, it is straightforward to derive the unique signal that induces these beliefs: it is given by
π(1 | 1) = 1 and π(1 | 0) = π^H. We conclude that the optimal slant is π(1 | 0) = π^H.

In Figure 1.6, 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 π(1 | 1) = 1 and π(1 | 0) = π^L. The optimal slant is thus π(1 | 0) = π^L.

In Figure 1.7, 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 π(1 | 1) = 1 and π(1 | 0) = π^D. Hence, the optimal slant is π(1 | 0) = π^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 ≤ β < β^H and 1 for β^H ≤ β ≤ 1. The posterior beliefs induced by the optimal signal would then be 0 and β^H, and so the optimal slant would be π(1 | 0) = π^H.

Figure 1.5: An illustration of the manipulator’s interim expected payoff \( \hat{V} \) and its concavification \( V \) such that the optimal slant is \( \pi(1 \mid 0) = \pi^H \).
Figure 1.6: An illustration of the manipulator’s interim expected payoff $\hat{V}$ and its concavification $V$ such that the optimal slant is $\pi(1 \mid 0) = \pi^L$.

Figure 1.7: An illustration of the manipulator’s interim expected payoff $\hat{V}$ and its concavification $V$ such that the optimal slant is $\pi(1 \mid 0) = \pi^D$. 
1.5 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 discuss 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.

1.5.1 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 $\mu_{t_0} : \{s\} \rightarrow \{s, \emptyset\}$, where $\mu_{t_0}(s) = s$ denotes passing $s$ on and $\mu_{t_0}(s) = \emptyset$ 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 $1 - r$.

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

**Proposition 1.6.** In the unique equilibrium, the agents’ beliefs and action strategies are the same as in Proposition 1.1, and the agents’ communication strategies are

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

\footnote{I use subscript “UT” (for “Unobservable Types”) to distinguish the notation in this extension from the one in the main setup.}
where $\beta(1)_{UT}^* = p + 2b\frac{1-r}{1+\delta-2r\delta}$, and

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

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

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 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 1.2. However, the effect on diffusion is now somewhat different. It is described in the following proposition.

**Proposition 1.7.** 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 $\pi(1 \mid 0) \leq \pi^D_{UT}$, where

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

(1.14)

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

Proposition 1.7 uses the equilibrium communication strategies in Proposition 1.6 and
the observation that \( \pi(1 \mid 1) = 1 \) must hold in the optimal information policy. As stated in Proposition 1.6, an \( L \)-type agent passes \( s = 1 \) on if and only if \( \beta(1) > p + 2b \frac{1-r}{1+4-2r\delta} \). Given that \( \pi(1 \mid 1) = 1 \), this condition is equivalent to \( \pi(1 \mid 0) \leq \pi^{D}_{UT} \), where \( \pi^{D}_{UT} \) is given by (1.14).

The main difference between the setups with unobservable and observable types of successors is that the threshold \( \pi^{D}_{UT} \) is increasing in assortativity \( r \), whereas \( \pi^{D} \) 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 \( \pi^{D}_{UT} \geq \pi^{H} \) holds, which follows from the fact that \( \pi^{H} \) does not depend on \( r \) and \( \pi^{D}_{UT} \) 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.

### 1.5.2 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 1.8 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 1.8.** 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 $\pi^H$ 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 $\pi^L$ 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 $\pi^D$ 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 $\pi(1 \mid 0) \leq \pi^D_{UT}$, where $\pi^D_{UT}$ 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 1.8 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 1.8: 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 \( \pi^D (\pi^D_{UT}) \) and increasing \( \pi^H \), it can make the relation between \( \pi^D (\pi^D_{UT}) \) and \( \pi^H \) be \( \pi^D < \pi^H \) \( (\pi^D_{UT} < \pi^H) \). Then, a slant \( \pi (1 \mid 0) = \pi^H \) corresponds to a partisan policy. Second, by decreasing both \( \pi^D (\pi^D_{UT}) \) and \( \pi^L \) while increasing \( \pi^H \), 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.

1.6 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.
Appendix to Chapter 1

A.1 Proofs

Proof of Proposition 1.1

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

\[
 u_H(\beta(1) + b, \omega_i | \beta(1)) > u_H(p + b, \omega_i | \beta(1)),
\]

(1.15)

\[
 u_H(\beta(1), \omega_i | \beta(1)) > u_H(p, \omega_i | \beta(1)).
\]

(1.16)

Given (1.15) and (1.16), for any \( \mu_L(1, t_1) \), a sequentially rational strategy must have

\[
 \mu_H(1, H) = 1 \quad \text{and} \quad \mu_H(1, L) = 1.
\]

From the payoff functions it also follows that

\[
 u_L(\beta(1), \omega_i | \beta(1)) > u_L(p, \omega_i | \beta(1)),
\]

(1.17)

Given (1.17), 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 \( \tilde{U}_L(a_0(1) | \{1, t_1\}) \) 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 \( \tilde{U}_L(a_0(0) | \{1, t_1\}) \) 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

\[
 \tilde{U}_L(a_0(1) | \{1, H\}) \geq \tilde{U}_L(a_0(0) | \{1, H\}),
\]

then a sequentially rational strategy must have \( \mu_L(1, H) = 1 \); otherwise it must have \( \mu_L(1, H) = 0 \).

We derive \( \tilde{U}_L(a_0(1) | \{1, H\}) \) by solving the simultaneous equations:

\[
 \tilde{U}_L(a_0(1) | \{1, H\}) = -b^2 + r\delta \tilde{U}_L(a_0(1) | \{1, H\}) + (1 - r)\delta \tilde{U}_L(a_0(1) | \{1, L\})
\]

(1.18)

\[
 \tilde{U}_L(a_0(1) | \{1, L\}) = r\delta \tilde{U}_L(a_0(1) | \{1, L\}) + (1 - r)\delta \tilde{U}_L(a_0(1) | \{1, H\}).
\]

(1.19)
We obtain
\[ \tilde{U}_{L}(a_0(1) \mid \{1, H\}) = \left(-b^2\right) \frac{1 - r\delta}{(1 - r\delta)^2 - (1 - r)^2\delta^2}. \] (1.20)

Similarly, we derive \( \tilde{U}_{L}(a_0(0) \mid \{1, H\}) \) by solving the simultaneous equations:
\[
\tilde{U}_{L}(a_0(0) \mid \{1, H\}) = -(\beta(1) - p - b)^2 + r\delta \tilde{U}_{L}(a_0(0) \mid \{1, L\}) + (1 - r)\delta \tilde{U}_{L}(a_0(0) \mid \{1, H\}) + (1 - r)\delta \tilde{U}_{L}(a_0(0) \mid \{1, H\}). \]
(1.21)

We obtain
\[
\tilde{U}_{L}(a_0(0) \mid \{1, H\}) = \left(-\beta(1) - p - b\right)^2 - r\delta \tilde{U}_{L}(a_0(0) \mid \{1, L\}) + (1 - r)\delta \tilde{U}_{L}(a_0(0) \mid \{1, H\}). \]
(1.22)

Then, \( \tilde{U}_{L}(a_0(1) \mid \{1, H\}) \geq \tilde{U}_{L}(a_0(0) \mid \{1, H\}) \) if and only if
\[
(\beta(1) - p - b)^2 + (1 - r)\delta (\beta(1) - p)^2 - b^2 \geq 0, \]
(1.24)
which can be rearranged to
\[
\beta(1) \geq p + 2b \frac{1 - \delta r}{1 + \delta - 2\delta r}. \]
(1.25)

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 \( \tilde{U}_H(a_0(0) \mid \{0, L\}) \geq \tilde{U}_H(a_0(0) \mid \{0, H\}) \) if and only if
\[
(p - \beta(0) - b)^2 + (1 - r)\delta (p - \beta(0))^2 - b^2 \geq 0, \]
(1.26)
which can be rearranged to
\[
\beta(0) \leq p - 2b \frac{1 - \delta r}{1 + \delta - 2\delta r}. \]
(1.27)
Proof of Proposition 1.2

The proof follows from the main text.

Proof of Proposition 1.3

The proof follows from the main text. To see that \( \partial \pi_D / \partial r < 0 \) and \( \partial \pi_D / \partial b < 0 \), 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 1.4

(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

\[
\begin{align*}
 b^* &= (\bar{a} - p) \left( 1 + \frac{\delta - 2\delta r}{3 + \delta - 4\delta r} \right), \quad (1.28) \\
 b^{**} &= (\bar{a} - p) \left( 1 + \frac{\delta - 2\delta r}{2 - 2\delta r} \right). \quad (1.29)
\end{align*}
\]

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

Proof of Proposition 1.5

Let \( \tilde{V}(\pi(1 \mid 0)) \) denote the manipulator’s expected payoff from an information policy with a slant \( \pi(1 \mid 0) \). The expected payoffs from slants \( \pi_D, \pi_L, \) and \( \pi_H \) are

\[
\begin{align*}
\tilde{V}(\pi_D) &= (p + (1 - p) \pi_D) V_D, \quad (1.30) \\
\tilde{V}(\pi_L) &= (p + (1 - p) \pi_L) V_L, \quad (1.31) \\
\tilde{V}(\pi_H) &= (p + (1 - p) \pi_H) V_H. \quad (1.32)
\end{align*}
\]
(a) The derivatives with respect to $r$ are

$$\frac{\partial \tilde{V}(\pi^D)}{\partial r} < 0,$$  \hspace{1cm} (1.33)

$$\frac{\partial \tilde{V}(\pi^L)}{\partial r} > 0,$$  \hspace{1cm} (1.34)

$$\frac{\partial \tilde{V}(\pi^H)}{\partial r} > 0.$$  \hspace{1cm} (1.35)

It follows from (1.33) and (1.34) 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 (1.33) and (1.35) 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 $r^{LH} = \frac{\delta a - (1+\delta)b}{\delta(a-2b)}$ and $\alpha > 2b$.

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 $r^{LH} = r^{DH} = r^{DL}$.

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

$$\frac{\partial \tilde{V}(\pi^D)}{\partial b} < 0,$$  \hspace{1cm} (1.36)

$$\frac{\partial \tilde{V}(\pi^L)}{\partial b} = 0,$$  \hspace{1cm} (1.37)

$$\frac{\partial \tilde{V}(\pi^H)}{\partial b} > 0.$$  \hspace{1cm} (1.38)

It follows from (1.36) and (1.37) 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 (1.36) and
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 (1.37) and (1.38) 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 $b < b^{LH}$; 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 1.1**

The proof follows from the main text.

**Proof of Proposition 1.6**

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

It is straightforward to show that $\tilde{U}_{H,H}(a_0(1) \mid 1) > \tilde{U}_{H,H}(a_0(\emptyset) \mid 1)$, 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 $\tilde{U}_{L,L}(a_0(1) \mid 1)$ by solving the simultaneous equations:

\[
\begin{align*}
\tilde{U}_{L,L}(a_0(1) \mid 1) &= (1 - r) \left(-b^2\right) + r\delta \tilde{U}_{L,L}(a_0(1) \mid 1) + \nonumber \\
&\quad + (1 - r) \delta \tilde{U}_{L,H}(a_0(1) \mid 1) , \\
\tilde{U}_{L,H}(a_0(1) \mid 1) &= r \left(-b^2\right) + r\delta \tilde{U}_{L,H}(a_0(1) \mid 1) + \nonumber \\
&\quad + (1 - r) \delta \tilde{U}_{L,L}(a_0(1) \mid 1) . 
\end{align*}
\]

We obtain

\[
\tilde{U}_{L,L}(a_0(1) \mid 1) = \left( (1 - r) \left(-b^2\right) + \frac{(1 - r) \delta}{1 - r\delta} \left(-b^2\right) \right) \frac{1 - r\delta}{(1 - r\delta)^2 - (1 - r)^2\delta^2} .
\]
Similarly, we derive \( \tilde{U}_{L,L}(a_0(0) \mid 1) \) by solving the simultaneous equations:

\[
\begin{align*}
\tilde{U}_{L,L}(a_0(0) \mid 1) &= r \left( -(\beta(1) - p)^2 \right) + (1 - r) \left( -(\beta(1) - p - b)^2 \right) + \\
&\quad + r \delta \tilde{U}_{L,L}(a_0(0) \mid 1) + (1 - r) \delta \tilde{U}_{L,H}(a_0(0) \mid 1), \\
\tilde{U}_{L,H}(a_0(0) \mid 1) &= r \left( -(\beta(1) - p - b)^2 \right) + (1 - r) \left( -(\beta(1) - p)^2 \right) + \\
&\quad + r \delta \tilde{U}_{L,H}(a_0(0) \mid 1) + (1 - r) \delta \tilde{U}_{L,L}(a_0(0) \mid 1).
\end{align*}
\]

We obtain

\[
\begin{align*}
\tilde{U}_{L,L}(a_0(0) \mid 1) &= [r \left( -(\beta(1) - p)^2 \right) + (1 - r) \left( -(\beta(1) - p - b)^2 \right) + \\
&\quad + \frac{(1 - r) \delta}{1 - r \delta} \left( r \left( -(\beta(1) - p - b)^2 \right) + (1 - r) \left( -(\beta(1) - p)^2 \right) \right)] \times \\
&\quad \times \frac{1 - r \delta}{(1 - r \delta)^2 - (1 - r)^2 \delta^2}.
\end{align*}
\]

Then, \( \tilde{U}_{L,L}(a_0(1) \mid 1) \geq \tilde{U}_{L,L}(a_0(0) \mid 1) \) if and only if

\[
\beta(1) \geq p + 2b \frac{1 - r}{1 + \delta - 2r \delta}.
\]

We repeat analogous steps to derive the equilibrium communication strategy for \( s = 0 \). It is straightforward to show that \( \tilde{U}_{L,L}(a_0(0) \mid 0) > \tilde{U}_{L,L}(a_0(0) \mid 0) \), 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 \( \tilde{U}_{H,L}(a_0(0) \mid 0) \geq \tilde{U}_{H,H}(a_0(0) \mid 0) \) if and only if

\[
\beta(0) \leq p - 2b \frac{1 - r}{1 + \delta - 2r \delta}.
\]

**Proof of Proposition 1.7**

The proof follows from the main text. To see that \( \frac{\partial \pi^D_{L,T}}{\partial r} > 0 \) and \( \frac{\partial \pi^D_{L}}{\partial b} < 0 \), note that the term \( 2b \frac{1 - r}{1 + \delta - 2r \delta} \) is decreasing in \( r \) and increasing in \( b \). In terms of \( b \), one can show that \( \frac{\partial^2 \pi^D_{L,T}}{\partial r \partial b} \geq 0 \) if and only if \( b \leq \frac{p(1 + \delta - 2\delta \rho)}{2(1 - r)} \) for \( 0 \leq r < 1 \) and \( b > 0 \) for \( r = 1 \), and \( \frac{\partial^2 \pi^D_{L,T}}{\partial r \partial b} < 0 \) otherwise. In terms of \( r \), one can show that \( \frac{\partial^2 \pi^D_{L,T}}{\partial r \partial b} \geq 0 \) if and only if \( 0 \leq r \leq 1 \) for \( b \leq \frac{1}{2} p (1 + \delta) \) and \( 2b - p(1 + \delta) \leq r \leq 1 \) for \( b > \frac{1}{2} p (1 + \delta) \), and \( \frac{\partial^2 \pi^D_{L,T}}{\partial r \partial b} < 0 \) otherwise.
Proof of Proposition 1.8

(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 $\pi(1 | 0) \leq \pi^H$, where

$$\pi^H = \frac{p}{1-p} \frac{1 - (\bar{a} - b - e)}{\bar{a} - b - e}. \quad (1.47)$$

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 $\pi(1 | 0) \leq \pi^L$, where

$$\pi^L = \frac{p}{1-p} \frac{1 - (\bar{a} + e)}{\bar{a} + e}. \quad (1.48)$$

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

(iii) We derive

$$\tilde{U}_L(a_0(1) | \{1, H\}) = \left(- (\beta_H(1) + b - \beta_L(1))^2\right) \times$$

$$\frac{1 - r\delta}{(1 - r\delta)^2 - (1 - r)^2 \delta^2}; \quad (1.49)$$

$$\tilde{U}_L(a_0(\emptyset) | \{1, H\}) = \left(- (\beta_L(1) - p - b)^2 + \frac{(1 - r)\delta}{1 - r\delta}(\beta_L(1) - p)^2\right) \times$$

$$\frac{1 - r\delta}{(1 - r\delta)^2 - (1 - r)^2 \delta^2}; \quad (1.50)$$

Then, $\tilde{U}_L(a_0(1) | \{1, H\}) \geq \tilde{U}_L(a_0(\emptyset) | \{1, H\})$ if and only if

$$\left(\beta_L(1) - p - b\right)^2 + \frac{(1 - r)\delta}{1 - r\delta}(\beta_L(1) - p)^2 - (\beta_H(1) + b - \beta_L(1))^2 \geq 0, \quad (1.51)$$

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

$$\left(\beta(1) - e - p - b\right)^2 + \frac{(1 - r)\delta}{1 - r\delta}(\beta(1) - e - p)^2 - b^2 \geq 0, \quad (1.52)$$

which holds if and only if

$$\beta(1) \geq p + 2b \frac{1 - \delta r}{1 + \delta - 2\delta r} + e. \quad (1.53)$$

We substitute $\frac{p}{p+(1-p)\pi(1|0)}$ for $\beta(1)$ to obtain that $\{1.53\}$ is equivalent to $\pi(1 | 0) \leq \pi^D$, where

43
where
\[
\pi^D = \frac{p}{1-p} \frac{1 - \left( p + 2b \frac{1-r}{1+\delta-2\delta r} + e \right)}{p + 2b \frac{1-r}{1+\delta-2\delta r} + e}.
\] (1.54)
The derivative with respect to $e$ is $\frac{\partial \pi^D}{\partial e} < 0$.

(iv) Using the notation from Proof of Proposition 1.6, we derive
\[
\tilde{U}_{L,L} (a_0(1) \mid 1) = \left( 1 - r + \frac{(1-r) \delta}{1-r\delta} r \right) \left( - (\beta_H(1) + b - \beta_L(1))^2 \right) \times \frac{1-r\delta}{(1-r\delta)^2 - (1-r)^2 \delta^2}.
\] (1.55)
\[
\tilde{U}_{L,L} (a_0(\emptyset) \mid 1) = \left[ r \left( - (\beta_L(1) - p)^2 \right) + (1-r) \left( - (\beta_L(1) - p - b)^2 \right) + (1-r) \left( - (\beta_L(1) - p)^2 \right) \right] \times \frac{1-r\delta}{(1-r\delta)^2 - (1-r)^2 \delta^2}.
\] (1.56)

Then, $\tilde{U}_{L,L} (a_0(1) \mid 1) \geq \tilde{U}_{L,L} (a_0(\emptyset) \mid 1)$ if and only if
\[
(1-r) \left( (\beta_L(1) - p - b)^2 - (\beta_H(1) + b - \beta_L(1))^2 \right) + (r + \delta - 2\delta r) (\beta_L(1) - p)^2 \geq 0
\] (1.57)
which—given that $\beta_L(1) = \beta(1) - e$ and $\beta_H(1) = \beta(1) + e$—can be rearranged to
\[
(1-r) \left( (\beta(1) - e - p - b)^2 - b^2 \right) + (r + \delta - 2\delta r) (\beta(1) - e - p)^2 \geq 0,
\] (1.58)
which holds if and only if
\[
\beta(1) \geq p + 2b \frac{1-r}{1+\delta-2\delta r} + e.
\] (1.59)
We substitute $\frac{p}{p+(1-p)\pi(1|0)}$ for $\beta(1)$ to obtain that (1.59) is equivalent to $\pi (1 \mid 0) \leq \pi^D_{UT}$, where
\[
\pi^D_{UT} = \frac{p}{1-p} \frac{1 - \left( p + 2b \frac{1-r}{1+\delta-2\delta r} + e \right)}{p + 2b \frac{1-r}{1+\delta-2\delta r} + e}.
\] (1.60)
The derivative with respect to $e$ is $\frac{\partial \pi^D_{UT}}{\partial e} < 0$. 

44
Chapter 2

Persuasion by Falsification of Evidence, and Measures Against It

2.1 Introduction

In recent decades, many interest groups have developed strategies to falsify scientific findings in order to defend their own agenda. Moreover, they have often done so while being aware that the scientific truth is most likely against them. One well-known example is that of tobacco producers, who knew about the risks which their products pose already in the 1950s, but they consistently denied adverse effects of active smoking in the 1950s and 1960s and of second-hand smoke exposure from the 1970s until the 1990s. The objective of the tobacco producers was to prevent any new regulation that would harm them or at least to delay it for as long as possible. There is also evidence of scientific evidence being falsified to counter changes in regulations on issues such as lead in petrol, asbestos, insecticides, acid rain, ozone-layer depletion, eating meat, and many others.\footnote{The many parallels between these controversies are the subject of Oreskes and Conway (2010).}  

Interest groups use various tactics to falsify scientific evidence. A common one is to develop an own research programme and publish only favourable results or simply to fabricate a research study with the desired results. An example of this is the Tobacco Institute created by the tobacco industry to foster research, but—as Oreskes and Conway (2010) report—there is evidence of scientific evidence being falsified to counter changes in regulations on issues such as lead in petrol, asbestos, insecticides, acid rain, ozone-layer depletion, eating meat, and many others.\footnote{See, for example: \url{http://www.theguardian.com/environment/planet-oz/2015/mar/05/doubt-over-climate-science-is-a-product-with-an-industry-behind-it}; \url{http://www.theguardian.com/environment/2015/feb/27/what-happened-to-lobbyists-who-tried-reshape-us-view-climate-change}.}
Conway (2010) point out—the primary objective of establishing this institution was in fact to develop “a cadre of experts who could be called upon in time of need”\(^3\). Philip Morris, one of the leading tobacco companies, was reported to have spent large amounts of money to fund scientists who would help the firm counter the emerging consensus on the health risks of second-hand tobacco smoke\(^4\).

Falsification can be effective because people are often not able to identify whether a study has been done by objective academic researchers or by an interest group. This may happen for several reasons, for example, people may have a limited amount of time to identify the source of research or they may have insufficient knowledge of how to do it. Another reason is that interest groups make a great effort to render their research hardly distinguishable from that done by objective scientists. For example, it has been reported that the tobacco industry financed conferences and workshops, as well as created newsletters, magazines, and scientific journals (e.g., Tobacco and Health, Science Fortnightly, and the Indoor Air Journal) where the results of industry-sponsored research could be reported, published, and cited, and an impression was made that these were independent (Proctor, 2012). There is even evidence of an interest group formatting an article to look like a reprint from the Proceedings of the National Academy of Sciences, which later featured in the Wall Street Journal\(^5\).

The problem is aggravated further by the fact that media often do not inform their readers, viewers, and listeners that the “experts” being quoted have links to interest groups. This may be because they feel that doing so would be “editorialising” or they simply are not aware of the connections themselves—as the links are often far from obvious (Oreskes and Conway, 2010). As a result, research produced by interest groups is harder to discern from that done by objective scientists, and it may end up having more impact on the decision makers than it deserves\(^6\).

In many cases, the strategy of falsifying scientific evidence has resulted in a significant delay in (or even a lack of) regulatory action, which has benefited a specific interest group but has had a negative effect on the wider public. For example, by advancing claims that were contrary to mainstream scientific evidence, tobacco producers...
managed to persuade many people that there was reasonable doubt about the harmful effects of smoking, even though they had been aware of their existence for decades. It was only in 2009 that the U.S. Congress gave the authority to the Food and Drug Administration to regulate tobacco as an addictive drug\textsuperscript{9}. A report by the European Environment Agency (2013) provides a long list of cases where regulatory action was delayed despite sufficient scientific evidence having been provided early: lead in petrol, Beryllium exposure, Vinyl Chloride, the pesticide DBCP, Bisphenol A, and others\textsuperscript{10}.

In this paper, we develop a theoretical model to study falsification of scientific evidence by interest groups and to investigate possible measures which policy makers could take against it. There are two players, a sender and a decision maker, and the state of the world is binary, either high or low. The decision maker chooses between two actions—accepting or rejecting—and would like to accept when the state is high, and to reject when the state is low. The sender, on the other hand, would like the decision maker always to accept, regardless of the state of the world. This illustrates a setting in which, for example, a policy maker would like to adjust its policy towards cigarettes and other tobacco products to the true scientific evidence on whether or not smoking has adverse effects on human health. At the same time, an interest group such as a tobacco company would always like this policy to be lenient.

The sender first observes the state of the world, interpreted as the scientific evidence on an issue. He then decides whether to falsify evidence at a cost. If he does falsify, then—with some predetermined probability—the decision maker sees the falsified evidence rather than the true scientific evidence. This captures the idea that falsification is not always effective (e.g., the decision maker may detect it) and, if this happens, the decision maker learns the scientific evidence. Furthermore, the fact that the sender observes the true state of the world while the decision maker only sees a potentially corrupted message from the sender reflects the fact that an interest group may be better informed about the scientific evidence on the issue. Finally, falsification of evidence is assumed to be costly to the sender because—as the example from the tobacco industry shows—fabricating research that is false but hardly distinguishable from that done by objective scientists requires significant effort.

We investigate two possible measures the decision maker could take to improve her chances of taking the action that matches the state of the world. The first one is to

\textsuperscript{9}http://www.wsj.com/articles/SB124474789599707175
acquire additional independent information about the issue, for example, a policy maker could conduct an independent study on the impact of smoking on health. The second one is to control how much information to acquire from the sender, e.g., a policy maker could choose how much time to spend on meetings with lobbyists.

Acquiring additional independent information has two important implications. First, it naturally provides the decision maker with extra information about the state of the world. Second, more interestingly, it also affects the sender's incentives to falsify the scientific evidence. This is because, when deciding whether to falsify, the sender compares the cost of doing so with the expected benefit, which depends on how likely the falsified evidence is to change the decision maker's action. For example, it may be that the sender's information on its own cannot possibly persuade the decision maker to accept, but it can persuade her if the decision maker's independent information also turns out to be favourable to the sender. In that case, if the cost of falsifying is low and the sender expects the decision maker's independent information to be indeed favourable, then his incentives to falsify can be boosted by the presence of the independent evidence.

We demonstrate that the effect of acquiring independent evidence on the sender's incentives to falsify scientific evidence is ambiguous. The sender's incentives to falsify are strengthened when (i) falsification is cheap, (ii) detecting it is difficult, (iii) the quality of the decision maker's independent evidence is low, and (iv) the prior belief is in favour of the low state. These increased incentives to falsify diminish the value of the independent evidence, and—in the extreme—they may even fully offset the informational benefit it brings. On the other hand, the sender's incentives to falsify are dampened when (i) falsification is expensive, (ii) detecting it is easy, (iii) the quality of the decision maker's independent evidence is high, and (iv) the prior belief is not too much in favour of the high state. In this case, since there is a side effect of less falsification, the value of the independent evidence exceeds the mere informational benefit it brings.

Despite its ambiguous effect on falsification by the sender, acquiring independent evidence always makes the decision maker better off in our model. We also show that this result still holds when the sender does not observe the true state of the world, but only a noisy signal of it. However, when the sender observes only a noisy signal, it becomes possible that acquiring independent evidence by the decision maker completely deters falsification by the sender. Moreover, if the sender is imperfectly informed, we demonstrate that the decision maker's welfare may be non-monotonic in the quality of her independent evidence, and this non-monotonicity can only occur when the prior
belief is to a sufficient extent in favour of the high state.

The second measure which the decision maker could take is to control how much attention she devotes to the sender. By committing to ignore the sender with some positive probability, the decision maker can lower the chances that the sender’s message will be pivotal to her decision, and this may in turn dampen his incentives to falsify scientific evidence. At the same time, the negative consequence of ignoring the sender is naturally that she has no information other than the prior when making the decision. In our model, the decision maker’s optimal strategy turns out to be either to pay full attention to the sender or to commit to ignore him just often enough to completely discourage him from falsifying evidence.

We identify the circumstances under which it is optimal for the decision maker to commit to ignore the sender with a positive probability. The impact of detectability of falsification on the decision maker’s incentives to commit to ignore the sender is twofold. One channel is that, as falsification becomes more difficult to detect, the decision maker’s benefit from committing to ignore (and thus deterring falsification altogether) increases. On the other hand, this also drives down the sender’s relative cost of falsification (i.e. the cost of falsification relative to the probability of being undetected), and therefore the level of ignorance required to disincentivise falsification goes up and makes commitment to ignorance less attractive for the decision maker. We show that, as falsification becomes very difficult to detect, the former effect dominates and commitment to ignorance is optimal for the decision maker regardless of the cost of falsification.

Regarding the cost of falsification, commitment to ignorance is optimal when this cost takes moderate values. When it is very high, the sender has no incentive to falsify evidence and therefore there is no benefit to the decision maker from committing to ignore him. When the cost is very small, however, the decision maker needs to commit to ignore the sender with a very high probability if she wants to disincentivise falsification. This makes commitment to ignorance prohibitively costly to the decision maker.

Lastly, we discuss how the decision maker’s incentives to acquire an additional private signal interact with her incentives to commit to ignore the sender’s message.

The paper is organised as follows. The rest of this section discusses the related literature. Section 2.2 presents the main model and analyses its equilibrium. Section 2.3 investigates the decision maker’s incentives to acquire private independent evidence, while Section 2.4 studies her incentives to acquire information from the sender. Section 2.5 concludes.
Related Literature

Although the focus of our paper is on falsification of evidence in the context of interest groups and policy makers, the model can be reinterpreted more generally as one about lying, and hence it naturally belongs to the literature on strategic communication. In cheap talk models (Crawford and Sobel, 1982), the sender can lie arbitrarily without any direct costs and the receiver cannot detect lying. In verifiable disclosure games (Grossman, 1981; Milgrom, 1981), the sender can withhold information but cannot lie, which could be because lying is prohibitively costly for the sender and/or the receiver can always detect it. In our model, we could say that the sender decides whether to tell the truth, which is costless, or to lie, which is costly. The decision maker is able to detect lying with some predetermined probability, and if she does detect a lie, she immediately observes the true information of the sender.

There are a number of theoretical papers which study the impact of the cost and the detectability of lying on communication. Dziuda and Salas (2017) and Balbuzanov (2017) focus on settings where lying has no exogenous cost to the sender. In Dziuda and Salas (2017), if the receiver detects a lie, she only observes that the sender’s message is inconsistent. In other words, unlike in our model, the receiver detects only that the sender is lying but not the true information possessed by the sender. On the other hand, Balbuzanov (2017) assumes that if the decision maker detects a lie, she observes the message the sender sent but still does not observe his true information. Thus, the type of lies which our model analyses is different from those in Dziuda and Salas (2017) and Balbuzanov (2017). Our model is more suited to a particular type of lies which we could call “obvious lies”, i.e. those where the receiver can easily determine the truth whenever she detects a lie.\textsuperscript{11}

In Dziuda and Salas (2017), like in our model, the receiver wants to take an action that matches the state, whereas the sender wants the receiver to take the highest possible action. She shows that, in any informative equilibrium, the sender types can be divided into three distinct groups. The lowest types lie and pretend to be the highest

\textsuperscript{11}For example, Donald Trump claimed he had “the biggest electoral college win since Ronald Reagan”, while Sean Spicer (White House Press Secretary, January-July 2017) said about Donald Trump’s inauguration that “This was the largest audience ever to witness an inauguration, period. Both in person and around the globe.” Both statements could easily be checked to be false:
types. The intermediate and the highest types do not lie; however, the messages of the latter are discounted to a degree because of untruthful messages sent by the lowest types. Balbuzanov (2017) makes a different assumption about the preferences of the sender and the receiver in that they are partially aligned. He shows that, in his setting, fully revealing equilibria are possible, and that this remains true even for low levels of detectability of lies.

Kartik, Ottaviani and Squintani (2007) and Kartik (2009) assume that lying has an exogenous cost but the receiver is unable to detect lying. Kartik et al. (2007) analyse a model of strategic communication between an informed sender and one or more uninformed receivers, where the sender has an upward bias. The message sent by the sender directly affects his payoff, which can be interpreted as a consequence of lying being costly or the receiver possibly being naive. Kartik et al. (2007) show that, under broad conditions, there exist separating equilibria with inflated communication, provided that the state space is unbounded above. Kartik (2009) assumes instead that the sender types are bounded above (as well as below). For any given type, it is cheapest for the sender to tell the truth, and as the magnitude of the lie increases, so does the marginal cost of a bigger lie. Kartik (2009) shows that this setting leads to an inflated language and equilibria have the following structure: low types of the sender use inflated language but they still separate, i.e. reveal their types through their equilibrium message, while the high types segment into one or more pools.

Our paper is also related to the literature on manipulation of information by interest groups. Bramoullé and Orset (2017) model how interest groups “manufacture doubt” about scientific research. In their model, firms can, at a cost, fabricate evidence in their favour in order to influence the citizens’ beliefs about pollution caused by the firms. Additional evidence about this pollution is provided by scientists. The citizens’ beliefs are then taken into account by the government when deciding on the regulation of pollution. Bramoullé and Orset (2017) study the firms’ optimal amount of fabrication. The key difference is that they assume that the citizens naively treat the fabricated evidence as independent scientific evidence, while in our paper the receiver is more sophisticated: she cannot verify whether the sender’s message is falsified, but is aware of the sender’s incentives to do so.

Shapiro (2016) models how media communicate policy-relevant information to a voter when (i) interest groups can make claims of fact that appear credible to the voter, and (ii) journalists have reputational concerns for appearing neutral. He shows that these two frictions can interact to prevent the voter from learning useful facts.
Stone (2011) analyses a simple model of policy making in the presence of an interest group that chooses strategically whether to fund and lobby research. In his model, if research is not lobbied, it enters the public domain and can still be randomly observed by the policy maker. The main result is that, for a range of parameter values, the interest group sometimes funds but never lobbies research. This behaviour effectively “jams” the public signal of the policy maker, making the policy choice worse on average.

### 2.2 Baseline Model

There are two players: a sender and a decision maker (DM). The state of the world is binary, $\theta \in \{\theta_L, \theta_H\}$, and the prior belief that the state is $\theta = \theta_H$ is $p \in (0, 1)$. The DM chooses between two actions: she can either accept (denoted by $a = A$) or reject (denoted by $a = R$). She receives a payoff of $v(a, \theta) = 1$ if $\theta = \theta_H$ and $a = A$, and if $\theta = \theta_L$ and $a = R$; she receives $v(a, \theta) = 0$ otherwise. In other words, the DM prefers to accept when the state of the world is “high” and to reject when the state is “low". The sender receives a payoff of $u(a) = 1$ if $a = A$ and $u(a) = 0$ if $a = R$. Thus, he wants the DM to accept regardless of the state of the world. Both the DM and the sender are expected payoff maximisers.

The game proceeds as follows. First, the sender privately observes a perfectly informative signal of the state of the world: $s = i$ if $\theta = \theta_i$, for $i \in \{L, H\}$. He then sends a message $m \in \{L, H\}$ to the DM. We will say that the sender does not falsify evidence if $m = s$, and that he falsifies evidence if $m \neq s$. The message received by the DM is denoted by $\tilde{m}$. If the sender does not falsify, then the DM receives $\tilde{m} = m = s$. If the sender falsifies, then with probability $x \in (0, 1]$ the DM receives $\tilde{m} = m$ (i.e. falsification is successful), and with probability $1 - x$ the DM receives $\tilde{m} = s$ (i.e. falsification is unsuccessful). The parameter $x$ can be interpreted as a measure of undetectability of falsification, or a measure of the sender’s skill in falsification. If the sender falsifies, he pays an exogenous cost $c \in (0, x)$.\(^{13}\)

After receiving $\tilde{m}$, the DM forms a belief $\beta$ about the state of the world and chooses an action, $a \in \{A, R\}$. The DM’s posterior belief about the state of the world upon observing $\tilde{m}$ is denoted by $\beta_{\tilde{m}} = \Pr (\theta = \theta_H \mid \tilde{m})$. Given the payoff function of the DM,

\[^{12}\text{More generally, we could assume that the sender privately observes a realisation of a noisy signal of the state of the world, } s \in \{L, H\}. \text{ The conditional probability that the signal correctly reveals the true state is } \Pr (s = i \mid \theta = \theta_i) = q \text{ for } i \in \{L, H\} \text{ with } q \in \left(\frac{1}{2}, 1\right], \text{ where } q \text{ is the quality of the signal. Here, we assume } q = 1, \text{ but we relax this assumption in Section 2.3.5 to show additional results.} \]

\[^{13}\text{The sender would never have an incentive to falsify if it were that } c > x.\]
it is straightforward to note that (i) if $\beta_m > \frac{1}{2}$, the DM accepts, (ii) if $\beta_m < \frac{1}{2}$, the DM rejects, and (iii) if $\beta_m = \frac{1}{2}$, then the DM is indifferent between accepting and rejecting.

The sender’s strategy, $\sigma$, describes the probability with which he falsifies, with mixed strategies being possible. It is a function

$$\sigma : \{L, H\} \rightarrow \Delta \{0, 1\}.$$  (2.1)

We will denote by $\sigma_s$ the probability with which the sender falsifies when he observes a signal realisation $s$.

The DM’s strategy, $\delta$, describes the probability with which she accepts, with mixed strategies possible. It is a function

$$\delta : \{L, H\} \rightarrow \Delta \{A, R\}.$$  (2.2)

We will denote by $\delta_m$ the probability with which the DM accepts when she observes $\bar{m}$.

### 2.2.1 Equilibrium

The solution concept is the perfect Bayesian equilibrium, which is defined in the usual way: a triple $(\sigma^*, \delta^*, \beta^*)$ is a perfect Bayesian equilibrium if (i) $\sigma^*_s$ maximises the expected value of $u(a)$ given the DM’s strategy $\delta^*$ and the belief $\beta^*$, (ii) $\delta^*_m$ maximises the expected value of $v(a, \theta)$ given the sender’s strategy $\sigma^*$ and the belief $\beta^*$, and (iii) $\beta^*$ is derived using $\sigma^*$ by Bayes’ rule from the strategy of the sender and the prior distribution over $\{\theta_H, \theta_L\}$, whenever it is possible. Henceforth, we refer to the perfect Bayesian equilibrium simply as an equilibrium. With the exception of knife-edge cases, the game has a unique equilibrium, which we state formally in the following proposition.

**Proposition 2.1.** The game has a unique equilibrium, except for knife-edge cases. The sender’s and the DM’s equilibrium strategies are:

(i) if $p \in \left(0, \frac{x}{1+x}\right)$, then $(\sigma^*_L, \sigma^*_H) = \left(\frac{p}{(1-p)x}, 0\right)$ and $(\delta^*_L, \delta^*_H) = \left(0, \frac{p}{x}\right)$;

(ii) if $p \in \left[\frac{x}{1+x}, 1\right)$, then $(\sigma^*_L, \sigma^*_H) = (1, 0)$ and $(\delta^*_L, \delta^*_H) = (0, 1)$.

The sender never has an incentive to falsify evidence upon observing $s = H$, and hence in equilibrium it must always be that $\sigma^*_H = 0$. For brevity, whenever we say that the sender’s equilibrium strategy is to falsify with some probability, we mean that he

---

14In knife-edge cases, we select the sender-preferred strategy profile whenever there are multiple equilibria.
falsifies with this probability upon observing \( s = L \) but does not falsify upon observing \( s = H \).

When prior belief \( p \) is sufficiently low, \( p < \frac{x}{1 + x} \), there is a mixed-strategy equilibrium in which the sender falsifies with a positive probability, whereas the DM accepts with a positive probability upon observing \( \tilde{m} = H \) and rejects otherwise. Let \( \beta_{\tilde{m}}(\sigma_L, \sigma_H) \) denote the DM’s belief upon observing \( \tilde{m} \) when the sender’s strategy is \( (\sigma_L, \sigma_H) \). The condition \( p < \frac{x}{1 + x} \) can be rewritten as \( \beta_H(\sigma_L = 1, \sigma_H = 0) < \frac{1}{2} \). This condition and the fact that \( \beta_H(\sigma_L = 0, \sigma_H = 0) > \frac{1}{2} \) together imply that there is no equilibrium in pure strategies. If the sender’s strategy were to falsify with probability zero, the DM’s optimal response would then be to accept with probability 1 upon receiving a message \( \tilde{m} = H \), which would create incentives for the sender to falsify. But if the sender’s strategy were to falsify with probability 1, then the DM would not be willing to accept upon observing \( \tilde{m} = H \), which would erase the sender’s incentives to falsify. Thus, there is instead a mixed-strategy equilibrium in which the sender is indifferent between falsifying and not falsifying when he sees \( s = L \), while the DM is indifferent between accepting and rejecting when she receives a message \( \tilde{m} = H \). The sender’s equilibrium strategy \( (\sigma^*_L, \sigma^*_H) \) satisfies the condition for the DM’s indifference:

\[
\beta_H(\sigma_L = \sigma^*_L, \sigma_H = 0) = \frac{p}{p + (1 - p)\sigma^*_L x} = \frac{1}{2},
\]

which implies that \( \sigma^*_L = \frac{p}{(1 - p)x} \). The DM’s equilibrium strategy \( (\delta^*_L, \delta^*_H) \) satisfies the condition for the sender’s indifference: \( \delta^*_H x = c \).

When prior belief \( p \) is sufficiently high, \( p \geq \frac{x}{1 + x} \), the sender falsifies with probability 1, while the DM accepts upon observing \( \tilde{m} = H \) and rejects upon observing \( \tilde{m} = L \). The condition that \( p > \frac{x}{1 + x} \) is equivalent to \( \beta_H(\sigma_L = 1, \sigma_H = 0) > \frac{1}{2} \), which implies that the DM’s optimal strategy is to accept upon observing \( \tilde{m} = H \) even when the sender falsifies with probability 1.

It follows from Proposition 1 that the probability that the sender falsifies evidence upon observing \( s = L \) is strictly increasing in prior belief \( p \) whenever \( p \in (0, \frac{x}{1 + x}) \), and is equal to 1 for \( p \in \left[ \frac{x}{1 + x}, 1 \right) \). The probability that the DM accepts upon observing \( \tilde{m} = H \) is weakly increasing in \( p \): it is equal to \( \frac{c}{x} \) when \( p \in (0, \frac{x}{1 + x}) \), and is equal to 1 for \( p \in \left[ \frac{x}{1 + x}, 1 \right) \).
2.2.2 Decision Maker’s Welfare

The DM’s ex-ante expected payoff in equilibrium is

\[ E(v(\sigma^*, \delta^*)) = \begin{cases} 1 - p & \text{for } p \in (0, \frac{x}{1+x}) \\ 1 - (1 - p)x & \text{for } p \in \left[\frac{x}{1+x}, 1\right] \end{cases} \]  

(2.4)

When \( p < \frac{x}{1+x} \), the DM is indifferent between the two actions when she observes a message \( \tilde{m} = H \), and she rejects otherwise. Thus, her expected payoff would remain the same if she rejected regardless of the message observed. Hence, her expected payoff is \( 1 - p \), which is as if she received a completely uninformative message.

On the other hand, when \( p \geq \frac{x}{1+x} \), the DM’s equilibrium strategy is such that a message \( \tilde{m} = H \) prompts her to accept and a message \( \tilde{m} = L \) prompts her to reject. Since the sender never falsifies when he observes \( \theta = \theta_H \), the DM takes an incorrect action only when the state is \( \theta = \theta_L \) and the sender successfully falsifies. The DM’s expected payoff is then \( E(v(\sigma^*, \delta^*)) = 1 - (1 - p)x \).

Thus, the DM’s expected payoff is non-monotonic in prior belief \( p \): it is strictly decreasing in \( p \) for \( p < \frac{x}{1+x} \) and is strictly increasing in \( p \) for \( p \geq \frac{x}{1+x} \). There are two effects here which lead to this U-shaped relationship. First, when \( p < \frac{x}{1+x} \), i.e. when the DM’s welfare is the same as if she relied only on her prior, a value of \( p \) that is further away from 0 or 1 means that there is more uncertainty about the state of the world and hence the action that the DM takes based on the prior is less likely to match the true state. Second, when \( p \geq \frac{x}{1+x} \), i.e. when the DM follows the message received from the sender, a higher value of \( p \) means that the sender is ex ante less likely to observe \( s = L \), and hence ex ante less likely to falsify.

Furthermore, the DM’s expected payoff is decreasing in the undetectability of falsification, \( x \). Intuitively, as the sender becomes more capable at corrupting the signal about the state of the world, the DM becomes less likely to take an action that matches the true state.

2.2.3 Other Models of Strategic Communication

Although the structure of the model is simple, it is worth noting that the limiting cases with respect to \( c \) and \( x \) capture the equilibrium outcomes (in particular, the DM’s expected payoff and the probability of her accepting conditional on the state of the world) of three canonical models of strategic communication: verifiable disclosure,
cheap talk, and Bayesian persuasion.

When \( x \to 0 \), the sender’s falsification is detectable with a probability approaching 1, and hence the sender is unable to corrupt the observed signal when sending a message to the DM. As a result, the outcome of the model is the same as in models of verifiable disclosure (Milgrom, 1981; Grossman, 1981). In these models, the sender cannot report false information to the DM but can suppress information and—by the unravelling result—the DM learns the state of the world in equilibrium.

When \( x \to 1 \) and \( c \to 0 \), falsification is virtually costless to the sender and undetectable by the DM, which makes our model related to the model of cheap talk (Crawford and Sobel, 1982), in which the sender sends a costless but unverifiable report to the DM about the state of the world. The outcome predicted by our model is then the same as in a babbling equilibrium, i.e. an equilibrium in which the sender sends an uninformative message while the DM ignores the sender’s message and makes a decision based on her prior belief. In our model, when \( x \to 1 \) and \( c \to 0 \), we have a mixed-strategy equilibrium for \( p < \frac{1}{2} \), and—as we have argued above—the equilibrium strategies are then such that the DM’s expected payoff is the same as if she relied only on her prior when making the decision. On the other hand, when \( p \geq \frac{1}{2} \), the DM’s equilibrium strategy is to follow the message \( \tilde{m} \) received from the sender, but since \( x \to 1 \), this message is essentially always \( \tilde{m} = H \) and hence completely uninformative about the state of the world, and thus again the DM’s expected payoff is the same as if she relied only on her prior.

The fact that, in equilibrium, the DM is indifferent between accepting and rejecting when \( p < \frac{1}{1+x} \) suggests that the model can also be related to Bayesian persuasion (Kamenica and Gentzkow, 2011). In models of Bayesian persuasion, the sender—before privately observing the state of the world—designs an experiment, i.e. a mapping from each possible state of the world to a distribution of signal realisations, and publicly commits to it. In our model, when \( x \to 1 \), the sender’s equilibrium strategy is identical to his optimal commitment strategy in a game of Bayesian persuasion. If the cost of falsification is very large, \( c \to 1 \), then we have \( \delta^*_H = \frac{c}{x} \to 1 \), which means that the DM’s equilibrium strategy is to always accept when she observes \( \tilde{m} = H \). This strategy is the same as the one she would have in a game of Bayesian persuasion, and—as a result—so is the probability of her accepting conditional on the state of the world as well as her expected payoff.
Figure 2.1: The relationship between the outcomes of the model in Section 2.2 (in particular, the DM’s expected payoff and the probability of her accepting conditional on the state of the world) and their equivalents in models of verifiable disclosure (VD), cheap talk (CT), and Bayesian persuasion (BP). The shaded area illustrates the condition $c < x$.

We summarise the above results in the following proposition, and illustrate them in Figure 2.1.

**Proposition 2.2.** The DM’s expected payoff, as well as the probability of her accepting conditional on the state of the world, approach:

(i) their equivalents in a model of verifiable disclosure when $x \to 0$;
(ii) their equivalents in a model of cheap talk when $x \to 1$ and $c \to 0$;
(iii) their equivalents in a model of Bayesian persuasion when $x \to 1$ and $c \to 1$.

### 2.3 Acquisition of Private Independent Evidence

In this section, we analyse a scenario in which the DM receives—in addition to the message from the sender—a private signal about the state of the world. The aim of the analysis is to provide insight into policy makers’ incentives to conduct private independent research, e.g., by consulting independent experts or hiring independent advisers.
The game now proceeds as follows. As before, the sender privately observes a perfectly informative signal of the state of the world: \( s = i \) if \( \theta = \theta_i \), for \( i \in \{L, H\} \). However, unlike in the baseline model, the DM now observes a realisation of a private signal of the state of the world. The fact that the DM receives the private signal is public knowledge. We denote the DM’s signal realisation by \( s \). The DM’s signal is assumed to be of quality \( q_{DM} \in \left( \frac{1}{2}, 1 \right) \), where the quality describes the conditional probability \( \Pr(s_{DM} = i \mid \theta = \theta_i) = q_{DM} \).

After observing \( s \), the sender sends a message \( m \in \{H, L\} \) to the DM. If the sender does not falsify, i.e. sends \( m = s \), then the DM receives \( \tilde{m} = m = s \). If the sender falsifies, i.e. sends \( m \neq s \), then with probability \( x \in [0, 1] \) the DM receives \( \tilde{m} = m \), and with probability \( 1 - x \) the DM receives \( \tilde{m} = s \). Thus, overall, the information observed by the DM is now a pair \((s_{DM}, \tilde{m})\) \( \in \{(L, L), (L, H), (H, L), (H, H)\} \). After observing \((s_{DM}, \tilde{m})\), the DM forms a belief \( \beta \) about the state of the world and chooses an action, \( a \in \{A, R\} \).

The strategy of the sender is still denoted by \((\sigma_L, \sigma_H)\). For the DM’s strategy, we will now use \( \delta_{s_{DM}\tilde{m}} \) to denote the probability with which the DM accepts when she observes \((s_{DM}, \tilde{m})\). The DM’s strategy is thus given by \((\delta_{LL}, \delta_{LH}, \delta_{HL}, \delta_{HH})\).

### 2.3.1 Impact of Private Independent Evidence on Equilibrium

The following proposition characterises the equilibrium in the setup in which the DM obtains a private signal.\(^{10}\)

**Proposition 2.3.** Suppose that, in addition to receiving a message from the sender, the DM observes a noisy private signal of quality \( q_{DM} \in \left( \frac{1}{2}, 1 \right) \). Except for knife-edge cases, there is a unique equilibrium. The equilibrium strategies of the sender and the DM satisfy \( \sigma^*_L = 0 \) and \( \delta^*_L = \delta^*_H = 0 \), and furthermore:

(i) if \( p^- \geq \frac{x}{1+x} \), then \( \sigma^*_L = 1 \) and \( (\delta^*_{LH}, \delta^*_H) = (1, 1) \);

(ii) if \( p^- < \frac{x}{1+x} \leq p^+ \) and \( q_{DM} < 1 - \frac{e}{x} \), then \( \sigma^*_L = 1 \) and \( (\delta^*_{LH}, \delta^*_H) = (0, 1) \);

(iii) if \( p^+ < \frac{x}{1+x} \) and \( q_{DM} < 1 - \frac{e}{x} \), then \( \sigma^*_L = \frac{p^+}{(1-p^+)^x} \) and \( (\delta^*_{LH}, \delta^*_H) = \left( \frac{0}{(1-q_{DM})^x}, 1 \right) \);

(iv) if \( p^- < \frac{x}{1+x} \) and \( q_{DM} \geq 1 - \frac{e}{x} \), then \( \sigma^*_L = \frac{p^-}{(1-p^-)^x} \) and \( (\delta^*_{LH}, \delta^*_H) = \left( \frac{e-(1-q_{DM})x}{q_{DM}^x}, 1 \right) \),

where \( p^- = \beta(\theta = \theta_H \mid s_{DM} = L) \) and \( p^+ = \beta(\theta = \theta_H \mid s_{DM} = H) \).

---

\(^{15}\)Equivalently, we could think of the baseline model as of one in which the DM’s private signal is uninformative, i.e. \( q_{DM} = \frac{1}{2} \).

\(^{16}\)As before, in knife-edge cases, we select the sender-preferred strategy profile whenever there are multiple equilibria.
The following table restates the results of Proposition 2.3 and compares them with those in the baseline model:

<table>
<thead>
<tr>
<th>Region</th>
<th>$p^+ &lt; \frac{x}{1+x}$</th>
<th>$p^+ \geq \frac{x}{1+x} &gt; p$</th>
<th>$p \geq \frac{x}{1+x} &gt; p^-$</th>
<th>$p^- \geq \frac{x}{1+x}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline Model</td>
<td>$\sigma_L^* = \frac{p}{1-p}$ and $(\delta_L^<em>, \delta_H^</em>) = (0, \frac{c}{2})$</td>
<td>$\sigma_L^* = 1$ and $(\delta_L^<em>, \delta_H^</em>) = (0, 1)$</td>
<td>$\sigma_L^* = 1$</td>
<td>$(\delta_L^<em>, \delta_H^</em>) = (1, 1)$</td>
</tr>
<tr>
<td>Private Signal</td>
<td>$\sigma_L^* = \frac{p^+}{1-p^+}$</td>
<td>$\sigma_L^* = \frac{p^-}{1-p^-}$</td>
<td>$\sigma_L^* = 1$</td>
<td>$\sigma_L^* = 1$</td>
</tr>
<tr>
<td>$q_{DM} &lt; 1 - \frac{c}{2}$</td>
<td>$(\delta_{LH}^<em>, \delta_{HH}^</em>) = (0, \frac{c}{1-q_{DM}})$</td>
<td>$(\delta_{LH}^<em>, \delta_{HH}^</em>) = (0, 1)$</td>
<td></td>
<td>$\sigma_L^* = 1$</td>
</tr>
<tr>
<td>Private Signal</td>
<td>$\sigma_L^* = \frac{p^+}{1-p^+}$</td>
<td>$\sigma_L^* = \frac{p^-}{1-p^-}$</td>
<td>$\sigma_L^* = 1$</td>
<td>$\sigma_L^* = 1$</td>
</tr>
<tr>
<td>$q_{DM} \geq 1 - \frac{c}{2}$</td>
<td>$(\delta_{LH}^<em>, \delta_{HH}^</em>) = \left(\frac{c-(1-q_{DM})}{q_{DM}}, 1\right)$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.1: A comparison of equilibrium strategy profiles specified in Proposition 2.3 with those in the baseline model (see Proposition 2.1).

Proposition 2.3 tells us that the equilibrium is described by four disjoint and collectively exhaustive regions in the $(p, q_{DM}, x, c)$ parameter space. We refer to the regions as:

1. $PF_{LH}$ (pure-strategy falsification with the DM accepting with probability 1 when $(s_{DM}, \bar{m}) \in \{(L, H), (H, H)\}$);
2. $PF_{HH}$ (pure-strategy falsification with the DM accepting with probability 1 when $(s_{DM}, \bar{m}) = (H, H)$);
3. $MF_{HH}$ (mixed-strategy falsification with the DM accepting with a positive probability when $(s_{DM}, \bar{m}) = (H, H)$);
4. $MF_{LH}$ (mixed-strategy falsification with the DM accepting with a positive probability when $(s_{DM}, \bar{m}) = (L, H)$ and probability 1 when $(s_{DM}, \bar{m}) = (H, H)$).

The four regions correspond to cases (i)-(iv) in Proposition 2.3. Figure 2.2 illustrates them in the $(p, q_{DM})$ parameter space while keeping $x$ and $c$ fixed. The baseline model is captured by $q_{DM} = \frac{1}{2}$.

The key difference between this setup and the baseline environment is that, here, the sender needs to make an inference about the DM’s belief: the sender knows the DM’s prior belief but he does not know her belief after she has observed her private signal. If the DM observes $s_{DM} = H$, then her belief becomes $\beta(\theta = \theta_H \mid s_{DM} = H) = \frac{pq_{DM}}{pq_{DM} + (1-p)(1-q_{DM})}$, which we denote by $p^+$. If she observes $s_{DM} = L$, then her belief
becomes $\beta(\theta = \theta_H \mid s = L) = \frac{p(1-q_{DM})}{p(1-q_{DM})+(1-p)q_{DM}}$, which we denote by $p^-$. We will say that a DM who has observed $s_{DM} = H$ ($s_{DM} = L$) is “optimistic” (“pessimistic”).

When $(p, q_{DM}) \in PF_{LH}$, the prior belief of the DM is so high that both an optimistic and a pessimistic DM’s belief is swayed towards acceptance when the sender falsifies with probability 1. Thus, given that $c < x$, the sender has an incentive to falsify with probability 1. The quality of the DM’s own private signal is low enough that, in equilibrium, her decision depends solely on the sender’s message, and not on the realisation of the private signal.

When $(p, q_{DM}) \in MF_{HH}$, the prior belief is at a moderate level, which means that only an optimistic DM’s belief is swayed towards acceptance when the sender falsifies with probability 1. The quality of the private signal is relatively low, which means that the sender realises upon observing $s = L$ that it is quite likely that the DM has observed $s_{DM} = H$, i.e. that he may well be facing an optimistic DM. In other words, when $q_{DM}$ is low, then the probability $\Pr(s_{DM} = H \mid s = L) = 1 - q_{DM}$ is high. In equilibrium, the DM accepts (and does so with probability 1) only if $s_{DM} = H$ and $\tilde{m} = H$, and the sender has an incentive to falsify with probability 1 as the expected benefit of falsification, $(1 - q_{DM})x$, exceeds the cost, $c$.

When $(p, q_{DM}) \in MF_{HH}$, then the prior belief is so low that even an optimistic DM’s belief is not swayed towards acceptance when the sender falsifies with probability 1—so
the sender needs to falsify with probability less than 1 in order to make an optimistic DM accept. Furthermore, the quality of the DM’s signal is low, so the sender realises upon observing \( s = L \) that the probability of facing an optimistic DM remains quite high. In equilibrium, the sender falsifies with a probability that makes an optimistic DM indifferent between accepting and rejecting upon observing \( \tilde{m} = H \), whereas a pessimistic DM rejects upon observing \( \tilde{m} = H \).

Finally, when \((p, q_{DM}) \in MF_{LH}\), then the quality of the private signal is high, so the sender realises upon observing \( s = L \) that he likely faces a pessimistic DM. In equilibrium, the sender falsifies with a probability that makes a pessimistic DM indifferent between accepting and rejecting upon observing \( \tilde{m} = H \), while an optimistic DM accepts upon observing \( \tilde{m} = H \).

### 2.3.2 How Is Falsification Affected?

Having described the equilibrium in the setup where the DM acquires a private signal, we can now analyse how the sender’s falsification changes compared to the setting in which she does not observe a private signal. It turns out that the effect on the sender’s falsification is ambiguous: it may increase, decrease, or remain unchanged.

The following proposition describes how this effect depends on the parameter values of the model:

**Proposition 2.4.** The impact of the DM’s acquisition of a private signal of quality \( q_{DM} \) on the sender’s falsification in equilibrium, \( \sigma^*_L \), is as follows:

(i) if \( p^- \geq \frac{x}{1+x} \) and if \( p^- < \frac{x}{1+x} \leq p \) and \( q_{DM} < 1 - \frac{c}{x} \), the sender’s falsification remains unchanged;

(ii) if \( p < \frac{x}{1+x} \) and \( q_{DM} < 1 - \frac{c}{x} \), the sender’s falsification increases;

(iii) if \( p^- < \frac{x}{1+x} \) and \( q_{DM} \geq 1 - \frac{c}{x} \), the sender’s falsification decreases.

Figure 2.3 illustrates the effect of the DM’s observation of a private signal on the sender’s falsification strategy, \( \sigma^*_L \), in the \((p, q_{DM})\) parameter space for fixed values of \( c \) and \( x \) (which are set to be the same as in Figure 2.2).

The fact that the DM observes a private signal has no effect on the sender’s falsification when the prior belief is high enough (region labelled as “no effect”). Intuitively, when the prior is high enough, then—regardless of whether the DM observes a private signal or not—she will likely accept even if the sender falsifies with probability 1. Thus, the sender has an incentive to falsify with the same probability (equal to 1) in both cases.
When the quality of the DM’s signal is low and the prior belief is low (region labelled as “more falsification”), the DM’s acquisition of a private signal results in more falsification by the sender. Since the DM’s signal is of low quality, the sender believes upon observing $s = L$ that he quite likely faces an optimistic DM, which makes him falsify with a higher probability than if he faced a neutral DM (i.e. a DM who has no private information and whose belief about the state of the world is $p$). More precisely, in equilibrium, the sender’s intensity of falsification is either 1 or such that an optimistic DM is indifferent between accepting and rejecting when she observes $\tilde{m} = H$. In either case, the intensity is higher than what we observe when the DM does not have private information—where the sender’s intensity of falsification is such that it makes a neutral DM indifferent when she observes $\tilde{m} = H$.

Lastly, when the quality of the DM’s signal is high and the prior belief is low (region labelled as “less falsification”), the DM’s observation of a private signal results in less falsification by the sender. Here, the sender realises upon observing $s = L$ that he likely faces a pessimistic DM, which makes him falsify with a smaller probability than if he faced a neutral DM. More precisely, in equilibrium, the sender’s intensity of falsification makes a pessimistic DM indifferent between accepting and rejecting when she sees $\tilde{m} = H$. This intensity is less than what we observe when the DM does not
have private information—where the sender’s intensity of falsification is equal to or higher than the one that makes a neutral DM indifferent when she observes $\tilde{m} = H$.

### 2.3.3 How Is the Decision Maker’s Welfare Affected?

In this subsection, we analyse how obtaining a private signal affects the DM’s welfare, i.e. her ex-ante expected payoff in equilibrium.

The private signal affects the DM’s welfare in two ways. First, observing a private signal obviously means that the DM has additional information, which boosts her expected payoff. Second, it has an impact on the falsification by the sender and thus on how likely the information received from the sender is manipulated. In the previous section, we have shown that this second effect can work in either direction: the sender’s incentives to falsify can be weakened or strengthened. If they are weakened, then a private signal brings an additional benefit—beyond the informational content of the signal itself—in that the sender’s message becomes less manipulated. If they are strengthened, the DM’s benefit from the informational content of the private signal is at least partly offset by the loss associated with more falsification by the sender.

We denote the value of the private signal by

$$V = E\left[ v\left(\sigma^*\left(q_{DM}\right), \delta^*\left(q_{DM}\right)\right) \right] - E\left[ v\left(\sigma^*\left(\frac{1}{2}\right), \delta^*\left(\frac{1}{2}\right)\right) \right], \quad (2.5)$$

where $\sigma^*\left(q_{DM}\right)$ and $\delta^*\left(q_{DM}\right)$ are the equilibrium strategies under a given value of $q_{DM}$.

**Proposition 2.5.** The DM’s acquisition of a private signal of quality $q_{DM}$ strictly increases her welfare if (i) $q_{DM} < 1 - \frac{c}{x}$ and $p^- < \frac{x}{1+x} \leq p^+$, or (ii) $q_{DM} \geq 1 - \frac{c}{x}$ and $p^- < \frac{x}{1+x}$. Otherwise, her welfare remains unchanged relative to the baseline model.

The results of Proposition 2.4 and Proposition 2.5 produce five mutually exclusive regions in the $(p, q)$ parameter space, which we illustrate in Figure 2.4. We describe these five regions below.

**Region $V = 0$ and more falsification.** This region is identical to region $M F^*_{HH}$ in Figure 2.2 and is defined by $q_{DM} < 1 - \frac{c}{x}$ and $p^+ < \frac{x}{1+x}$. Here, the private signal does not change the DM’s expected payoff in equilibrium. She is indifferent between accepting and rejecting when $s_{DM} = H$ and $\tilde{m} = H$, and rejects otherwise. Hence, her equilibrium strategy is such that she would have the same expected payoff if she always
Figure 2.4: An illustration of how the DM’s acquisition of a private signal affects the DM’s welfare (compared to the baseline model with no DM’s private signal). The simulation uses $c = 0.2$ and $x = 0.75$; “m.f.” and “n.f.” are abbreviations for, respectively, “more falsification” and “no effect on falsification”.

rejected regardless of $s_{DM}$ and $\tilde{m}$. In the baseline model, her equilibrium strategy is also such that she would have the same expected payoff if she always rejected. Therefore, in both cases, her expected payoff is the same as if she relied on her prior only. The increased falsification by the sender (which we have established in Proposition 2.4) thus fully offsets any benefits brought by the additional information in the private signal.

Region $V > 0$ and more falsification. This region is a part of region $PF_{HH}^*$ in Figure 2.2 and is defined by $q_{DM} < 1 - \frac{c}{2}$ and $p < \frac{p}{1 + p} \leq p^*$. The private signal strictly increases the DM’s expected payoff in equilibrium. The DM accepts only when $s_{DM} = H$ and $\tilde{m} = H$, and rejects otherwise. In other words, the realisation of the private signal is essential to determining her optimal action in equilibrium, and she would not be able to achieve the same expected payoff if she relied on the same decision rule as in the baseline model (which was only a function of $\tilde{m}$ and was thus coarser than here). As we have seen in Proposition 2.4 there is more falsification than in the baseline model; however, unlike in the previous region, the increased falsification does not fully offset the benefits brought by the information contained in the private signal.
Region $V > 0$ and no effect on falsification. This region is a part of region $PF_{HH}^*$ in Figure 2.2 and is defined by $q_{DM} < 1 - \frac{c}{x}$ and $p^- < \frac{x}{1+x} \leq p$. The private signal strictly increases the DM’s expected payoff in equilibrium. The DM’s equilibrium strategy is the same as in the previous region: she accepts only when $s_{DM} = H$ and $\tilde{m} = H$, and rejects otherwise. This means that her decision-making relies on realisations of both $s_{DM}$ and $\tilde{m}$, and thus is finer than in the baseline model, where the DM accepted when $\tilde{m} = H$ and rejected otherwise. At the same time, the sender’s falsification strategy remains the same as in the baseline model, which means that—unlike in the previous two regions—the informational benefit from the private signal is not dampened here by increased falsification.

Region $V = 0$ and no effect on falsification. This region is identical to region $PF_{LH}^*$ in Figure 2.2 and is defined by $p^- \geq \frac{x}{1+x}$. The private signal does not change the DM’s expected payoff in equilibrium. This is because the sender’s equilibrium strategy does not change relative to the baseline model and neither does the DM’s decision rule: her equilibrium strategy in the model with a private signal is such that she accepts if and only if $\tilde{m} = H$. Thus, despite bringing some additional information about the state of the world, the DM’s expected payoff remains the same as in the baseline model.

Region $V > 0$ and less falsification. This region is identical to region $MF_{LH}^*$ in Figure 2.2 and is defined by $q_{DM} \geq 1 - \frac{c}{x}$ and $p^- < \frac{x}{1+x}$. The private signal strictly increases the DM’s expected payoff in equilibrium. The DM’s equilibrium strategy specifies that she is indifferent between accepting and rejecting when $s_{DM} = L$ and $\tilde{m} = H$, accepts when $s_{DM} = H$ and $\tilde{m} = H$, and rejects otherwise. Hence, the probability of her accepting upon observing $\tilde{m} = H$ differs depending on whether the realisation of the private signal is $s_{DM} = L$ or $s_{DM} = H$. The benefit brought by the additional information in the private signal is further boosted by the reduced falsification by the sender (which we established in Proposition 2.4).

Despite the potentially increased falsification by the sender, the DM is always weakly better off with a private signal in this setting: the fact that the DM obtains a private signal never boosts the sender’s incentives to falsify by so much as to make the DM worse off with a signal of quality $q_{DM}$ than without it. In fact, a stronger statement follows from Propositions 2.3 and 2.4.
Corollary 2.1. The welfare of the decision maker is increasing in the quality of her private signal, $q_{DM}$.

As we will show later, the result that $V \geq 0$ still holds when the quality of the sender’s private signal is $q < 1$; however, the statement in Corollary 2.1 may then not be satisfied.

2.3.4 Private vs. Public Independent Evidence

In this section, we briefly consider how the DM’s welfare is affected when the additional signal obtained by the DM is publicly observed by both players, as opposed to being privately observed only by the DM. The aim here is to shed light on policy makers’ benefits from conducting public rather than private independent research.

We will use $s_{pub}$ to denote the realisation of the public signal. Since the sender decides whether to falsify evidence after the public signal is realised, the analysis here amounts to repeating the steps in the baseline model for two cases: (i) the public signal is $s_{pub} = H$ and therefore the new prior is $p^+ = \Pr(\theta = \theta_H | s_{pub} = H)$, and (ii) the public signal is $s_{pub} = L$ and therefore the new prior is $p^- = \Pr(\theta = \theta_H | s_{pub} = L)$.

We will say that the DM is “optimistic” (“pessimistic”) if $s_{pub} = H$ ($s_{pub} = L$); however, since the signal is public, the sender now knows whether the DM is optimistic or pessimistic.

The following table shows the DM’s and the sender’s equilibrium strategies when the DM obtains a public signal of quality $q_{DM}$:

<table>
<thead>
<tr>
<th>$p^+ &lt; \frac{x}{1+x}$</th>
<th>$p^+ \geq \frac{x}{1+x} &gt; p^-$</th>
<th>$p^- \geq \frac{x}{1+x}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S: (\sigma_{LL}^<em>, \sigma_{HL}^</em>)$</td>
<td>$(\frac{p^-}{1-p^-}, \frac{p^+}{1-p^+})$, $(1, 1)$</td>
<td>$(0, 0, \frac{c}{x}, \frac{c}{x})$, $(0, 0, 1, 1)$</td>
</tr>
<tr>
<td>$\text{DM: } (\delta_{LL}^<em>, \delta_{HL}^</em>, \delta_{LH}^<em>, \delta_{HH}^</em>)$</td>
<td>$(0, 0, \frac{c}{x}, \frac{c}{x})$, $(0, 0, \frac{c}{x}, 1)$</td>
<td>$(0, 0, 1, 1)$</td>
</tr>
</tbody>
</table>

Table 2.2: A summary of the equilibrium strategy profiles when the DM obtains a public signal of quality $q_{DM}$.

We thus obtain the following proposition:

Proposition 2.6. When $q_{DM} \geq 1 - \frac{c}{x}$ and $p^- < \frac{x}{1+x}$, the DM’s welfare with a public signal of quality $q_{DM}$ is strictly lower than that with a private signal of quality $q_{DM}$. Otherwise, the DM’s welfare is the same regardless of whether the additional signal is public or private.
The value of a public signal is constructed analogously to (2.5) in the analysis of the case with a private signal. Proposition 2.6 implies that, by obtaining a public signal, the DM can increase her welfare but the increase is never higher than it would have been with a private signal. Moreover, for a sufficiently high quality of the signal and a sufficiently low prior, the benefit from obtaining a signal is strictly higher if it is private rather than public.

It is worth noting, however, that when the signal is public, there is weakly less falsification of evidence than when it is private and \( q_{DM} < 1 - \frac{c}{x} \). More precisely, whenever \( p^- < \frac{x}{1+x} \), there is strictly less falsification in the former case than in the latter, and whenever \( p^- > \frac{x}{1+x} \), the amount of falsification is the same whether the signal is public or private.

As we have seen in Table 2.1, when \( p^+ < \frac{x}{1+x} \), \( q_{DM} < 1 - \frac{c}{x} \) and the signal is private, the sender falsifies with a probability that makes an optimistic DM indifferent between accepting and rejecting upon observing \( \tilde{m} = H \). On the other hand, when the signal is public, then we can see from Table 2.2 that the probability of falsification depends on the realisation of the public signal: if \( s_{pub} = L \), the sender falsifies with a probability that makes a pessimistic DM indifferent (which is lower than the probability needed to make an optimistic DM indifferent), while if \( s_{pub} = H \), he falsifies with a probability that makes an optimistic DM indifferent. For \( p^+ \geq \frac{x}{1+x} > p^- \) and \( q_{DM} < 1 - \frac{c}{x} \), the sender falsifies with probability 1 when the DM’s signal is private; however, when it is public, then again if \( s_{pub} = L \), he falsifies with a probability that makes a pessimistic DM indifferent (which is lower than 1), and if \( s_{pub} = H \), he falsifies with probability 1.

Despite there being weakly less falsification when the signal is public than when it is private and \( q_{DM} < 1 - \frac{c}{x} \), the DM’s welfare is the same in both cases. The reason for this is that the reduced falsification is actually on a margin that is irrelevant to the DM’s ex ante expected payoff.

To see why, consider for example a prior belief such that \( p^+ < \frac{x}{1+x} \). When the signal is private and \( q_{DM} < 1 - \frac{c}{x} \), the DM’s equilibrium strategy is such that a pessimistic DM always rejects and an optimistic DM is indifferent between accepting and rejecting upon observing \( \tilde{m} = H \) and rejects otherwise. When the signal is public, then regardless of whether \( s_{pub} = L \) or \( s_{pub} = H \), the DM’s equilibrium strategy is such that she is indifferent upon receiving \( \tilde{m} = H \) and rejects otherwise. Therefore, when \( q_{DM} < 1 - \frac{c}{x} \), no matter whether the signal is public or private, the DM’s ex ante expected payoff is the same as if she rejected regardless of the observed information, i.e. as if she effectively relied on her prior only. On the other hand, when \( q_{DM} \geq 1 - \frac{c}{x} \), her ex ante expected
payoff is strictly higher with a private signal. The DM’s equilibrium strategy is then to accept when \((s_{DM}, \tilde{m}) = (H, H)\), and she is indifferent between the two actions when \((s_{DM}, \tilde{m}) = (L, H)\), which must yield a strictly higher expected payoff than if she relied on her prior only.

### 2.3.5 Extension: Sender Is Imperfectly Informed About \(\theta\)

We have assumed so far that the sender observes a signal which is perfectly informative about the state of the world. In this subsection, we relax this assumption in order to show that when the sender’s evidence is of imperfect quality, the DM may completely deter falsification by obtaining a private signal. Furthermore, we demonstrate that a higher quality of the DM’s signal may hurt her expected payoff. We first analyse how relaxing this assumption affects the baseline model and we then consider the DM’s incentives to acquire private independent evidence.

#### Adapting the Baseline Model

Here, we consider a model which is the same as the baseline model except that we assume that the sender observes a noisy signal of the state of the world, \(s\), where the quality of the signal is \(\Pr(s = i | \theta = \theta_i) = q\) for \(i \in \{L, H\}\) with \(q \in \left(\frac{1}{2}, 1\right]\). The following proposition states the equilibrium:

**Proposition 2.7.** Suppose that the sender’s private signal is of quality \(q \in \left(\frac{1}{2}, 1\right]\). The game has a unique equilibrium in which the sender’s and the DM’s equilibrium strategies are as follows:

- (i) if \(p \in (0, 1 - q)\), then \((\sigma_L^*, \sigma_H^*) = (0, 0)\) and \((\delta_L^*, \delta_H^*) = (0, 0)\);
- (ii) if \(p \in \left[1 - q, \frac{1-q+q(1-x)^3}{1+x}\right]\), then \((\sigma_L^*, \sigma_H^*) = \left(\frac{p+q-1}{(q-p)x}, 0\right)\) and \((\delta_L^*, \delta_H^*) = \left(0, \frac{x}{x+1}\right)\);
- (iii) if \(p \in \left[\frac{1-q+q(1-x)^3}{1+x}, q\right]\), then \((\sigma_L^*, \sigma_H^*) = (1, 0)\) and \((\delta_L^*, \delta_H^*) = (0, 1)\);
- (iv) if \(p \in [q, 1]\), then \((\sigma_L^*, \sigma_H^*) = (0, 0)\) and \((\delta_L^*, \delta_H^*) = (1, 1)\).

The imperfect quality of the sender’s signal means that it is now possible in equilibrium that the sender never falsifies, i.e. he falsifies with probability zero even upon observing \(s = L\). When the prior belief is very low or very high, the DM is so confident ex ante about the state of the world that her action does not depend on the information that she observes, and thus the sender has no incentive to falsify. However, when the DM’s prior belief is at a moderate level, her action can be swayed in one or the other direction by the information she observes. This provides an incentive for the sender
to falsify. In equilibrium, the sender falsifies with a positive probability if and only if \(1 - q < p < q\).

Another feature of the equilibrium is that the higher the quality of the signal observed by the sender, the more likely (at least weakly) the sender is to falsify upon observing \(s = L\). Intuitively, a high quality means that if the information observed by the DM happens to be not falsified, it reveals the true state with high precision. Consequently, the DM becomes more willing to make her action dependent on the information that she observes, which in turn boosts the sender’s incentive to falsify.

**Acquisition of Private Independent Evidence**

Suppose now that the DM acquires a private signal of quality \(q_{DM} \leq q\). Thus, the sender’s signal is not perfectly informative but is more informative than the DM’s private signal. In this setting, two new results arise relative to the setting where the sender is perfectly informed about \(\theta\).

The first new result is that, by obtaining a private signal, the DM can completely disincentivise falsification of evidence, i.e. she can make the sender falsify evidence with probability zero in equilibrium. We are naturally interested here in values of prior belief \(p\) such that \(1 - q < p < q\), which guarantees that the sender falsifies with a positive probability in the absence of DM’s private signal. The following proposition describes the circumstances under which obtaining a private signal by the DM completely disincentivises falsification of evidence by the sender:

**Proposition 2.8.** Suppose that the sender’s private signal is of quality \(q \in \left(\frac{1}{2}, 1\right]\) and that the DM observes a private signal of quality \(q_{DM} \in \left(\frac{1}{2}, q\right]\). Consider values of prior belief \(p\) such that \(1 - q < p < q\). Falsification of evidence can be then completely deterred in equilibrium (i.e. \(\sigma^*_L = 0\)) only if \(q < 1\). Complete deterrence occurs whenever the prior belief is such that:

\[(i)\ \beta(s_{DM} = H, \tilde{m} = H \mid \sigma_L = 0) > \frac{1}{2}, \ \beta(s_{DM} = L, \tilde{m} = H \mid \sigma_L = 0) < \frac{1}{2}, \ \text{and} \ \pi_{H|L}x \leq c, \ \text{or} \]

\[(ii)\ \beta(s_{DM} = H, \tilde{m} = L) > \frac{1}{2}, \ \beta(s_{DM} = L, \tilde{m} = L) < \frac{1}{2}, \ \text{and} \ \pi_{L|L}x \leq c, \]

where \(\pi_{i|j} = \Pr(s_{DM} = i \mid s = j)\).

The second new result is that now the DM’s welfare is not necessarily increasing in the quality of her private signal:
Proposition 2.9. Suppose that the sender’s private signal is of quality $q \in \left(\frac{1}{2}, 1\right]$ and that the DM observes a private signal of quality $q_{DM} \in \left(\frac{1}{2}, q\right)$. Then the DM’s welfare is increasing in $q_{DM}$ if $p \leq \frac{1}{2}$ but can be non-monotonic with respect to $q_{DM}$ if $p > \frac{1}{2}$.

We now discuss the results in Propositions 2.8 and 2.9 in more detail. Generally speaking, for the complete deterrence to occur, two elements are necessary. First, it must be that the sender’s message is pivotal to the DM’s decision only for one realisation of the DM’s signal. Second, the sender—upon observing his own signal—must conclude that it is unlikely that the DM has observed that particular signal realisation and hence the sender’s message is unlikely to be pivotal. The sender’s expected benefit from falsification is then lower than its cost, and so falsification is deterred. Proposition 2.8 shows that there are two instances in which falsification by the sender is completely deterred.

First, consider part (i) of Proposition 2.8 which requires a prior belief such that

$$\beta(s_{DM} = H, \tilde{m} = H \mid \sigma_L = 0) > \frac{1}{2},$$

(2.6)

and

$$\beta(s_{DM} = L, \tilde{m} = H \mid \sigma_L = 0) < \frac{1}{2},$$

(2.7)

where the latter together with $q > q_{DM}$ implies that $p < \frac{1}{2}$.

Conditions (2.6) and (2.7) define a range of values of prior belief such that the sender’s message can be pivotal to the DM’s decision only if the DM’s private signal is $s_{DM} = H$. To see this, note that (i) the condition in (2.7) and $q > q_{DM}$ imply that $\beta(s_{DM} = H, \tilde{m} = L) < \frac{1}{2}$, and (ii) the condition (2.7) also implies that $\beta(s_{DM} = L, \tilde{m} = L) < \frac{1}{2}$. However, the sender chooses to falsify with a positive probability if and only if the expected benefit from falsification exceeds the cost, i.e. $\pi_{H|L}x \geq c$ must be satisfied. If (2.6) and (2.7) are satisfied but $\pi_{H|L}x < c$, the sender realises that her message is not likely enough to be pivotal to the DM’s decision, and therefore he chooses not to falsify evidence. Falsification is thus deterred by the fact that the DM obtains a private signal.\textsuperscript{17}

An improvement in the quality of the DM’s private signal has two distinct effects on the sender’s incentives to falsify evidence. The first effect is that, for any given

\textsuperscript{17}When the sender’s private signal perfectly reveals the state of the world, i.e. $q = 1$, we have $\beta(s_{DM} = H, \tilde{m} = H \mid \sigma_L = 0) = 1$ and $\beta(s_{DM} = L, \tilde{m} = H \mid \sigma_L = 0) = 1$, which implies that only one of (2.6) and (2.7) is satisfied.
message from the sender and any given strategy of his, the DM’s belief in response to $s_{DM} = L$ decreases and her belief in response to $s_{DM} = H$ increases. In other words, (2.6) and (2.7) are more likely to be satisfied. The second effect of improving the quality of the DM’s signal is that the higher $q_{DM}$ is, the less likely it is that the DM’s private signal is $s_{DM} = H$ when the sender’s signal is $s = L$, i.e. $\pi_{H|L}$ decreases. When $q_{DM}$ becomes high enough, the condition $\pi_{H|L}x \geq c$ is no longer satisfied and the sender has no incentive to falsify.

Thus, both effects of an increase in $q_{DM}$ work to disincentivise falsification. In other words, when the prior belief is less than $\frac{1}{2}$, there may be a twofold benefit to the DM from increasing the quality of her private signal. Not only does it make her private information more precise, but it also unambiguously boosts the chances that the sender will not find it worthwhile to falsify evidence. This result explains the statement in Proposition 2.9 that the DM’s welfare is always increasing in $q_{DM}$ when $p < \frac{1}{2}$.

Let us now consider part (ii) of Proposition 2.8, which requires a prior belief such that

$$\beta (s_{DM} = H, \bar{m} = L) > \frac{1}{2}$$

and

$$\beta (s_{DM} = L, \bar{m} = L) < \frac{1}{2},$$

where the former together with $q > q_{DM}$ implies that $p > \frac{1}{2}$.

Conditions (2.8) and (2.9) define a range of values of prior belief such that the sender’s message can be pivotal to the DM’s decision only if the DM’s private signal is $s_{DM} = L$. To see this, note that (i) the condition in (2.8) implies $\beta (s_{DM} = H, \bar{m} = H) > \frac{1}{2}$, and (ii) the condition in (2.8) and $q > q_{DM}$ imply that $\beta (s_{DM} = L, \bar{m} = H | \sigma_L = 0) > \frac{1}{2}$. The sender chooses to falsify with a positive probability if and only if the expected benefit from falsification exceeds the cost, which means here that $\pi_{L|L}x \geq c$ must be satisfied. If (2.8) and (2.9) are satisfied but $\pi_{L|L}x < c$, the sender realises that her message is not likely enough to be pivotal to the DM’s decision, and therefore falsification is deterred when the DM obtains a private signal.\footnote{When the sender’s private signal perfectly reveals the state of the world, i.e. $q = 1$, we have $\beta (s_{DM} = H, \bar{m} = L) = 0$ and $\beta (s_{DM} = L, \bar{m} = L) = 0$, which implies that (2.8) and (2.9) are not simultaneously satisfied.}

As in part (i) of Proposition 2.8, a higher quality of the DM’s private signal has two effects on the sender’s incentives to falsify evidence. First, $\beta (s_{DM} = H, \bar{m} = L)$ increases while $\beta (s_{DM} = L, \bar{m} = L)$ decreases, which means that (2.8) and (2.9) are
more likely to be satisfied. In other words, we observe an increase in the range of values of the prior belief such that the sender’s message is pivotal to the DM’s action only if $s_{DM} = L$. However, the second effect works in the opposite direction to what we observed in part (i) of Proposition 2.8. The higher $q_{DM}$ is, the more likely it is that the DM’s private signal is $s_{DM} = L$ when the sender’s signal is $s = L$, i.e. $\pi_{L|L}$ increases. Thus, an increase in $q_{DM}$ can result in $\pi_{L|L} \geq c$ being satisfied. Hence, improving the quality of the DM’s signal may dramatically (and, in this binary setting, discontinuously) boost the sender’s incentives to falsify.

Overall, an increase in $q_{DM}$ has two effects on deterrence of falsification, which—unlike in part (i) of Proposition 2.8—work in opposite directions. The first one widens, while the second one narrows the range of parameter values for which the DM’s private information deters falsification. The two opposing effects can lead to a non-monotonic relationship between the quality of the DM’s signal and the DM’s expected payoff in equilibrium, as stated in Proposition 2.9 for $p > \frac{1}{2}$. In Figure 2.5, we illustrate this possibility with a numerical example:

![Figure 2.5](image)

Figure 2.5: The DM’s expected payoff in equilibrium as a function of the quality of the DM’s signal. The numerical example assumes $x = \frac{3}{4}$, $c = \frac{2}{5}$, $p = \frac{4}{5}$, and $q = \frac{17}{20}$ (depicted with the dashed line).

### 2.4 Paying Less Attention to the Sender

In the previous section, we have analysed how acquisition of private independent information can improve the DM’s decision making. In this section, we study whether she
can improve the decision making by committing to pay less attention to the sender. We then briefly discuss the interactions between this strategy and acquisition of private independent evidence.

2.4.1 Commitment to Ignorance

We assume that paying more attention is costless and allow the DM to choose ex ante a level of attention that determines the probability of her receiving the sender’s message. More formally, suppose that the DM can commit to pay attention to the sender’s message with probability $\tau$ and to ignore it with probability $1 - \tau$. If she does so, she observes the sender’s message $\tilde{m}$ with probability $\tau$, and does not observe it otherwise (and thus she keeps her prior belief about the state of the world). The baseline model corresponds to the DM choosing $\tau = 1$. We refer to the value of $\tau$ chosen by the DM as her level of attention, and to $1 - \tau$ as her level of ignorance.\footnote{In the appendix, we consider a model with a continuous—rather than a binary—choice of falsification by the sender, which yields additional intuition.}

Committing to $\tau < 1$ has two opposing effects on the DM’s expected payoff. On the one hand, it means that with probability $1 - \tau$ the DM has to rely only on her prior when making the decision. However, the benefit is that it makes the sender’s message less likely to be observed by the DM and thus less likely to influence her decision, which may weaken the sender’s incentives to falsify.

In order to disincentivise falsification, the DM needs to commit to ignore the sender’s message with a sufficiently high probability. To be more precise, $\tau$ must be below $\frac{c}{x}$. To see why, note that, from the sender’s perspective, when the DM commits to ignore his message with probability $1 - \tau$, the expected benefit from falsifying when he observes that $\theta = \theta_L$ is at most $\tau x$ (which happens when the DM’s strategy is $\delta^*_H = 1$ and $\delta^*_L = 0$), whereas the cost is $c$. If $\tau < \frac{c}{x}$, the cost of falsification exceeds the expected benefit, and therefore it is optimal for the sender not to falsify.

Therefore, from the DM’s point of view, if committing to $\tau < 1$ is optimal, the DM should choose $\tau = \frac{c}{x} - \varepsilon$, where $\varepsilon$ is arbitrarily close to zero. We will henceforth assume that when $\tau = \frac{c}{x}$, the sender does not falsify. If the DM chose $\tau < \frac{c}{x}$, she would disincentivise falsification, but she would be receiving no information other than her prior more often than necessary; if she chose $\frac{c}{x} < \tau < 1$, she would not disincentivise falsification and would be sometimes making her decision based on the prior rather than on a possibly falsified—but still to some degree informative—message from the sender.
These observations yield the following lemma:

**Lemma 2.1.** The DM’s optimal strategy is either (a) to never ignore the sender’s message, i.e. to set $\tau = 1$, or (b) to commit to ignore the sender’s message just often enough to discourage the sender from falsifying evidence, i.e. to set $\tau = \frac{c}{x}$.

### 2.4.2 Optimal Level of Attention and Ignorance

Let $\mathbb{E}(v_\tau (\sigma^*, \delta^*))$ denote the DM’s expected payoff when she pays attention with probability $\tau$ and the equilibrium strategies are $(\sigma^*, \delta^*)$. If the DM chooses not to ignore any information, i.e. she sets $\tau = 1$, her expected payoff is the same as in the equilibrium of the baseline model:

$$
\mathbb{E}(v_\tau = 1 (\sigma^*, \delta^*)) = \begin{cases} 
1 - p & \text{for } p \in \left(0, \frac{x}{1+x}\right) \\
1 - (1 - p) x & \text{for } p \in \left[\frac{x}{1+x}, 1\right]
\end{cases}
$$

(2.10)

On the other hand, if she commits to pay attention to the sender’s message with probability $\tau = \frac{c}{x}$, it is as if she received a perfectly informative signal with probability $\frac{c}{x}$ and a completely uninformative signal with probability $1 - \frac{c}{x}$. Thus, with probability $\frac{c}{x}$, she makes the correct decision, and with probability $1 - \frac{c}{x}$, she makes the decision that is dictated by the prior. Her expected payoff is then

$$
\mathbb{E}\left(v_{\tau = \frac{c}{x} (\sigma^*, \delta^*)}\right) = \begin{cases} 
1 \times \frac{c}{x} + (1 - p) \times \left(1 - \frac{c}{x}\right) & \text{for } p \in \left(0, \frac{1}{2}\right) \\
1 \times \frac{c}{x} + p \times \left(1 - \frac{c}{x}\right) & \text{for } p \in \left[\frac{1}{2}, 1\right]
\end{cases}
$$

(2.11)

The following table summarises the DM’s expected payoff in the two scenarios:

<table>
<thead>
<tr>
<th>$\frac{0}{1} &lt; p &lt; \frac{x}{1+x}$</th>
<th>$\frac{x}{1+x} &lt; p &lt; \frac{1}{2}$</th>
<th>$\frac{1}{2} &lt; p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathbb{E}(v_{\tau = 1 (\sigma^<em>, \delta^</em>)})$</td>
<td>$1 - p$</td>
<td>$p + (1 - p)(1 - x)$</td>
</tr>
<tr>
<td>$\mathbb{E}(v_{\tau = \frac{c}{x} (\sigma^<em>, \delta^</em>)})$</td>
<td>$\frac{c}{x} + (1 - p) \left(1 - \frac{c}{x}\right)$</td>
<td>$\frac{c}{x} + p \left(1 - \frac{c}{x}\right)$</td>
</tr>
</tbody>
</table>

Table 2.3: The DM’s expected payoff when she does not ignore the sender’s message (i.e. $\tau = 1$) and when she commits to ignore it with probability $1 - \frac{c}{x}$ (i.e. $\tau = \frac{c}{x}$).

By comparing the DM’s expected payoffs when $\tau = 1$ and when $\tau = \frac{c}{x}$, we can pin down the circumstances under which it is optimal for the DM to set $\tau = 1$ or $\tau = \frac{c}{x}$. The DM’s optimal choice of $\tau$ is denoted by $\tau^*$. 

74
Proposition 2.10. It is optimal for the DM to commit to ignore the sender’s message with probability \( \tau^* = \frac{c}{x} \) if and only if

\[
\min \left\{ x(1-x), x \left(1 - \frac{1-p}{p}x\right) \right\} \leq c; \tag{2.12}
\]

otherwise it is optimal for the DM not to ignore the sender’s message, i.e. to set \( \tau^* = 1 \).

Naturally, if it were that \( c > x \), the sender would never falsify evidence, so the DM would have no incentive to commit to ignore his message and hence would set \( \tau^* = 1 \). Together with Proposition 2.10, this implies that committing to ignore the sender’s message is optimal for the DM when \( \min \left\{ x(1-x), x \left(1 - \frac{1-p}{p}x\right) \right\} \leq c \leq x \), i.e. when the cost of falsification \( c \) takes moderate values.

Table 2.4 summarises the DM’s optimal choices of \( \tau \) as defined by Proposition 2.10. The analysis is performed for three different cases in which we compare the values of \( x \) and \( 1 - \frac{c}{x} \). As we will see in the next section—where we provide more intuition—\( x \) can be seen as capturing the DM’s “benefit” from commitment to ignorance, while \( 1 - \frac{c}{x} \) reflects the DM’s “cost” of commitment to ignorance. When the undetectability of falsification is such that \( x > 1 - \frac{c}{x} \), it is always optimal for the DM to commit to ignore by setting \( \tau = \frac{c}{x} \). On the other hand, when \( x < 1 - \frac{c}{x} \), the DM finds it optimal to pay full attention, i.e. to set \( \tau = 1 \), as long as prior belief \( p \) is high enough. Furthermore, this threshold value of \( p \) above which the DM optimally chooses to pay full attention, \( \tilde{p} = \frac{x}{1+x-c/x} \), decreases as the difference between \( 1 - \frac{c}{x} \) (the “cost” of ignorance) and \( x \) (the “benefit” from ignorance) increases.

<table>
<thead>
<tr>
<th>( 0 &lt; p &lt; \frac{x}{1+x} )</th>
<th>( \frac{x}{1+x} \leq p &lt; \frac{1}{2} )</th>
<th>( \frac{1}{2} \leq p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x &lt; 1 - \frac{c}{x} )</td>
<td>( \tau^* = \frac{c}{x} )</td>
<td>( \tau^* = \frac{c}{x} ) for ( p &lt; \tilde{p} ), ( \tau^* = 1 ) for ( p &gt; \tilde{p} ), where ( \tilde{p} = \frac{x}{1+x-c/x} )</td>
</tr>
<tr>
<td>( x = 1 - \frac{c}{x} )</td>
<td>( \tau^* = \frac{c}{x} )</td>
<td>( \tau^* = \frac{c}{x} )</td>
</tr>
<tr>
<td>( \text{indifferent between} )</td>
<td></td>
<td>( \tau^* = \frac{c}{x} ) and ( \tau^* = 1 )</td>
</tr>
<tr>
<td>( x &gt; 1 - \frac{c}{x} )</td>
<td></td>
<td>( \tau^* = \frac{c}{x} )</td>
</tr>
</tbody>
</table>

Table 2.4: The DM’s optimal choices of \( \tau \) as a function of \( p, c, \) and \( x \), when she can commit to ignore the sender’s message.
2.4.3 The Role of Parameters

Figure 2.6 illustrates, in the \((x, c)\) parameter space, the regions where the DM chooses not to ignore the sender’s message and where she chooses to commit to ignore it with probability \(1 - \tau^*\). There are two regions in which it is optimal for the DM not to ignore the sender’s message: the top-left triangular region and the bottom semicircular region. The former corresponds to the values of \(c\) and \(x\) which satisfy \(c > x\). The latter corresponds to the values of \(c\) and \(x\) which satisfy \(\min\{x(1-x), x(1 - \frac{1-p}{p} x)\} > c\).

We now discuss in more detail how the DM’s incentives to commit to ignore the sender’s message are affected by parameters \(c\), \(x\), and \(p\).

**Role of the cost of falsification, \(c\).** It is optimal to commit to ignore the sender’s message only if the cost of falsification is moderate. As mentioned earlier, when \(c > x\), the cost of falsification is so high that the sender never has an incentive to falsify, and hence the DM naturally prefers not to ignore his message. The more surprising result is that the DM optimally chooses not to ignore the sender’s message also when falsification is sufficiently cheap, i.e. when the condition in Proposition 2.10 is not satisfied. This occurs because a decrease in \(c\) lowers the sender’s relative cost of falsification, \(\frac{c}{x}\), which means that the DM needs to commit to ignore the sender’s message with a higher
probability if she wants to disincentivise falsification. Eventually, when \( c \) is sufficiently low, the DM prefers to observe a possibly falsified message from the sender rather than to commit to almost always ignore his message.

**Role of the undetectability of falsification, \( x \).** It can be optimal for the DM to pay full attention to the sender’s message both under high and low undetectability of falsification relative to its cost.

Naturally, it is optimal to pay full attention to the sender’s message when falsification has low undetectability relative to the cost, i.e. when \( x < c \). In that case, falsification is so easily detectable by the DM that the sender has no incentive to falsify and thus the DM prefers not to ignore the sender’s message.

However, even when \( x \geq c \), Proposition 2.10 shows that it may still be optimal for the DM to pay full attention to the sender’s message. This is because an increase in the undetectability of falsification, \( x \), has two conflicting effects on the DM’s incentives to commit to ignore the sender’s message. First, a higher value of \( x \) means that falsification is less detectable. Since the DM’s commitment to ignorance can deter falsification by the sender, this channel drives up the DM’s relative benefit from commitment to ignorance. Second, as \( x \) increases, the relative cost of falsification incurred by the sender decreases, and therefore the level of ignorance required to disincentivise falsification, \( 1 - \frac{c}{x} \), goes up. Thus, this channel makes commitment to ignorance less attractive to the DM. When \( x \) is sufficiently high, the channel which makes commitment to ignorance more attractive to the DM dominates and therefore the condition in Proposition 2.10 is satisfied. As a result, when \( x \) is sufficiently high, the DM optimally chooses to ignore the sender’s message with the required probability for all \( x \geq c \).

**Role of the prior belief, \( p \).** As for the impact of prior belief \( p \) on the DM’s incentives to commit to ignore the sender’s message, note that the condition \( c \geq x \left(1 - \frac{1-p}{p} x\right)\) in Proposition 2.10 is binding if and only if \( p < \frac{1}{2} \). This means that when \( p \geq \frac{1}{2} \), the DM prefers to commit to ignore the sender’s message only if \( c \geq x (1 - x) \). Thus, when \( p \geq \frac{1}{2} \), the DM optimally chooses \( \tau = 1 \) whenever \( c \) is below the dashed line in Figure 2.6.

On the other hand, when \( p < \frac{1}{2} \), the condition in Proposition 2.10 becomes tighter as \( p \) decreases—the cost of falsification \( c \) must then be lower and lower for it to be satisfied. Therefore, a lower value of \( p \) decreases the size of the region where the DM chooses not to ignore. Figure 2.6 shows that as \( p \) decreases to \( \frac{2}{3} \), the value of \( c \) must be
below the dotted line for the DM to prefer $\tau = 1$. Intuitively, a low value of $p$ means that the state is ex ante likely to be $\theta = \theta_L$, and hence it is likely that the sender is put in a position where he might want to falsify. Therefore, as $p$ decreases, falsification becomes more likely ex ante, which in turn makes paying full attention to the sender’s message less attractive than ignoring the sender with a probability that disincentivises falsification.

### 2.4.4 How Does the Ability to Commit to Ignore Affect the Incentives to Acquire Independent Evidence?

In Section 2.3, we have identified the circumstances under which the acquisition of private independent evidence leads to more or less falsification by the sender. But if the DM is able to commit to ignore the sender’s message, she may disincentivise falsification via this instrument, which then means that the benefit from acquiring independent evidence changes. With falsification already deterred, the benefit from acquiring a private signal derives only from its informational value and not from how it affects the sender’s falsification strategy.

Thus, the ability to commit to ignore the sender’s message may boost or reduce the DM’s incentives to acquire independent evidence. It is useful to consider here Figure 2.3. In this example, we have $c = 0.2$ and $x = 0.75$, which means that the condition $x > 1 - \frac{c}{x}$ from Table 2.4 is satisfied, and therefore it is optimal for the DM to commit to ignore the sender for all values of prior belief $p$. Whenever the acquisition of a private signal by the DM leads to more falsification by the sender (see part (ii) of Proposition 2.4 and the region labelled “more falsification” in Figure 2.3), the DM’s ability to commit to ignore is likely to increase the DM’s benefit from acquiring a private signal. On the other hand, when the acquisition of a private signal by the DM weakens the sender’s incentives to falsify (see part (iii) of Proposition 2.4 and the region labelled “less falsification” in Figure 2.3), the DM’s ability to commit to ignore is likely to decrease the DM’s benefit from acquiring a private signal. In other words, the two measures discussed in Sections 2.3 and 2.4 are likely to be complements in the former case, and substitutes in the latter case.

---

20A more complete analysis of the interactions between the DM’s incentives to acquire a private signal and her incentives to commit to ignore the sender is, however, beyond the scope of this paper.
2.5 Conclusion

Falsification of scientific evidence can lead to a significant delay in, or even prevent, regulatory action in a way that is beneficial to a specific interest group but harms the wider public. A prominent example is the impact of the tobacco industry on the general acknowledgement that smoking has a detrimental effect on human health; however, similar measures taken by interest groups have also been observed in other areas of scientific research.

In this paper, we have investigated the incentives of interest groups to falsify scientific evidence in a communication game between a sender and a decision maker where falsification is costly and can be detected with some predetermined probability. We have also analysed the measures the policymakers could take in order to prevent that. In particular, we have looked at the benefits from acquiring independent evidence and from controlling how much information to receive from the sender.

The obvious benefit from conducting one’s own independent research is that it brings additional information that can be used in the decision making process; however, it also affects the sender’s incentives to falsify evidence. As a result, the value of acquiring independent evidence may be smaller or larger than the pure information it contains. We have identified the conditions for both possibilities by first determining the circumstances under which the sender’s incentives to falsify are strengthened or weakened. The former is more likely when the quality of the decision maker’s research is low and the prior belief is against the sender’s interests, while the latter occurs when the evidence acquired by the decision maker is of high quality and the prior belief is not too much in favour of the sender’s agenda. Finally, we have demonstrated how acquisition of independent evidence may completely deter falsification, and have shown that better quality of one’s own research may not always benefit the decision maker.

In the second part of the paper, we have assumed that the decision maker can commit to ignore the sender’s message with some probability. By doing so, the decision maker can decrease the chances that the sender’s message will be pivotal to the decision, and hence may decrease his incentives to falsify evidence. We have fully characterised the conditions under which such a strategy is optimal for the sender. In general, for it to be optimal, the cost of falsification cannot be too high or too low if the undetectability of falsification is relatively low, and it cannot be too high if the undetectability of falsification is relatively high. Moreover, other things being equal, the prior belief must be to a sufficient extent against the sender’s interests. Finally, we have suggested
how this commitment ability of the decision maker can affect her incentives to acquire independent evidence.

Future research could investigate in more detail how the decision maker’s ability to control the amount of information she receives from an interest group interacts with her incentives to conduct independent research. This analysis would shed more light on the circumstances under which the two measures analysed in this paper are substitutes or complements.

Appendix to Chapter 2

B.1 Manufacturing Doubt about Scientific Consensus

In the baseline model, we have implicitly assumed that the sender can completely falsify the scientific evidence. However, in the real world, complete falsification is often not possible. Therefore, interest groups have developed strategies to “manufacture doubt”, i.e. to increase the uncertainty of the public about scientific consensus on an issue. By doing so, the interest groups attempt to persuade the public that the scientific evidence is mixed and thus delay or prevent any unfavourable regulation.

In this section, we extend the baseline model to analyse “manufacturing doubt”. We are interested in answering under what circumstances interest groups are most likely to engage in “manufacturing doubt”, in particular, how it depends on the quality of research and on the public opinion on the issue at hand.

Model

The basic structure of the game is the same as in the baseline model. There are two players: a sender and a decision maker (DM). The state of the world is binary, \( \theta \in \{ \theta_L, \theta_H \} \), and the prior belief that the state is \( \theta = \theta_H \) is \( p \in (0, 1) \). The DM chooses between accepting (denoted by \( a = A \)) and rejecting (denoted by \( a = R \)). She receives a payoff of \( v(a, \theta) = 1 \) if \( \theta = \theta_H \) and \( a = A \), and if \( \theta = \theta_L \) and \( a = R \); she receives \( v(a, \theta) = 0 \) otherwise.

The game now proceeds as follows. First, the sender privately observes a signal of the state of the world, \( s \), with a different signal structure than in the baseline model. There are three possible signal realisations: \( H \) (which denotes favourable scientific consensus), \( M \) (mixed scientific evidence), and \( L \) (unfavourable scientific consensus).
The signal structure is: \( \Pr (s = i \mid \theta = \theta_i) = q^2 \), \( \Pr (s = M \mid \theta = \theta_i) = 2q(1 - q) \), and \( \Pr (s = j \mid \theta = \theta_i) = (1 - q)^2 \), where \( i, j \in \{L, H\} \) and \( i \neq j \). Therefore, for example, if the state is \( \theta_H \), the sender observes a favourable scientific consensus with probability \( q^2 \), mixed scientific evidence with probability \( 2q(1 - q) \), and an unfavourable scientific consensus with probability \( (1 - q)^2 \). The parameter \( q \) measures here the quality of scientific research.

After observing \( s \), the sender sends a message \( m \in \{L, M, H\} \) to the DM. If \( s = L \), the sender can either send \( m = L \) at no cost or falsify evidence by sending \( m = M \) at a cost of \( c \in (0, x) \). This falsification of evidence can be interpreted as “manufacturing doubt”. If \( s = M \) or \( s = H \), the sender can only send \( m = s \) at no cost. The message received by the DM is denoted by \( \tilde{m} \). If the sender does not falsify, then the DM observes \( \tilde{m} = m = s \). If the sender falsifies, then with probability \( x \in (0, 1] \) the DM observes \( \tilde{m} = m \) and with probability \( 1 - x \) the DM observes \( \tilde{m} = s \).

After observing \( \tilde{m} \), like in the baseline model, the DM forms a belief \( \beta \) about the state of the world and chooses an action, \( a \in \{A, R\} \). The DM’s posterior belief about the state of the world upon observing \( \tilde{m} \) is denoted by \( \beta_{\tilde{m}} = \Pr (\theta = \theta_H \mid \tilde{m}) \). Given the payoff function of the DM, it is straightforward to note that (i) if \( \beta_{\tilde{m}} > \frac{1}{2} \), the DM accepts, (ii) if \( \beta_{\tilde{m}} < \frac{1}{2} \), the DM rejects, and (iii) if \( \beta_{\tilde{m}} = \frac{1}{2} \), then the DM is indifferent between accepting and rejecting.

The sender’s strategy, \( \sigma \), describes the probability with which he falsifies, with mixed strategies being possible. Since the sender can only falsify if \( s = L \), his strategy is a function

\[
\sigma : \{L\} \to \Delta \{0,1\}.
\]

We will denote by \( \sigma_L \) the probability with which the sender falsifies when he observes a signal realisation \( s = L \).

The DM’s strategy, \( \delta \), describes the probability with which she accepts, with mixed strategies possible. It is a function

\[
\delta : \{L, M, H\} \to \Delta \{A, R\}.
\]

We will denote by \( \delta_{\tilde{m}} \) the probability with which the DM accepts when she observes \( \tilde{m} \).

Thus, there are two main differences between this setup and the baseline model. First, the sender’s signal structure is different: there are three possible signal realisations, rather than two, which allows for a distinction between scientific consensus and
mixed evidence. Second, the sender has a limited ability to falsify evidence: the only possible falsification is to send \( m = M \) when \( s = L \), which can be interpreted as “manufacturing doubt”. What are the underlying assumptions that can justify this particular setup?

For the DM’s signal structure, suppose that the sender observes results of two scientific reports on the state of the world, \( \theta \in \{ \theta_L, \theta_H \} \). The results of the reports are denoted by \((r_1, r_2)\), where \( r_i \in \{ L, H \} \) for \( i = 1, 2 \). The quality of each report is given by \( \Pr(r_i = j \mid \theta = \theta_j) = q \) for \( j \in \{ L, H \} \) with \( q \in (\frac{1}{2}, 1] \). Then, if the state is \( \theta_H \), the sender observes a favourable scientific consensus \((r_1, r_2) = (H, H)\) with probability \( q^2 \), mixed scientific evidence \((r_1, r_2) \in \{(H, L), (L, H)\}\) with probability \( 2q(1 - q) \), and an unfavourable scientific consensus \((r_1, r_2) = (L, L)\) with probability \( (1 - q)^2 \). Thus, we obtain the same signal structure—with \((r_1, r_2) = (H, H), (r_1, r_2) \in \{(H, L), (L, H)\}\), and \((r_1, r_2) = (L, L)\) corresponding to \( s = H \), \( s = M \), and \( s = L \) respectively.

For the DM’s limited ability to falsify evidence, suppose that the sender can falsify it by fabricating an additional report \( r_3 = H \) at a cost \( c \in (0, x) \). If the sender fabricates, then with probability \( x \in (0, 1) \) the vector of reports becomes \((r_1, r_2, r_3) = (r_1, r_2, H)\) and with probability \( 1 - x \) it becomes \((r_1, r_2, r_3) = (r_1, r_2, \emptyset)\). If the sender does not fabricate, then the vector of reports becomes \((r_1, r_2, r_3) = (r_1, r_2, \emptyset)\). The DM observes a coarse summary, \( \hat{m} \in \{ L, M, H \} \), of the vector of reports, \((r_1, r_2, r_3)\), according to the following rules: (i) if \((r_1, r_2, r_3) = (H, H, \cdot)\), then \( \hat{m} = H \), (ii) \((r_1, r_2, r_3) \in \{(H, L, \cdot), (L, H, \cdot), (L, L, H)\}\), then \( \hat{m} = M \), and (iii) if \((r_1, r_2, r_3) = (L, L, \emptyset)\), then \( \hat{m} = L \). In other words, (i) if all existing (scientific and fabricated) reports are \( H \), then the DM observes \( \hat{m} = H \), i.e. “favourable scientific consensus”, (ii) if among all existing reports there is at least one realisation \( H \) and at least one \( L \), then the DM observes \( \hat{m} = M \), i.e. “mixed scientific evidence”, and (iii) if all existing reports are \( L \), then the DM observes \( \hat{m} = L \), i.e. “unfavourable scientific consensus”\(^{21}\).

Therefore, fabricating a report \( r_3 = H \) has no effect on the DM’s coarse summary if \((r_1, r_2) = \{(H, H), (H, L), (L, H)\}\). However, if \((r_1, r_2) = (L, L)\), then it has an effect because the DM might observe \( \hat{m} = M \) instead of \( \hat{m} = L \). In other words, the DM might have an impression of mixed scientific evidence instead of unfavourable scientific consensus.

\(^{21}\)The coarseness of the DM’s summary is a natural assumption given that interest groups make their fabricated evidence hard to distinguish from scientific evidence and given the way media report scientific evidence.
Equilibrium

We now analyse the perfect Bayesian equilibrium, which is defined analogously to the one in the baseline model.

**Proposition 2.11.** The game has a unique equilibrium, in which the sender’s and the DM’s strategies are:

(i) if \( p_1 - p \in \left( 0, \frac{(1-q^2)^2}{q^2} \right) \), then \( \sigma_L^* = 0 \) and \( (\delta_L^*, \delta_M^*, \delta_H^*) = (0, 0, 0) \);

(ii) if \( p_1 - p \in \left[ \frac{(1-q^2)^2}{q^2}, 1 \right) \), then \( \sigma_L^* = 0 \) and \( (\delta_L^*, \delta_M^*, \delta_H^*) = (0, 0, 1) \);

(iii) if \( p_1 - p \in \left[ 1, \frac{q^2 + 2q(1-q)}{(1-q^2)^2+2q(1-q)} \right] \), then \( \sigma_L^* = \frac{2q(1-q)(2p-1)}{x(q^2(1-p)-(1-q)^2p)} \) and \( (\delta_L^*, \delta_M^*, \delta_H^*) = \left( 0, \frac{x}{2}, 1 \right) \);

(iv) if \( p_1 - p \in \left[ \frac{q^2 + 2q(1-q)}{(1-q^2)^2+2q(1-q)}, \frac{q^2}{(1-q)^2} \right) \), then \( \sigma_L^* = 1 \) and \( (\delta_L^*, \delta_M^*, \delta_H^*) = (0, 1, 1) \);

(v) if \( p_1 - p \in \left[ \frac{q^2}{(1-q)^2}, +\infty \right) \), then \( \sigma_L^* = 0 \) and \( (\delta_L^*, \delta_M^*, \delta_H^*) = (1, 1, 1) \).

The conditions in parts (i)-(v) in Proposition 2.11 form five disjoint and collectively exclusive regions in the \((p, q, x, c)\) parameter space. It turns out that the sender plays a mixed strategy of “manufacturing doubt”, i.e. \( \sigma_L^* \in (0, 1) \), in part (iii) and a pure strategy, i.e. \( \sigma_L^* = 1 \), in part (iv). In parts (i), (ii), and (v), the sender does not “manufacture doubt”. Figure 2.7 illustrates the five regions in the \((p, q)\) parameter space for fixed values of \( x \) and \( c \).

![Figure 2.7](image-url)

**Figure 2.7:** An illustration of the regions defining the equilibrium strategy profiles in the extended (“manufacturing doubt”) model. The simulation uses \( x = 0.5 \).

It is instructive to analyse how the “manufacturing doubt” strategy of the sender depends on the values of parameters \( p \) and \( q \).
Corollary 2.2. (i) The probability that the sender “manufactures doubt”, i.e. sends $m = M$ upon observing $s = L$, is non-monotonic (quasi-hump-shaped) in the prior belief, $p$: it is constant and equal to 0 for low $p$ (more precisely, $p < \frac{1}{2}$), strictly increasing for moderately low $p$, constant and equal to 1 for moderately high $p$, and constant and equal to 0 for high $p$.

(ii) The probability that the sender “manufactures doubt” is non-monotonic (quasi-hump-shaped) in the quality of research, $q$, for $p \geq \frac{1}{2}$: it is constant and equal to 0 for low $q$, constant and equal to 1 for moderate $q$, and strictly decreasing for high $q$, and eventually equals 0 for $q = 1$. For $p < \frac{1}{2}$, the probability that the sender “manufactures doubt” is equal to 0 and hence does not depend on $q$.

The message behind Corollary 2.2 is that “manufacturing doubt” occurs for intermediate values of $p$ and $q$. Thus, the model predicts that interest groups will engage in “manufacturing doubt” on issues characterised by (i) high uncertainty of the public about the truth, and (ii) not too high and not too low ability of scientific research to determine the truth.

More precisely, part (i) of Corollary 2.2 says that the sender would only engage in “manufacturing doubt” when the prior belief is at an intermediate level and in favour of the sender (i.e. $p \geq \frac{1}{2}$). When the prior belief is too low or too high, the DM is so confident ex ante about the state of the world that her action does not depend on the information that she observes, and so the sender has no incentive to “manufacture doubt”. In particular, if $p < \frac{1}{2}$, then—regardless of the values of other parameters—the sender cannot have an incentive to “manufacture doubt” because $\tilde{m} = M$ cannot bring the DM’s belief above $\frac{1}{2}$ even in the absence of fabrication.

Part (ii) of Corollary 2.2 tells us that the sender would engage in “manufacturing doubt” only when the quality of research is at an intermediate level (the prior belief must also be $p \geq \frac{1}{2}$). This contrasts with Section 2.3.5 where Proposition 2.7 states that falsification is increasing in the quality of the sender’s signal. The reason for the non-monotonicity here is as follows.

If the quality of research is low, then the sender’s message $\tilde{m}$ is inevitably rather uninformative about the state of the world, so the DM updates her belief very little upon observing $\tilde{m}$ and her action does not depend on $\tilde{m}$. Thus, the sender has no incentive to “manufacture doubt”. An analogous effect leads the sender not to falsify evidence when the quality of her signal is low in the setup of Section 2.3.5.

If the quality of research is high, then the DM expects $s$ to be $L$ or $H$ with a high
probability (i.e. expects scientific consensus with a high probability), and thus realises that her observation of $\tilde{m} = M$ is likely due to the “manufacturing of doubt” by the sender. Therefore, “manufacturing doubt” decreases when the quality of research is high. This effect is absent in the setup of Section 2.3.5 where only signal realisations $L$ and $H$ are possible.

B.2 Commitment to Ignorance in a Model with a Continuous Choice of Falsification by the Sender

In order to gain a better intuition for the results on the DM’s incentives to commit to ignore the sender’s message (which are the subject of Section 2.4), it is useful to consider a model with a continuous—rather than a binary—choice of falsification by the sender. Suppose that the sender can choose the “intensity” of falsification, which we denote by $x \in [0, 1]$. When he falsifies with intensity $x$, the probability that the DM observes $\tilde{m} \neq s$ from the sender is $x$, and the probability that she observes $\tilde{m} = s$ is $1 - x$.\footnote{Thus, parameter $x$ has now a slightly different interpretation than before.} Since we assume here that $q = 1$, the signal received by the sender perfectly reveals the state of the world. The cost of falsifying with intensity $x$ is $c(x)$, where the marginal cost is $c'(x) = MC(x) > 0$ with $MC(0) = 0$ and $MC(1) \geq 1$.

The sender’s strategy, $\sigma_s$, now effectively describes the probability with which the sender falsifies upon observing a signal $s$. It is a function

$$\sigma : \{L, H\} \to [0, 1]. \quad (2.15)$$

As before, the DM’s strategy, $\delta_{\tilde{m}}$, describes the probability with which she accepts, with mixed strategies allowed. It is a function

$$\delta : \{L, H\} \to \Delta \{A, R\}. \quad (2.16)$$

**Proposition 2.12.** In the model with a continuous choice of falsification intensity by the sender, it is optimal for the DM to commit to ignore the sender’s message more often (i.e. to decrease $\tau$) if and only if

$$\frac{\partial [MC^{-1}(\tau) \cdot \tau]}{\partial \tau} \geq \min \left\{ \frac{p}{1 - p}, 1 \right\}. \quad (2.17)$$
Consider first the case where $p < \frac{1}{2}$. If the DM commits to pay attention to the sender’s message with probability $\tau$, her expected payoff is:

$$E(v_\tau(\sigma^*, \delta^*)) = \left( p + (1 - p) \left( 1 - MC^{-1}(\tau) \right) \right) \tau + (1 - p) (1 - \tau).$$

(2.18)

With probability $\tau$, the DM makes her decision based on the message received from the sender, who—upon seeing that the state is $\theta = \theta_L$—falsifies evidence with intensity $x$ such that $MC(x) = \tau$, i.e. with intensity $x^*(\tau) = MC^{-1}(\tau)$. Thus, when the state is $\theta = \theta_L$, the DM receives $\tilde{m} = H$ with probability $x^* = MC^{-1}(\tau)$, and $\tilde{m} = L$ otherwise. When the state is $\theta = \theta_H$, the DM receives $\tilde{m} = H$ with probability 1.

With probability $1 - \tau$, the DM ignores the sender’s message and therefore makes her decision solely based on her prior. Given that $p < \frac{1}{2}$, she then chooses to reject, which is the “correct” action with probability $1 - p$.

The trade-off faced by the DM is as follows. By committing to ignore the sender’s message with some probability, she ensures that the signal she receives is more precise—since the sender is falsifying with less intensity. This is because the sender’s optimal choice of the intensity, $x^*(\tau)$, is increasing in $\tau$. However, ignoring the sender’s message more often also has a cost: the DM needs to make the decision based only on her prior information more often. If the rate at which the signal becomes more precise as $\tau$ falls is sufficiently high, it is optimal for the DM to commit to ignore the sender’s message more often.

To be more concrete, for any given level of $\tau$, it is optimal for the DM to commit to ignore the sender’s message more often if and only if $\frac{\partial E(v_\tau(\sigma^*, \delta^*))}{\partial \tau} \leq 0$, which yields

$$\frac{\partial [MC^{-1}(\tau)]}{\partial \tau} \geq \frac{p}{1 - p}.$$  

(2.19)

Figure 2.8 helps visualise this. When the DM commits to ignore the sender’s message slightly more often, she marginally decreases the value of $\tau$. This leads to a decrease in the area of the grey rectangle (the area is equal to $MC^{-1}(\tau) \times \tau$). How large this decrease is depends on the slope of the marginal cost function, $MC(x)$. If $MC(x)$ has a steep slope, i.e. $MC'(x)$ is high, then a small decrease in $\tau$ results in a small decrease in $MC^{-1}(\tau)$, and so the area of the rectangle falls by little. If the slope of the $MC(x)$ curve is shallow, i.e. $MC'(x)$ is low, then a small decrease in $\tau$ results in a large decrease in $MC^{-1}(\tau)$, and thus the area of the rectangle falls significantly. Hence, committing to ignore more often is likely to be effective from the DM’s point
of view when the sender’s marginal cost of increasing \( x \) does not rise too fast as \( x \) goes up.

An analogous analysis holds for the case where \( p \geq \frac{1}{2} \). If the DM commits to pay attention to the sender’s message with probability \( \tau \), her expected payoff is:

\[
\mathbb{E}(v_{\tau}(\sigma^*, \delta^*)) = \left( p + (1 - p) \left( 1 - MC^{-1}(\tau) \right) \right) \tau + p(1 - \tau).
\]

As before, with probability \( \tau \), the DM makes her decision based on the message received from the sender, who falsifies evidence with intensity \( x^*(\tau) = MC^{-1}(\tau) \) upon seeing that the state is \( \theta_L \). However, when the DM ignores the sender’s message (which happens with probability \( 1 - \tau \)) and makes her decision based on the prior only, she now chooses to accept, which is the “correct” action with probability \( p \). Given \( (2.20) \), \( \frac{\partial \mathbb{E}(v_{\tau}(\sigma^*, \delta^*))}{\partial \tau} \leq 0 \) holds if and only if

\[
\frac{\partial [MC^{-1}(\tau) \tau]}{\partial \tau} \geq 1.
\]

**Consistency with the binary choice model.** The results in Proposition 2.12 are consistent with those for the model with a binary choice by the sender. To see this, note that when the DM considers whether to increase \( \tau \) from \( \frac{c}{x} \) to 1, we have

\[
\frac{\Delta [x^*(\tau) \tau]}{\Delta \tau} = \frac{x^*(1) \times 1 - x^*(\frac{c}{x}) \times \frac{c}{x}}{1 - \frac{c}{x}} = \frac{x \times 1 - 0 \times \frac{c}{x}}{1 - \frac{c}{x}} = \frac{x}{1 - \frac{c}{x}}.
\]
Hence, it follows from Proposition 2.12 that if \( p < \frac{1}{2} \), the DM finds it optimal to choose \( \tau = 1 \) rather than \( \tau = \frac{2}{x} \) if

\[
\frac{\Delta [x^*(\tau)]}{\Delta \tau} = \frac{x}{1 - \frac{x}{2}} \geq \frac{p}{1 - p},
\]

(2.23)

which can be rearranged to yield \( c \leq x \left( 1 - \frac{1-p}{p} x \right) \).

Furthermore, Proposition 2.12 implies that if \( p \geq \frac{1}{2} \), the DM finds it optimal to set \( \tau = 1 \) rather than \( \tau = \bar{c} \bar{x} \) if

\[
\frac{\Delta [x^*(\tau)]}{\Delta \tau} = \frac{x}{1 - \frac{x}{2}} \geq 1,
\]

(2.24)

which is equivalent to \( c \leq x (1 - x) \). We thus obtain exactly the same result as the one in Proposition 2.10.

B.3 Proofs

Proof of Proposition 2.1

It is straightforward to show that it must be that \( \sigma_H^* = 0 \) in any equilibrium.

Since \( \beta_L (\sigma_L, \sigma_H = 0) = 0 \) for any \( \sigma_L \), the DM’s sequentially rational (henceforth, abbreviated to “s.r.”) strategy satisfies \( \delta_L = 0 \). Thus, it must be \( \delta_L^* = 0 \) in any equilibrium. It remains to derive \( \sigma_L^* \) and \( \delta_H^* \). To do so, we consider two mutually exclusive and exhaustive regions of parameter values, and show that there is a unique equilibrium strategy profile in each of them.

1. Consider parameter values such that \( \beta_H (\sigma_L = 1, \sigma_H = 0) \geq \frac{1}{2} \), which is equivalent to \( p \geq \frac{x}{1+2x} \). For any \( \sigma_L \in [0, 1] \) and \( \sigma_H = 0 \), given that \( \beta_H (\sigma_L = 1, \sigma_H = 0) \geq \frac{1}{2} \), the DM’s s.r. strategy satisfies \( \delta_H = 1 \). Now, given that \( c \leq x \), \( \delta_L = 0 \), and \( \delta_H = 1 \), the sender’s expected payoff from falsifying upon observing \( s = L \) is \( x (\delta_H - \delta_L) - c > 0 \), and thus the sender’s s.r. strategy satisfies \( \sigma_L = 1 \). Thus, for \( p \geq \frac{x}{1+2x} \), the unique equilibrium strategies are \( \sigma_L^* = 1 \) and \( \delta_H^* = 1 \).

2. Consider parameter values such that \( \beta_H (\sigma_L = 1, \sigma_H = 0) < \frac{1}{2} \), which is equivalent to \( p < \frac{x}{1+2x} \).

(a) If we suppose that \( \sigma_L = 1 \), the DM’s s.r. strategy satisfies \( \delta_H = 0 \), which makes \( \sigma_L = 1 \) not s.r., a contradiction.
(b) If we suppose that $\sigma_L = 0$, the DM’s s.r. strategy satisfies $\delta_H = 1$, which makes $\sigma_L = 0$ not s.r., a contradiction.

(c) Suppose now that $\sigma_L \in (0, 1)$. If $\beta_H (\sigma_L, \sigma_H = 0) > \frac{1}{2}$, then the DM’s s.r. strategy satisfies $\delta_H = 1$, which makes $\sigma_L \in (0, 1)$ not s.r., a contradiction. If $\beta_H (\sigma_L, \sigma_H = 0) < \frac{1}{2}$, then the DM’s s.r. strategy satisfies $\delta_H = 0$, which makes $\sigma_L \in (0, 1)$ not s.r., a contradiction. Suppose then that $\beta_H (\sigma_L, \sigma_H = 0) = \frac{1}{2}$, which pins down $\sigma_L \in (0, 1)$ to $\sigma_L = \frac{p}{(1-p)x}$. If $\delta_H - \delta_L < 0$, where $\delta_L = 0$, then $\sigma_L \in (0, 1)$ is not s.r. ($\sigma_L = 1$ is). If $\delta_H - \delta_L > 0$, then $\sigma_L \in (0, 1)$ is not s.r. ($\sigma_L = 0$ is). If $\delta_H - \delta_L = 0$, then the sender is indifferent between falsifying and not falsifying upon observing $s = L$, so $\sigma_L = \frac{p}{(1-p)x} \in (0, 1)$ is s.r. This uniquely pins down $\delta_H = \frac{c}{x}$. Thus, for $p < \frac{x}{1+x}$, the unique equilibrium strategies are $\sigma_L^* = \frac{p}{(1-p)x}$ and $\delta_H^* = \frac{c}{x}$.

**Proof of Proposition 2.2**

**Part (i).** When $x \to 0$, the equilibrium strategies are $(\sigma_L^*, \sigma_H^*) = (1, 0)$ and $(\delta_L^*, \delta_H^*) = (0, 1)$ for all $p \in (0, 1)$. Given that $x \to 0$ and these equilibrium strategies, we have $\mathbb{E} (v (\sigma^*, \delta^*)) \to p \cdot 1 + (1-p) \cdot 1 = 1$, $\Pr (a = A \mid \theta = \theta_H) \to 1$, and $\Pr (a = A \mid \theta = \theta_L) \to 0$.

In a model of verifiable disclosure, in equilibrium, the sender effectively truthfully reveals $s = H$ and $s = L$, so the DM learns $s$ and takes action accordingly. This gives us $\mathbb{E} (v (\sigma^*, \delta^*)) = p \cdot 1 + (1-p) \cdot 1 = 1$, $\Pr (a = A \mid \theta = \theta_H) = 1$, and $\Pr (a = A \mid \theta = \theta_L) = 0$.

**Part (ii).** When $x \to 1$ and $c \to 0$, the equilibrium strategies are $\sigma_L^* \to \frac{p}{1-p}$, $\sigma_H^* = 0$, $\delta_L^* = 0$, $\delta_H^* \to 0$ for $p \in \left(0, \frac{1}{2}\right)$, and $(\sigma_L^*, \sigma_H^*) = (1, 0)$ and $(\delta_L^*, \delta_H^*) = (0, 1)$ for $p \in \left[\frac{1}{2}, 1\right)$.

Given that $x \to 1$ and $c \to 0$ and these equilibrium strategies, $\mathbb{E} (v (\sigma^*, \delta^*)) \to 1 - p$ for $p \in \left(0, \frac{1}{2}\right)$, $\mathbb{E} (v (\sigma^*, \delta^*)) \to p$ for $p \in \left[\frac{1}{2}, 1\right)$, $\Pr (a = A \mid \theta = \theta_H) \to 0$ for $p \in \left(0, \frac{1}{2}\right)$, $\Pr (a = A \mid \theta = \theta_L) \to 1$ for $p \in \left[\frac{1}{2}, 1\right)$, $\Pr (a = A \mid \theta = \theta_H) \to 1$ for $p \in \left[\frac{1}{2}, 1\right)$, $\Pr (a = A \mid \theta = \theta_L) \to 0$ for $p \in \left(0, \frac{1}{2}\right)$.

In a model of cheap talk, given the misaligned preferences of the sender and the DM, in equilibrium, the sender sends an uninformative (“babbling”) message to the DM regardless of the observed $s$. Effectively, the DM does not learn anything about $s$ and keeps her prior belief, and takes action accordingly. This gives us $\mathbb{E} (v (\sigma^*, \delta^*)) = 1 - p$ for $p \in \left(0, \frac{1}{2}\right)$, $\mathbb{E} (v (\sigma^*, \delta^*)) = p$ for $p \in \left[\frac{1}{2}, 1\right)$, $\Pr (a = A \mid \theta = \theta_H) = 0$ for $p \in \left(0, \frac{1}{2}\right)$, $\Pr (a = A \mid \theta = \theta_L) = 1$ for $p \in \left[\frac{1}{2}, 1\right)$, $\Pr (a = A \mid \theta = \theta_L) = 0$ for $p \in \left(0, \frac{1}{2}\right)$,
\[ \Pr(a = A \mid \theta = \theta_L) = 1 \text{ for } p \in \left[\frac{1}{2}, 1\right). \]

**Part (iii).** When \( x \to 1 \) and \( c \to 1 \), the equilibrium strategies are \( \sigma^*_L \rightarrow \frac{p}{1-p}, \sigma^*_H = 0, \delta^*_L = 0, \delta^*_H \to 1 \) for \( p \in (0, \frac{1}{2}) \), and \((\sigma^*_L, \sigma^*_H) = (1,0)\) and \((\delta^*_L, \delta^*_H) = (0,1)\) for \( p \in \left[\frac{1}{2}, 1\right) \).

Given that \( x \to 1 \) and \( c \to 1 \) and these equilibrium strategies, \( \mathbb{E}(v(\sigma^*, \delta^*)) \to 1 - p \) for \( p \in \left(0, \frac{1}{2}\right) \), \( \mathbb{E}(v(\sigma^*, \delta^*)) \to p \) for \( p \in \left[\frac{1}{2}, 1\right) \), \( \Pr(a = A \mid \theta = \theta_H) \to 1 \) for \( p \in (0, 1) \), \( \Pr(a = A \mid \theta = \theta_L) \to 1 \) for \( p \in \left[\frac{1}{2}, 1\right). \)

In a model of Bayesian persuasion, note that the receiver is willing to take action \( a = A \) if \( \Pr(\theta = \theta_H \mid m = H) \geq \frac{1}{2} \) (we assume here that the DM takes action \( a = A \) when indifferent). Therefore, the sender (the designer) commits to the following signal structure: \( \Pr(m = H \mid \theta = \theta_H) = 1 \) for \( p \in (0, 1) \), \( \Pr(m = H \mid \theta = \theta_L) = \frac{p}{1-p} \) for \( p \in \left(0, \frac{1}{2}\right) \), and \( \Pr(m = H \mid \theta = \theta_L) = 1 \) for \( p \in \left[\frac{1}{2}, 1\right) \). Given this strategy of the sender, the DM’s posterior beliefs are \( \Pr(\theta = \theta_H \mid m = H) = \frac{1}{2} \) when \( p < \frac{1}{2} \), and \( \Pr(\theta = \theta_H \mid m = H) = p \) when \( p \geq \frac{1}{2} \), which means that (i) the DM’s strategy is to take action \( a = A \) upon observing \( m = H \) and \( a = R \) upon observing \( m = L \), and (ii) the probability of the DM accepting conditional on the state of the world is maximised for any given \( p \). This gives us \( \mathbb{E}(v(\sigma^*, \delta^*)) = 1 - p \) for \( p \in \left(0, \frac{1}{2}\right) \), \( \mathbb{E}(v(\sigma^*, \delta^*)) = p \) for \( p \in \left[\frac{1}{2}, 1\right) \), \( \Pr(a = A \mid \theta = \theta_H) = 1 \) for \( p \in (0, 1) \), \( \Pr(a = A \mid \theta = \theta_L) = \frac{p}{1-p} \) for \( p \in \left(0, \frac{1}{2}\right) \), and \( \Pr(a = A \mid \theta = \theta_L) = 1 \) for \( p \in \left[\frac{1}{2}, 1\right) \).

**Proof of Proposition 2.3**

Let \( p^- \) denote the DM’s belief upon observing \( s_{DM} = L \), i.e. \( p^- = \frac{p(1-q_{DM})}{p(1-q_{DM}) + (1-p)q_{DM}} \), and let \( p^+ \) denote the DM’s belief upon observing \( s_{DM} = H \), i.e. \( p^+ = \frac{p q_{DM}}{p q_{DM} + (1-p)(1-q_{DM})} \).

It is straightforward to show that it must be that \( \sigma^*_H = 0 \) in any equilibrium. Since \( \beta_{LL}(\sigma_L, \sigma_H = 0) = \beta_{HL}(\sigma_L, \sigma_H = 0) = 0 \) for any \( \sigma_L \), the DM’s sequentially rational (henceforth, abbreviated to “s.r.”) strategy satisfies \( \delta_{LL} = \delta_{HL} = 0 \). Thus, it must be \( \delta^*_L = \delta^*_H = 0 \) in any equilibrium. It remains to derive \( \sigma^*_L, \delta^*_LH, \) and \( \delta^*_HH \). To do so, we consider four mutually exclusive regions of parameter values, and show that there is a unique equilibrium strategy profile in each of them. In the remaining knife-edge cases, we select the sender-preferred strategy profile whenever there are multiple equilibria.

1. Consider parameter values such that \( \beta_{LH}(\sigma_L = 1, \sigma_H = 0) > \frac{1}{2} \), which is equivalent to \( p^- > \frac{p}{1+p} \). For any \( \sigma_L \in (0, 1] \) and \( \sigma_H = 0 \), given that \( \beta_{LH}(\sigma_L = 1, \sigma_H = 0) > \frac{1}{2} \) (and thus \( \beta_{HH}(\sigma_L = 1, \sigma_H = 0) > \frac{1}{2} \)), the DM’s s.r. strategy satisfies \( \delta_{LH} = 1 \).
and $\delta_{HH} = 1$. Now, given that $c < x$, the sender’s expected payoff from falsifying upon observing $s = L$ is $q_{DM}x(\delta_{LH} - \delta_{LL}) + (1 - q_{DM})x(\delta_{HH} - \delta_{HL}) - c = x - c > 0$, and thus the sender’s s.r. strategy satisfies $\sigma_L = 1$. Thus, for $p^- > \frac{x}{1 + x}$, the unique equilibrium strategy profile satisfies $\sigma^*_L = 1$, $\delta^*_LH = 1$, and $\delta^*_{HH} = 1$.

2. Consider parameter values such that $\beta_{LH}(\sigma_L = 1, \sigma_H = 0) < \frac{1}{2}$ (which is equivalent to $p^- < \frac{x}{1 + x}$), $\beta_{HH}(\sigma_L = 1, \sigma_H = 0) > \frac{1}{2}$ (which is equivalent to $p^+ > \frac{x}{1 + x}$), and $q_{DM} < 1 - \frac{c}{x}$. For any $\sigma_L \in [0,1]$ and $\sigma_H = 0$, given that $\beta_{HH}(\sigma_L = 1, \sigma_H = 0) > \frac{1}{2}$, the DM’s s.r. strategy satisfies $\delta_{HH} = 1$. Now, given that $q_{DM} < 1 - \frac{c}{x}$, the sender’s expected payoff from falsifying upon observing $s = L$ is $q_{DM}x(\delta_{LH} - \delta_{LL}) + (1 - q_{DM})x(\delta_{HH} - \delta_{HL}) - c = q_{DM}x\delta_{LH} + (1 - q_{DM})x - c > 0$, and thus the sender’s s.r. strategy satisfies $\sigma_L = 1$. Thus, for $p^- < \frac{x}{1 + x}, p^+ > \frac{x}{1 + x},$ and $c < (1 - q_{DM})x$, the unique equilibrium strategy profile satisfies $\sigma^*_L = 1$, $\delta^*_LH = 0$, and $\delta^*_{HH} = 1$.

3. Consider parameter values such that $\beta_{LH}(\sigma_L = 1, \sigma_H = 0) < \frac{1}{2}$ (which is equivalent to $p^- < \frac{x}{1 + x}$), $\beta_{HH}(\sigma_L = 1, \sigma_H = 0) < \frac{1}{2}$ (which is equivalent to $p^+ < \frac{x}{1 + x}$), and $q_{DM} < 1 - \frac{c}{x}$. If we suppose that $\sigma_L = 1$, the DM’s s.r. strategy satisfies $\delta_{LH} = 0$ and $\delta_{HH} = 0$, which makes $\sigma_L = 1$ not s.r., a contradiction. If we suppose that $\sigma_L = 0$, the DM’s s.r. strategy satisfies $\delta_{LH} = 1$ and $\delta_{HH} = 1$, which makes $\sigma_L = 0$ not s.r., a contradiction. Suppose now that $\sigma_L \in (0,1)$. There are then five possibilities to consider.

(a) If $\beta_{LH}(\sigma_L, \sigma_H = 0) > \frac{1}{2}$ and $\beta_{HH}(\sigma_L, \sigma_H = 0) > \frac{1}{2}$, then the DM’s s.r. strategy satisfies $\delta_{LH} = 1$ and $\delta_{HH} = 1$, so the sender’s expected payoff from falsifying upon observing $s = L$ is $q_{DM}x(\delta_{LH} - \delta_{LL}) + (1 - q_{DM})x(\delta_{HH} - \delta_{HL}) - c = x - c > 0$, which makes $\sigma_L \in (0,1)$ not s.r., a contradiction.

(b) If $\beta_{LH}(\sigma_L, \sigma_H = 0) < \frac{1}{2}$ and $\beta_{HH}(\sigma_L, \sigma_H = 0) > \frac{1}{2}$, then the DM’s s.r. strategy satisfies $\delta_{LH} = 0$ and $\delta_{HH} = 1$, so the sender’s expected payoff from falsifying upon observing $s = L$ is $q_{DM}x(\delta_{LH} - \delta_{LL}) + (1 - q_{DM})x(\delta_{HH} - \delta_{HL}) - c = (1 - q_{DM})x - c > 0$, which makes $\sigma_L \in (0,1)$ not s.r., a contradiction.

(c) If $\beta_{LH}(\sigma_L, \sigma_H = 0) < \frac{1}{2}$ and $\beta_{HH}(\sigma_L, \sigma_H = 0) < \frac{1}{2}$, then the DM’s s.r. strategy satisfies $\delta_{LH} = 0$ and $\delta_{HH} = 0$, so the sender’s expected payoff from falsifying upon observing $s = L$ is $q_{DM}x(\delta_{LH} - \delta_{LL}) + (1 - q_{DM})x(\delta_{HH} - \delta_{HL}) - c = -c < 0$, which makes $\sigma_L \in (0,1)$ not s.r., a contradiction.
(d) If $\beta_{LH}(\sigma_L, \sigma_H = 0) = \frac{1}{2}$ and $\beta_{HH}(\sigma_L, \sigma_H = 0) > \frac{1}{2}$, then the DM’s s.r. strategy satisfies $\delta_{HH} = 1$, so the sender’s expected payoff from falsifying upon observing $s = L$ is $q_{DM} x \delta_{LH} + (1 - q_{DM}) x - c > 0$, which makes $\sigma_L \in (0, 1)$ not s.r., a contradiction.

(e) If $\beta_{LH}(\sigma_L, \sigma_H = 0) < \frac{1}{2}$ and $\beta_{HH}(\sigma_L, \sigma_H = 0) = \frac{1}{2}$, then the DM’s s.r. strategy satisfies $\delta_{LH} = 0$ and $\sigma_L$ is pinned down to $\sigma_L = \frac{p^+}{(1 - p^+) x}$. If either $(1 - q_{DM}) x \delta_{HH} - c < 0$ or $(1 - q_{DM}) x \delta_{HH} - c > 0$, then $\sigma_L \in (0, 1)$ is not s.r., a contradiction. If $(1 - q_{DM}) x \delta_{HH} - c = 0$, the sender is indifferent between falsifying and not falsifying upon observing $s = L$, so $\sigma_L = \frac{p^+}{(1 - p^+) x} \in (0, 1)$ is s.r. This uniquely pins down $\delta_{HH} = \frac{c}{(1 - q_{DM}) x}$. Thus, for $p^- < \frac{x}{1 + x}$, $p^+ < \frac{x}{1 + x}$, and $q_{DM} < 1 - \frac{c}{x}$, the unique equilibrium strategy profile satisfies $\sigma_L^* = \frac{p^+}{(1 - p^+) x}$, $\delta_{LH}^* = 0$, and $\delta_{HH}^* = \frac{c}{(1 - q_{DM}) x}$.

4. Consider parameter values such that $\beta_{LH}(\sigma_L = 1, \sigma_H = 0) < \frac{1}{2}$, which is equivalent to $p^- < \frac{x}{1 + x}$, and $q_{DM} > 1 - \frac{c}{x}$. If we suppose that $\sigma_L = 1$, the DM’s s.r. strategy satisfies $\delta_{LH} = 0$, so the sender’s expected payoff from falsifying upon observing $s = L$ is $(1 - q_{DM}) x \delta_{HH} - c < 0$, which makes $\sigma_L = 1$ not s.r., a contradiction. If we suppose that $\sigma_L = 0$, the DM’s unique s.r. action is $\delta_{LH} = 1$, so the sender’s expected payoff from falsifying upon observing $s = L$ is $x - c > 0$, which makes $\sigma_L = 0$ not s.r., a contradiction. Suppose now that $\sigma_L \in (0, 1)$. There are then five possibilities to consider.

(a) If $\beta_{LH}(\sigma_L, \sigma_H = 0) > \frac{1}{2}$ and $\beta_{HH}(\sigma_L, \sigma_H = 0) > \frac{1}{2}$, then the DM’s s.r. strategy satisfies $\delta_{LH} = 1$ and $\delta_{HH} = 1$, so the sender’s expected payoff from falsifying upon observing $s = L$ is $q_{DM} x (\delta_{LH} - \delta_{LL}) + (1 - q_{DM}) x (\delta_{HH} - \delta_{HL}) - c = x - c > 0$, which makes $\sigma_L \in (0, 1)$ not s.r., a contradiction.

(b) If $\beta_{LH}(\sigma_L, \sigma_H = 0) < \frac{1}{2}$ and $\beta_{HH}(\sigma_L, \sigma_H = 0) > \frac{1}{2}$, then the DM’s s.r. strategy satisfies $\delta_{LH} = 0$ and $\delta_{HH} = 1$, so the sender’s expected payoff from falsifying upon observing $s = L$ is $q_{DM} x (\delta_{LH} - \delta_{LL}) + (1 - q_{DM}) x (\delta_{HH} - \delta_{HL}) - c = (1 - q_{DM}) x - c < 0$, which makes $\sigma_L \in (0, 1)$ not s.r., a contradiction.

(c) If $\beta_{LH}(\sigma_L, \sigma_H = 0) < \frac{1}{2}$ and $\beta_{HH}(\sigma_L, \sigma_H = 0) < \frac{1}{2}$, then the DM’s s.r. strategy satisfies $\delta_{LH} = 0$ and $\delta_{HH} = 0$, so the sender’s expected payoff from falsifying upon observing $s = L$ is $q_{DM} x (\delta_{LH} - \delta_{LL}) + (1 - q_{DM}) x (\delta_{HH} - \delta_{HL}) - c = -c < 0$, which makes $\sigma_L \in (0, 1)$ not s.r., a contradiction.
(d) If \( \beta_{LH}(\sigma_L, \sigma_H = 0) < \frac{1}{2} \) and \( \beta_{HH}(\sigma_L, \sigma_H = 0) = \frac{1}{2} \), the DM’s s.r. strategy satisfies \( \delta_{LH} = 0 \), so the sender’s expected payoff from falsifying upon observing \( s = L \) is \( (1 - q_{DM})x\delta_{HH} - c < 0 \), which makes \( \sigma_L \in (0, 1) \) not s.r., a contradiction.

(e) If \( \beta_{LH}(\sigma_L, \sigma_H = 0) = \frac{1}{2} \) and \( \beta_{HH}(\sigma_L, \sigma_H = 0) > \frac{1}{2} \), then the DM’s s.r. strategy satisfies \( \delta_{HH} = 1 \) and \( \sigma_L \) is pinned down to \( \sigma_L = \frac{p^\ast}{(1-p^\ast)x} \). If either \( qx\delta_{LH} - c < 0 \) or \( (1 - q_{DM})x\delta_{HH} - c > 0 \), then \( \sigma_L \in (0, 1) \) is not s.r., a contradiction. If \( qx\delta_{LH} - c = 0 \), the sender is indifferent between falsifying and not falsifying upon observing \( s = L \), so \( \sigma_L = \frac{p^\ast}{(1-p^\ast)x} \in (0, 1) \) is s.r. This uniquely pins down \( \delta_{LH} = \frac{c}{q_{DM}x} \). Thus, for \( p^\ast < \frac{x}{1+x} \) and \( q_{DM} > 1 - \frac{c}{x} \), the unique equilibrium strategy profile satisfies \( \sigma_L^\ast = \frac{p^\ast}{(1-p^\ast)x}, \delta_{LH}^\ast = \frac{c}{q_{DM}x} \), and \( \delta_{HH}^\ast = 1 \).

We thus obtain Table 2.1

**Proof of Proposition 2.4**

Using the results from Proposition 2.1 and 2.3 we can compare the sender’s equilibrium strategy in the baseline model and in the model in which the DM observes a private signal, for all possible regions of parameter values. The results in Proposition 3 follow from Table 2.1. Note here that, since \( p^- < p < p^+ \), which follows from \( q_{DM} > \frac{1}{2} \), we also have \( \frac{p^-}{(1-p^-)x} < \frac{p}{(1-p)x} < \frac{p^+}{(1-p^+)x} \).

**Proof of Proposition 2.5**

Using the results from Proposition 2.1 and 2.3, we can compare the DM’s welfare in the baseline model and in the model in which the DM observes a private signal of quality \( q_{DM} \), for all possible regions of parameter values. For brevity, let \( W\left(\frac{1}{2}\right) = \mathbb{E}\left[ v\left(\sigma^\ast\left(\frac{1}{2}\right), \delta^\ast\left(\frac{1}{2}\right)\right)\right] \) be the DM’s welfare in the baseline model, and let \( W(q_{DM}) = \mathbb{E}\left[ v\left(\sigma^\ast(q_{DM}), \delta^\ast(q_{DM})\right)\right] \) be the DM’s welfare in the model where she observes a private signal of quality \( q_{DM} \). Thus, \( V = W(q_{DM}) - W\left(\frac{1}{2}\right) \). We consider five mutually exclusive and exhaustive regions of parameter values.

1. Consider parameter values such that \( p^- \geq \frac{x}{1+x} \). When \( q_{DM} = \frac{1}{2} \), the equilibrium strategies are \( (\sigma_L^\ast, \sigma_H^\ast) = (1, 0) \) and \( (\delta_L^\ast, \delta_H^\ast) = (0, 1) \). When \( q_{DM} > \frac{1}{2} \), they are \( (\sigma_L^\ast, \sigma_H^\ast) = (1, 0) \) and \( (\delta_L^\ast, \delta_H^\ast, \delta_L^\ast, \delta_H^\ast, \delta_{HH}^\ast) = (0, 0, 1, 1) \). We can then derive \( W\left(\frac{1}{2}\right) = p + (1-p)(1-x) \) and \( W(q_{DM}) = p + (1-p)(1-x) \). Hence, \( V = 0 \).
2. Consider parameter values such that \( q_{DM} < 1 - \frac{c}{x} \) and \( p^+ < \frac{x}{1+\xi} \). When \( q_{DM} = \frac{1}{2} \), the equilibrium strategies are \( (\sigma^*_L, \sigma^*_H) = \left( \frac{p}{(1-p)x}, 0 \right) \) and \( (\delta^*_L, \delta^*_H) = \left( 0, \frac{c}{x} \right) \). When \( q_{DM} > \frac{1}{2} \), they are \( (\sigma^*_L, \sigma^*_H) = \left( \frac{p^+}{(1-p^+)^x}, 0 \right) \) and \( (\delta^*_L, \delta^*_H) = \left( 0, 0, 0, \frac{c}{1-q_{DM}x} \right) \). We can then derive \( W \left( \frac{1}{2} \right) = 1 - p \) and \( W(q_{DM}) = 1 - p \). Hence, \( V = 0 \).

3. Consider parameter values such that \( q_{DM} < 1 - \frac{c}{x} \) and \( p < \frac{x}{1+\xi} \leq p^+ \). When \( q_{DM} = \frac{1}{2} \), the equilibrium strategies are \( (\sigma^*_L, \sigma^*_H) = \left( \frac{p}{(1-p)x}, 0 \right) \) and \( (\delta^*_L, \delta^*_H) = \left( 0, \frac{c}{x} \right) \). When \( q_{DM} > \frac{1}{2} \), they are \( (\sigma^*_L, \sigma^*_H) = (1, 0) \) and \( (\delta^*_L, \delta^*_H) = (0, 0, 0, 1) \). We can then derive \( W \left( \frac{1}{2} \right) = 1 - p \), \( W(q_{DM}) = p[q_{DM} + (1 - p) (1 - x) + q_{DM}x] \). The condition \( p^+ \geq \frac{x}{1+\xi} \) implies that \( V = W(q_{DM}) - W \left( \frac{1}{2} \right) = pq_{DM} + (1 - p^+ (1 - q_{DM}) x \geq 0 \), with the inequality being strict whenever \( p^+ > \frac{x}{1+\xi} \).

4. Consider parameter values such that \( q_{DM} < 1 - \frac{c}{x} \) and \( p^- < \frac{x}{1+\xi} \leq p \). When \( q_{DM} = \frac{1}{2} \), the equilibrium strategies are \( (\sigma^*_L, \sigma^*_H) = (1, 0) \) and \( (\delta^*_L, \delta^*_H) = (0, 1) \). When \( q_{DM} > \frac{1}{2} \), they are \( (\sigma^*_L, \sigma^*_H) = (1, 0) \) and \( (\delta^*_L, \delta^*_H) = (0, 0, 0, 1) \). We can then derive \( W \left( \frac{1}{2} \right) = p + (1 - p) (1 - x) \) and \( W(q_{DM}) = p[q_{DM} + (1 - p) [1 - (1 - \sigma^*_L) + \sigma^*_L (1 - x)] \), where \( \sigma^*_L = \frac{p^-}{(1-p^+)^x} \). It follows that \( V = W(q_{DM}) - W \left( \frac{1}{2} \right) = (1 - p) (1 - \sigma^*_L) x > 0 \), where \( \sigma^*_L = \frac{p^-}{(1-p^+)^x} \).

5. We need to consider two subregions:

(a) Consider parameter values such that \( q_{DM} \geq 1 - \frac{c}{x} \) and \( p^- < \frac{x}{1+\xi} \leq p \). When \( q_{DM} = \frac{1}{2} \), the equilibrium strategies are \( (\sigma^*_L, \sigma^*_H) = (1, 0) \) and \( (\delta^*_L, \delta^*_H) = (0, 1) \). On the other hand, when \( q_{DM} > \frac{1}{2} \), they are \( (\sigma^*_L, \sigma^*_H) = \left( \frac{p^-}{(1-p^+)^x}, 0 \right) \) and \( (\delta^*_L, \delta^*_H, \delta^*_L, \delta^*_H) = \left( 0, 0, \frac{c(1-q_{DM}x)}{q_{DM}x}, 1 \right) \). We can then derive \( W \left( \frac{1}{2} \right) = p + (1 - p) (1 - x) \) and \( W(q_{DM}) = p + (1 - p) [(1 - \sigma^*_L) + \sigma^*_L (1 - x)] \), where \( \sigma^*_L = \frac{p^-}{(1-p^+)^x} \). It follows that \( V = W(q_{DM}) - W \left( \frac{1}{2} \right) = (1 - p) (1 - \sigma^*_L) x > 0 \), where \( \sigma^*_L = \frac{p^-}{(1-p^+)^x} \).

(b) Consider parameter values such that \( q_{DM} \geq 1 - \frac{c}{x} \) and \( p^- < \frac{x}{1+\xi} \leq p \). When \( q_{DM} = \frac{1}{2} \), the equilibrium strategies are \( (\sigma^*_L, \sigma^*_H) = \left( \frac{p^-}{(1-p)^x}, 0 \right) \) and \( (\delta^*_L, \delta^*_H) = \left( 0, \frac{c}{x} \right) \). When \( q_{DM} > \frac{1}{2} \), they are \( (\sigma^*_L, \sigma^*_H) = \left( \frac{p^-}{(1-p)^x}, 0 \right) \) and \( (\delta^*_L, \delta^*_H, \delta^*_L, \delta^*_H) = \left( 0, 0, \frac{c(1-q_{DM}x)}{q_{DM}x}, 1 \right) \). We can then derive \( W \left( \frac{1}{2} \right) = 1 - p \) and \( W(q_{DM}) = p + (1 - p) [(1 - \sigma^*_L) + \sigma^*_L (1 - x)] \), where \( \sigma^*_L = \frac{p^-}{(1-p)^x} \). It follows that \( V = W(q_{DM}) - W \left( \frac{1}{2} \right) = p - \frac{1-p^-}{1-p} p^- > 0 \), where \( \sigma^*_L = \frac{p^-}{(1-p)^x} \).
Proof of Corollary 2.1

First, we consider the DM’s expected payoff in five mutually exclusive and exhaustive regions of the parameter values. Let $W(q_{DM})$ denote the DM’s welfare in the model in which she observes a private signal of quality $q_{DM}$, $\mathbb{E}[v(\sigma^*(q_{DM}), \delta^*(q_{DM}))].$

1. Consider parameter values such that $p^- \geq \frac{x}{1+x}$. The DM’s welfare is $W(q_{DM}) = p + (1 - p)(1 - x)$. Then, $\frac{\partial W(q_{DM})}{\partial q_{DM}} = 0$.

2. Consider parameter values such that $q_{DM} < 1 - \frac{x}{x}$ and $p^+ < \frac{x}{1+x}$. The DM’s welfare is $W(q_{DM}) = 1 - p$. Then $\frac{\partial W(q_{DM})}{\partial q_{DM}} = 0$.

3. Consider parameter values such that $q_{DM} < 1 - \frac{x}{x}$ and $p < \frac{x}{1+x} \leq p^+$. The DM’s welfare is $W(q_{DM}) = p [q_{DM}] + (1 - p) [(1 - x) + q_{DM} x]$. Then, $\frac{\partial W(q_{DM})}{\partial q_{DM}} = p + (1 - p)x > 0$.

4. Consider parameter values such that $q_{DM} < 1 - \frac{x}{x}$ and $p^- < \frac{x}{1+x} \leq p$. The DM’s welfare is $W(q_{DM}) = p [q_{DM}] + (1 - p) [(1 - x) + q_{DM} x]$. Then, $\frac{\partial W(q_{DM})}{\partial q_{DM}} = p + (1 - p)x > 0$.

5. Consider parameter values such that $q_{DM} \geq 1 - \frac{x}{x}$ and $p^- < \frac{x}{1+x}$. The DM’s welfare is $W(q_{DM}) = p + (1 - p) [(1 - \sigma^*_L - x) + \sigma^*_L (1 - x)]$, where $\sigma^*_L = \frac{p^-}{(1-p^-)}$. Then, $\frac{\partial W(q_{DM})}{\partial q_{DM}} = (1 - p)(-x) \frac{\partial \sigma^*_L}{\partial q_{DM}}$, where $\frac{\partial \sigma^*_L}{\partial q_{DM}} = \frac{p(p-q_{DM} x_i)}{(1-p)q_{DM} x_i}$. From $p^- < \frac{x}{1+x}$ it follows that $p^- < \frac{1}{2}$, which is equivalent to $p < q_{DM}$. Hence, $\frac{\partial \sigma^*_L}{\partial q_{DM}} < 0$, and so $\frac{\partial W(q_{DM})}{\partial q_{DM}} > 0$.

Second, we consider whether the DM’s expected payoff is increasing in $q_{DM}$ at the boundaries of the above five regions. It suffices to check here (as can also be seen from Figure 2.4) whether

$$
\lim_{q_{DM} \rightarrow (1 - \frac{x}{x})^+} W(q_{DM}) \geq \lim_{q_{DM} \rightarrow (1 - \frac{x}{x})^-} W(q_{DM})
$$

(2.25)

when $p^- < \frac{x}{1+x} < p^+$. To start with, note that the DM’s equilibrium strategy is $(\delta^*_{LL}, \delta^*_{HH}, \delta^*_{LH}, \delta^*_{HL}) = (0, 0, 0, 1)$ when $q_{DM} < 1 - \frac{x}{x}$, and $(\delta^*_{LL}, \delta^*_{HH}, \delta^*_{LH}, \delta^*_{HL}) = (0, 0, \frac{-1-q_{DM} x_i}{q_{DM} x_i}, 1)$ when $q_{DM} \geq 1 - \frac{x}{x}$. Because in the latter case the DM is indifferent between the two actions upon observing $s_{DM} = L$ and $\tilde{m} = H$, the DM’s expected pay-off would remain unchanged if she took action $a = A$ with prob. 1 if and only if $\tilde{m} = H$, in which case the mapping from $s_{DM}$ and $\tilde{m}$ to actions would be exactly the same as
when $q_{DM} < 1 - \frac{c}{x}$. However, the sender’s equilibrium strategy when $q_{DM} < 1 - \frac{c}{x}$ is $(\sigma^*_L, \sigma^*_H) = (1, 0)$, which means that it involves more falsification than the sender’s equilibrium strategy when $q_{DM} \geq 1 - \frac{c}{x}$, which is $(\sigma^*_L, \sigma^*_H) = (\frac{p^-}{(1-p^-)x}, 0)$. Hence, (2.25) holds.

Proof of Proposition 2.6

We first derive the DM’s welfare in the model in which the DM observes a public signal of quality $q_{DM}$. This amounts to replicating the analysis in Sections 2.2.1 and 2.2.2 for $s_{pub} = L$ and $s_{pub} = H$. We consider three mutually exclusive regions of parameter values; as before, in the knife-edge cases, we select the sender-preferred strategy profile whenever there are multiple equilibria. We then compare the results with the DM’s welfare in the model in which the DM observes a private signal of quality $q_{DM}$, already derived in Proof of Proposition 2.5.

Let us denote by $\sigma_{s_{pub}}^*$ the sender’s strategy for given realisations of the public signal, $s_{pub}$, and the sender’s private signal, $s$. Similarly, let us denote by $\delta_{s_{pub}}^*$ the sender’s strategy for given realisations of $s_{pub}$ and $s$.

1. Consider parameter values such that $p^+ < \frac{x}{1+x}$. In the setup with a public signal of quality $q_{DM}$, the equilibrium strategies depend on the realisation of the public signal: if $s_{pub} = L$, then the sender falsifies with prob. $\frac{p^-}{(1-p^-)x}$, and if $s_{pub} = H$, he falsifies with prob. $\sigma^*_L = \frac{p^+}{(1-p^+)x}$. Thus, his equilibrium strategy is $(\sigma^*_L, \sigma^*_H) = (\frac{p^-}{(1-p^-)x}, \frac{p^+}{(1-p^+)x})$. The DM’s equilibrium strategy is $(\delta^*_{LL}, \delta^*_{HL}, \delta^*_{LH}, \delta^*_{HH}) = (0, 0, \frac{c}{x}, \frac{c}{x})$. Since the DM either strictly prefers to reject or is indifferent between accepting and rejecting, her expected payoff would remain unchanged if she rejected regardless of $s_{pub}$ and $\tilde{m}$. It follows that her expected payoff is $W_{pub}(q_{DM}) = \mathbb{E}[v(\sigma^*(q_{DM}), \delta^*(q_{DM}))] = 1 - p$. The algebraic expression for $W_{pub}(q_{DM})$ is thus the same as the one for $W(q_{DM})$ in (2) in the proof of Proposition 2.5 (which applied to the parameter space defined by $q_{DM} < 1 - \frac{c}{x}$ and $p^+ < \frac{x}{1+x}$).

2. Consider parameter values such that $p^- < \frac{x}{1+x} \leq p^+$. The equilibrium strategies are $(\sigma^*_{LL}, \sigma^*_{HL}) = (\frac{p^-}{(1-p^-)x}, 1)$ for the sender, and $(\delta^*_{LL}, \delta^*_{HL}, \delta^*_{LH}, \delta^*_{HH}) = (0, 0, \frac{c}{x}, 1)$ for the DM. The DM’s expected payoff is then $W_{pub}(q_{DM}) = pq_{DM} + (1-p)[(1-q_{DM})(1-x) + q_{DM}] = p[q_{DM}] + (1-p)[(1-x) + q_{DM}x]$. The algebraic expression for $W_{pub}(q_{DM})$ is thus the same as the ones for $W(q_{DM})$ in
When DM’s expected payoff is strictly lower with a public signal than with a private signal

The equilibrium strategies of the DM and the sender yield Table 2.2. To show that the profile whenever there are multiple equilibria.

1. Consider parameter values such that \( p^+ < \frac{x}{1+x} \). The DM’s expected payoff is then \( W_{\text{pub}}(q_{DM}) = p + (1-p)(1-x) \). The algebraic expression for \( W_{\text{pub}}(q_{DM}) \) is thus the same as the one for \( W(q_{DM}) \) in (1) in the proof of Proposition 2.5 (which applied to the parameter space defined by \( p^+ < \frac{x}{1+x} \)).

2. Consider parameter values such that \( p^+ \geq \frac{x}{1+x} \). The DM’s expected payoff is then \( W_{\text{pub}}(q_{DM}) = 1-p \) with a public signal, and \( W(q_{DM}) = p^+ + (1-p)[(1-x) + q_{DM}(1-x)] = p + (1-p) \left( 1 - \frac{p^-}{1-p^+} \right) \) with a private signal. It can be easily shown that the latter is strictly higher than the former (using the property that \( p^- < p \)).

3. Consider parameter values such that \( p^- \geq \frac{x}{1+x} \). The equilibrium strategies are \( (\sigma^*_{LL}, \sigma^*_{HL}) = (1, 1) \) for the sender, and \( (\delta^*_{LL}, \delta^*_{HL}, \delta^*_{HH}) = (0, 0, 1, 1) \) for the DM. The DM’s expected payoff is then \( W_{\text{pub}}(q_{DM}) = p + (1-p)(1-x) \). The algebraic expression for \( W_{\text{pub}}(q_{DM}) \) is thus the same as the one for \( W(q_{DM}) \) in (1) in the proof of Proposition 2.5 (which applied to the parameter space defined by \( p^- \geq \frac{x}{1+x} \)).

The equilibrium strategies of the DM and the sender yield Table 2.2. To show that the DM’s expected payoff is strictly lower with a public signal than with a private signal when \( q_{DM} \geq 1 - \frac{c}{x} \) and \( p^- < \frac{x}{1+x} \), consider two cases:

1. Consider parameter values such that \( q_{DM} \geq 1 - \frac{c}{x} \) and \( p^+ < \frac{x}{1+x} \). The DM’s expected payoff is then \( W_{\text{pub}}(q_{DM}) = 1-p \) with a public signal, and \( W(q_{DM}) = p^+ + (1-p)[(1-x) + q_{DM}(1-x)] = p + (1-p) \left( 1 - \frac{p^-}{1-p^+} \right) \) with a private signal. It can be easily shown that the latter is strictly higher than the former (using the property that \( p^- < p \)).

Proof of Proposition 2.7

It is straightforward to show that it must be that \( \sigma^*_{H} = 0 \) in any equilibrium. It remains to derive \( \sigma^*_{L}, \delta^*_{L}, \) and \( \delta^*_{H} \). To do so, we consider four mutually exclusive regions of parameter values, and show that there is a unique equilibrium strategy profile in each of them. As before, in the knife-edge cases, we select the sender-preferred strategy profile whenever there are multiple equilibria.

1. Consider parameter values such that \( \beta_H(\sigma_L = 0, \sigma_H = 0) < \frac{1}{2} \), which is equivalent to \( p < 1 - q \). For any \( \sigma_L \in [0, 1] \) and \( \sigma_H = 0 \), given that \( \beta_H(\sigma_L = 0, \sigma_H = 0) < \frac{1}{2} \),
the DM’s sequentially rational (henceforth, abbreviated to “s.r.”) strategy satisfies \( \delta_H = 0 \). From \( \beta_H(\sigma_L = 0, \sigma_H = 0) < \frac{1}{2} \), it also follows that \( \beta_L(\sigma_L = 0, \sigma_H = 0) < \frac{1}{2} \), so the DM’s s.r. strategy satisfies \( \delta_L = 0 \). Given that \( c > 0, \delta_L = 0, \) and \( \delta_H = 0 \), the sender’s expected payoff from falsifying upon observing \( s = L \) is \( x(\delta_H - \delta_L) - c > 0 \), so the sender’s s.r. strategy satisfies \( \sigma_L = 0 \). Thus, for \( p < 1 - q \), the unique equilibrium strategies are \( \sigma_L^* = 0, \delta_L^* = 0, \) and \( \delta_H^* = 0 \).

2. Consider parameter values such that \( \beta_H(\sigma_L = 0, \sigma_H = 0) > \frac{1}{2} \) and \( \beta_H(\sigma_L = 1, \sigma_H = 0) < \frac{1}{2} \), which together are equivalent to \( p \in \left( 1 - q, \frac{1-(1-x)q}{1+x} \right) \).

   (a) If we suppose that \( \sigma_L = 0 \), then the DM’s s.r. strategy satisfies \( \delta_H = 1 \), which makes \( \sigma_L = 0 \) not s.r., a contradiction.

   (b) If we suppose that \( \sigma_L = 1 \), then the DM’s s.r. strategy satisfies \( \delta_H = 0 \), which makes \( \sigma_L = 1 \) not s.r., a contradiction.

   (c) Suppose now that \( \sigma_L \in (0,1) \). If \( \beta_H(\sigma_L, \sigma_H = 0) > \frac{1}{2} \), then the DM’s s.r. strategy satisfies \( \delta_H = 1 \), which makes \( \sigma_L \in (0,1) \) not s.r., a contradiction. If \( \beta_H(\sigma_L, \sigma_H = 0) < \frac{1}{2} \), then the DM’s s.r. strategy satisfies \( \delta_H = 0 \), which makes \( \sigma_L \in (0,1) \) not s.r., a contradiction. Therefore, suppose now that \( \beta_H(\sigma_L, \sigma_H = 0) = \frac{1}{2} \), which pins down \( \sigma_L \in (0,1) \) to \( \sigma_L = \frac{p+q-1}{(q-p)x} \). If \( x \delta_H - c > 0 \), then \( \sigma_L \in (0,1) \) is not s.r. (\( \sigma_L = 1 \) is). If \( x \delta_H - c < 0 \), then \( \sigma_L \in (0,1) \) is not s.r. (\( \sigma_L = 0 \) is). If \( x \delta_H - c = 0 \), then the sender is indifferent between falsifying and not falsifying upon observing \( s = L \), so \( \sigma_L = \frac{p+q-1}{(q-p)x} \in (0,1) \) is s.r. This uniquely pins down \( \delta_H = \frac{c}{x} \). Thus, for \( p \in \left( 1 - q, \frac{1-(1-x)q}{1+x} \right) \), the unique equilibrium strategies satisfy \( \sigma_L^* = \frac{p+q-1}{(q-p)x}, \delta_L^* = 0, \) and \( \delta_H^* = \frac{c}{x} \).

3. Consider parameter values such that \( \beta_H(\sigma_L = 1, \sigma_H = 0) > \frac{1}{2} \) and \( \beta_L(\sigma_L = 0, \sigma_H = 0) < \frac{1}{2} \), which together are equivalent to \( p \in \left( \frac{1-(1-x)q}{1+x}, q \right) \). For any \( \sigma_L \in [0,1] \) and \( \sigma_H = 0 \), given that \( \beta_H(\sigma_L = 1, \sigma_H = 0) > \frac{1}{2} \), the DM’s s.r. strategy satisfies \( \delta_H = 1 \). Given that \( \beta_L(\sigma_L = 0, \sigma_H = 0) < \frac{1}{2} \), the DM’s s.r. strategy satisfies \( \delta_L = 0 \). Given that \( c \in (0, x) \), \( \delta_L = 0 \), and \( \delta_H = 1 \), the sender’s expected payoff from falsifying upon observing \( s = L \) is \( x(\delta_H - \delta_L) - c > 0 \), and thus the sender’s s.r. strategy satisfies \( \sigma_L = 1 \). Thus, for \( p \in \left( \frac{1-(1-x)q}{1+x}, q \right) \), the unique equilibrium strategies satisfy \( \sigma_L^* = 1, \delta_L^* = 0, \) and \( \delta_H^* = 1 \).
4. Consider parameter values such that $\beta_L(\sigma_L = 0, \sigma_H = 0) > \frac{1}{2}$, which is equivalent to $p \in (q, 1)$. For any $\sigma_L \in [0, 1]$ and $\sigma_H = 0$, given that $\beta_L(\sigma_L = 0, \sigma_H = 0) > \frac{1}{2}$, the DM’s s.r. strategy satisfies $\delta_L = 1$ and $\delta_H = 1$. Given that $c > 0$, $\delta_L = 1$, and $\delta_H = 1$, the sender’s expected payoff from falsifying upon observing $s = L$ is $x(\delta_H - \delta_L) - c < 0$, so the sender’s unique s.r. action is $\sigma_L = 0$. Thus, for $p \in (q, 1)$, the unique equilibrium strategies satisfy $\sigma_L^* = 0$, $\delta_{LH}^* = 1$, and $\delta_{HH}^* = 1$.

**Proof of Proposition 2.8**

We consider here $p \in (1 - q, q)$, i.e. the values of prior belief such that the sender’s equilibrium strategy in the absence of DM’s private signal satisfies $\sigma_L^* = 0$. Falsification can be fully deterred only if (i) $\tilde{m}$ changes the DM’s decision only for one realisation of $s_{DM} \in \{L, H\}$, and (ii) the sender—upon observing $s$—must conclude that it is unlikely that the DM has observed that particular signal realisation and, given $c$ and $x$, the sender’s message is sufficiently unlikely to change the DM’s decision.

1. First, consider a prior belief such that (a) $\beta(s_{DM} = H, \tilde{m} = H | \sigma_L = 0) > \frac{1}{2}$, (b) $\beta(s_{DM} = L, \tilde{m} = H | \sigma_L = 0) < \frac{1}{2}$, (c) $\beta(s_{DM} = H, \tilde{m} = L) < \frac{1}{2}$, and (d) $\beta(s_{DM} = L, \tilde{m} = L) < \frac{1}{2}$. Note that condition (b) and the assumption $q > q_{DM}$ imply that $p < \frac{1}{2}$. Given that $p < \frac{1}{2}$, $q > q_{DM}$ implies condition (c). Finally, condition (d) is implied by condition (b). Thus, conditions (c) and (d) are redundant. We need to analyse two cases here:

- Consider $\beta(s_{DM} = H, \tilde{m} = H | \sigma_L = 1) > \frac{1}{2}$. The DM’s equilibrium strategy is then $(\delta_{LH}, \delta_{HH}, \delta_{LH}, \delta_{HH}) = (0, 0, 0, 1)$ regardless of whether the sender falsifies or not, and hence the sender’s expected benefit from falsification upon observing $s = L$ is $\pi_{H|x} - c$. Thus, the sender’s equilibrium strategy is $\sigma_L^* = 1$ if and only if $c < \pi_{H|x}$; otherwise, it is $\sigma_L^* = 0$ (in the knife-edge case of $c = \pi_{H|x}$, we assume that the sender does not falsify).

- Consider now $\beta(s_{DM} = H, \tilde{m} = H | \sigma_L = 1) < \frac{1}{2}$.

  - Suppose $\sigma_L = 0$. Given condition (a), the DM’s s.r. strategy then satisfies $\delta_{HH} = 1$, so the sender’s expected benefit from falsification upon observing $s = L$ is $\pi_{H|x} - c$. If $c > \pi_{H|x}$, the sender’s s.r. strategy is $\sigma_L = 0$, and therefore the equilibrium strategies of the DM and the sender satisfy $\sigma_L = 0$ and $\delta_{HH} = 1$. If $c < \pi_{H|x}$, the sender’s s.r.
strategy is $\sigma_L > 0$, a contradiction. In the knife-edge case of $c = \pi_{H|L}x$, we assume that the sender does not falsify, i.e. $\sigma_L = 0$.

- Suppose $\sigma_L = 1$. The DM’s s.r. strategy then satisfies $\delta_{HH} = 0$, in which case $\sigma_L = 1$ is not s.r., a contradiction.

- Suppose $\sigma_L \in (0, 1)$. If $\beta(s_{DM} = H, \tilde{m} = H \mid \sigma_L) > \frac{1}{2}$, then $\sigma_L = 1$ is s.r., a contradiction. If $\beta(s_{DM} = H, \tilde{m} = H \mid \sigma_L) < \frac{1}{2}$, then $\sigma_L = 0$ is s.r., a contradiction. It must then be that $\beta(s_{DM} = H, \tilde{m} = H \mid \sigma_L) = \frac{1}{2}$, which implies that any $\delta_{HH} \in [0, 1]$ is s.r. for the DM. Now, if either $\pi_{H|L}x\delta_{HH} - c > 0$ or $\pi_{H|L}x\delta_{HH} - c < 0$, then $\sigma_L \in (0, 1)$ is not s.r. for the sender. If $\pi_{H|L}x\delta_{HH} - c = 0$, then $\sigma_L \in (0, 1)$ is s.r. Hence, this pins down $\delta_{HH} = \frac{c}{\pi_{H|L}x}$. If $c > \pi_{H|L}x$, then $\delta_{HH} = \frac{c}{\pi_{H|L}x} > 1$, a contradiction. If $c < \pi_{H|L}x$, then $\delta_{HH} = \frac{c}{\pi_{H|L}x} < 1$, and hence the equilibrium strategies satisfy $\delta_{HH} = \frac{c}{\pi_{H|L}x}$ and $\sigma_L^*$ such that $\beta(s_{DM} = H, \tilde{m} = H \mid \sigma_L^*) = \frac{1}{2}$.

The case of $c = \pi_{H|L}x$ yields $\delta_{HH} = \frac{c}{\pi_{H|L}x} = 1$ and has been analysed in the first bullet point.

Finally, we need to show that there exist values of $p$ which satisfy (a) and (b) as well as $p > 1 - q$ (so that the sender would falsify with a positive probability if the DM did not acquire a private signal). Note here that $p > 1 - q$ is equivalent to $\beta(\tilde{m} = H \mid \sigma_L = 0) > \frac{1}{2}$. Furthermore, $\beta(\tilde{m} = H \mid \sigma_L = 0) > \frac{1}{2}$ implies $\beta(s_{DM} = H, \tilde{m} = H \mid \sigma_L = 0) > \frac{1}{2}$, and $\beta(s_{DM} = L, \tilde{m} = H \mid \sigma_L = 0) > \frac{1}{2}$ implies $\beta(\tilde{m} = H \mid \sigma_L = 0) > \frac{1}{2}$. Hence, there exists a range of $p$ such that (a) and (b) are satisfied and $\sigma_L^* > 0$ when $q_{DM} = \frac{1}{2}$.

2. Second, consider a prior belief such that (a) $\beta(s_{DM} = H, \tilde{m} = H \mid \sigma_L) > \frac{1}{2}$, (b) $\beta(s_{DM} = L, \tilde{m} = H \mid \sigma_L = 0) > \frac{1}{2}$, (c) $\beta(s_{DM} = H, \tilde{m} = L) > \frac{1}{2}$, and (d) $\beta(s_{DM} = L, \tilde{m} = L) < \frac{1}{2}$. Note that condition (c) implies that condition (a) holds for any $\sigma_L$. Furthermore, condition (c) and the assumption $q > q_{DM}$ imply that $p > \frac{1}{2}$. Finally, given that $p > \frac{1}{2}$, $q > q_{DM}$ implies condition (b). Thus, conditions (a) and (b) are redundant here. We need to analyse two cases here:

- Consider $\beta(s_{DM} = L, \tilde{m} = H \mid \sigma_L = 1) > \frac{1}{2}$. The DM’s equilibrium strategy is then $(\delta_{LL}^*, \delta_{HL}^*, \delta_{LH}^*, \delta_{HH}^*) = (0, 1, 1, 1)$ regardless of whether the sender falsifies or not, and hence the sender’s expected benefit from falsification upon observing $s = L$ is $\pi_{L|L}x - c$. Thus, the sender’s equilibrium strategy
is $\sigma^*_L = 1$ if and only if $c \leq \pi_{L|x}$; otherwise, it is $\sigma^*_L = 0$ (in the knife-edge case of $c = \pi_{L|x}$, we assume that the sender does not falsify).

• Consider $\beta(s_{DM} = L, \tilde{m} = H \mid \sigma_L = 1) < \frac{1}{2}$. The analysis proceeds analogously to that in the second bullet point of (1).

Finally, we need to show that there exist values of $p$ which satisfy (c) and (d) as well as $p < q$ (so that the sender would falsify with a positive probability if the DM did not acquire a private signal). Note here that $p < q$ is equivalent to $\beta(\tilde{m} = L) < \frac{1}{2}$. Furthermore, $\beta(\tilde{m} = L) < \frac{1}{2}$ implies $\beta(s_{DM} = L, \tilde{m} = L) < \frac{1}{2}$, and $\beta(s_{DM} = H, \tilde{m} = L) < \frac{1}{2}$ implies $\beta(\tilde{m} = L) < \frac{1}{2}$. Hence, there exists a range of $p$ such that (c) or (d) are satisfied and $\sigma^*_L > 0$ when $q_{DM} = \frac{1}{2}$.

Proof of Proposition 2.9

The proof follows from the main text.

Proof of Lemma 2.1

1. Suppose that the DM sets $\tau = \tau'$ where $\tau' \leq \frac{c}{x}$, and solve for the equilibrium strategies of the DM and the sender. Consider any $\delta_H \in [0, 1]$. When $\tau' < \frac{c}{x}$, the sender’s expected payoff from falsifying upon observing $s = L$ is $\tau'x\delta_H - c < 0$, so the sender’s s.r. strategy satisfies $\sigma_L = 0$. In the knife-edge case of $\tau' = \frac{c}{x}$, we assume that whenever the sender is indifferent between falsifying and not falsifying, he chooses not to falsify. Given that the sender’s equilibrium strategy satisfies $\sigma^*_L = 0$, the DM’s welfare is $(1 - \tau')E[v(\sigma^*, \delta^* = (0, 0))] + \tau'E[v(\sigma^* = (0, 0), \delta^*)]$ for $p \leq \frac{1}{2}$ and $(1 - \tau')E[v(\sigma^*, \delta^* = (1, 1))] + \tau'E[v(\sigma^* = (0, 0), \delta^*)]$ for $p > \frac{1}{2}$.

The DM’s welfare is increasing in $\tau'$. When $\tau' = \frac{c}{x}$, the DM’s welfare is

$$
\left(1 - \frac{c}{x}\right)E[v(\sigma^*, \delta^* = (0, 0))] + \frac{c}{x}E[v(\sigma^* = (0, 0), \delta^*)] \tag{2.26}
$$

for $p \leq \frac{1}{2}$, and

$$
\left(1 - \frac{c}{x}\right)E[v(\sigma^*, \delta^* = (1, 1))] + \frac{c}{x}E[v(\sigma^* = (0, 0), \delta^*)] \tag{2.27}
$$

for $p > \frac{1}{2}$.
2. Suppose that the DM sets \( \tau = \tau' \) where \( \tau' > \frac{c}{x} \), and solve for the equilibrium strategies of the DM and the sender.

(a) Consider parameter values such that \( \beta_H(\sigma_L = 1) \geq \frac{1}{2} \), which is equivalent to \( p \geq \frac{x}{1+x} \). Then, for any \( \sigma_L \in [0, 1] \), the DM’s s.r (“sequentially rational”) strategy satisfies \( \delta_H = 1 \). The sender’s expected payoff from falsifying upon observing \( s = L \) is \( \tau' x - c > 0 \), so the sender’s s.r. strategy satisfies \( \sigma_L = 1 \). Thus, for \( p \geq \frac{x}{1+x} \), the sender’s equilibrium strategy is \( \sigma^*_L = 1 \). The DM’s welfare for \( p \leq \frac{1}{2} \) is then \( (1 - \tau') \mathbb{E}[v(\sigma^*, \delta^* = (0, 0))] + \tau' \mathbb{E}[v(\sigma^* = (1, 0), \delta^* = (0, 1))] \), and for \( p > \frac{1}{2} \) it is \( (1 - \tau') \mathbb{E}[v(\sigma^*, \delta^* = (1, 1))] + \tau' \mathbb{E}[v(\sigma^* = (1, 0), \delta^* = (0, 1))] \). The DM’s welfare is increasing in \( \tau' \). When \( \tau' = 1 \), there is a unique equilibrium and the DM’s welfare is \( \mathbb{E}[v(\sigma^* = (1, 0), \delta^* = (0, 1))] = 1 - (1 - p) x \) for all \( p \in (\frac{x}{1+x}, 1) \).

(b) Consider parameter values such that \( \beta_H(\sigma_L = 1) < \frac{1}{2} \), which is equivalent to \( p < \frac{x}{1+x} \).

i. If \( \sigma_L = 1 \), then the DM’s s.r strategy satisfies \( \delta_H = 0 \), so the sender’s expected payoff from falsifying upon observing \( s = L \) is \( \tau' x \cdot 0 - c < 0 \), which makes \( \sigma_L = 1 \) not s.r., a contradiction.

ii. If \( \sigma_L = 0 \), then the DM’s s.r strategy satisfies \( \delta_H = 1 \), so the sender’s expected payoff from falsifying upon observing \( s = L \) is \( \tau' x \cdot 1 - c > 0 \), which makes \( \sigma_L = 0 \) not s.r., a contradiction.

iii. Suppose now that \( \sigma_L \in (0, 1) \). If either \( \beta_H(\sigma_L) > \frac{1}{2} \) and \( \beta_H(\sigma_L) < \frac{1}{2} \), then \( \sigma_L \in (0, 1) \) is not s.r. If \( \beta_H(\sigma_L) = \frac{1}{2} \), then any \( \delta_H \in [0, 1] \) is s.r for the DM. If either \( \tau' x \delta_H - c > 0 \) or \( \tau' x \delta_H - c < 0 \), then \( \sigma_L \in (0, 1) \) is not s.r. If \( \tau' x \delta_H - c = 0 \), then \( \sigma_L \in (0, 1) \) is s.r. for the sender. This uniquely pins down \( \delta_H = \frac{c}{\tau' x} \). Also, \( \beta_H(\sigma_L) = \frac{1}{2} \) uniquely pins down \( \sigma_L = \frac{p}{(1-p) x} \). Thus, for \( p < \frac{x}{1+x} \), the equilibrium strategies satisfy \( \sigma^*_L = \frac{p}{(1-p) x} \) and \( \delta^*_H = \frac{c}{\tau' x} \). For \( p \leq \frac{1}{2} \), the DM’s welfare is \( (1 - \tau') \mathbb{E}[v(\sigma^*, \delta^* = (0, 0))] + \tau' \mathbb{E}[v(\sigma^*, \delta^* = (0, 0))] \) which is smaller than \( \boxed{2.26} \). For \( p > \frac{1}{2} \), the DM’s welfare is \( (1 - \tau') \mathbb{E}[v(\sigma^*, \delta^* = (1, 1))] + \tau' \mathbb{E}[v(\sigma^*, \delta^* = (0, 0))] \), which is smaller than \( \mathbb{E}[v(\sigma^*, \delta^* = (1, 1))] \), which in turn is smaller than \( \boxed{2.27} \).
Proof of Proposition 2.10

Using the results from Table 2.3:

1. Consider $0 < p < \frac{x}{1 + x}$. Then $E(v_{\tau = c/\hat{m}}(\sigma^*, \delta^*)) = \frac{c}{x} + (1 - p) \left(1 - \frac{c}{x}\right)$ and $E(v_{\tau = 1}(\sigma^*, \delta^*)) = 1 - p$. It can be easily verified that $\frac{c}{x} + (1 - p) \left(1 - \frac{c}{x}\right) > 1 - p$ for all parameter values.

2. Consider $\frac{x}{1 + x} < p < \frac{1}{2}$. Then $E(v_{\tau = c/\hat{m}}(\sigma^*, \delta^*)) = \frac{c}{x} + (1 - p) \left(1 - \frac{c}{x}\right)$ and $E(v_{\tau = 1}(\sigma^*, \delta^*)) = p + (1 - p) (1 - x)$. The condition $\frac{c}{x} + (1 - p) \left(1 - \frac{c}{x}\right) \geq p + (1 - p) (1 - x)$ can be rearranged to $c \geq x \left(1 - \frac{1-p}{p} x\right)$, where the right-hand side of the inequality is increasing in $p$.

3. Consider $\frac{1}{2} < p$. Then $E(v_{\tau = c/\hat{m}}(\sigma^*, \delta^*)) = \frac{c}{x} + p \left(1 - \frac{c}{x}\right)$ and $E(v_{\tau = 1}(\sigma^*, \delta^*)) = p + (1 - p) (1 - x)$. The condition $\frac{c}{x} + p \left(1 - \frac{c}{x}\right) \geq p + (1 - p) (1 - x)$ can be rearranged to $c \geq x (1 - x)$.

Note that $x \left(1 - \frac{1-p}{p} x\right) \leq x (1 - x)$ if and only if $p \leq \frac{1}{2}$, and that $x \left(1 - \frac{1-p}{p} x\right) \leq 0$ if and only if $p \leq \frac{x}{1 + x}$. This yields the condition in (2.12) and can be used to construct all the cases in Table 2.4.

Proof of Proposition 2.11

Let $\beta_{\hat{m}}(\sigma_L)$ denote the DM’s belief upon observing $\hat{m}$ when the sender’s strategy is $\sigma_L$. Since the sender can only falsify $s = L$ into $\hat{m} = M$, $\beta_L(\sigma_L)$ and $\beta_H(\sigma_L)$ do not depend on $\sigma_L$, so we can simply write $\beta_L$ and $\beta_H$. We consider five mutually exclusive and exhaustive regions of parameter values, and show that there is a unique equilibrium strategy profile in each of them.

1. Consider parameter values such that $\beta_H < \frac{1}{2}$, which is equivalent to $\frac{p}{1-p} \in \left(0, \frac{(1-q)^2}{q^2}\right]$. Then, for any $\sigma_L \in [0, 1]$, the DM’s s.r. strategy satisfies $(\delta_L, \delta_M, \delta_H) = (0, 0, 0)$. Given that $c < x$, the sender’s s.r. strategy satisfies $\sigma_L = 0$. Thus, for $\frac{p}{1-p} \in \left(0, \frac{(1-q)^2}{q^2}\right]$, the unique equilibrium strategies are $\sigma_L^* = 0$ and $(\delta_L^*, \delta_M^*, \delta_H^*) = (0, 0, 0)$.

2. Consider parameter values such that $\beta_H > \frac{1}{2}$ and $\beta_M(\sigma_L = 0) < \frac{1}{2}$, which together are equivalent to $\frac{p}{1-p} \in \left(\frac{(1-q)^2}{q^2}, 1\right]$. Then, for any $\sigma_L \in [0, 1]$, the DM’s s.r. strategy satisfies $(\delta_L, \delta_M, \delta_H) = (0, 0, 1)$. Given that $c < x$, the sender’s s.r.
strategy satisfies $\sigma_L = 0$. Thus, for $\frac{p}{1-p} \in \left(\frac{(1-q)^2}{q^2}, 1\right]$, the unique equilibrium strategies are $\sigma_L^* = 0$ and $(\delta_L^*, \delta_M^*, \delta_H^*) = (0, 0, 1)$.

3. Consider parameter values such that $\beta_M(\sigma_L = 0) > \frac{1}{2}$ and $\beta_M(\sigma_L = 1) < \frac{1}{2}$, which together are equivalent to $\frac{p}{1-p} \in \left(1, \frac{q^2 + 2q(1-q)}{(1-q)^2 + 2q(1-q)}\right]$. If we suppose that $\sigma_L = 1$, the DM’s s.r. strategy satisfies $(\delta_L, \delta_M, \delta_H) = (0, 0, 1)$, which makes $\sigma_L = 1$ not s.r., a contradiction. If we suppose that $\sigma_L = 0$, the DM’s s.r. strategy satisfies $(\delta_L, \delta_M, \delta_H) = (0, 1, 1)$, which makes $\sigma_L = 0$ not s.r., a contradiction. Suppose now that $\sigma_L \in (0, 1)$. If $\beta_M(\sigma_L) > \frac{1}{2}$, then the DM’s s.r. strategy satisfies $(\delta_L, \delta_M, \delta_H) = (0, 1, 1)$, which makes $\sigma_L \in (0, 1)$ not s.r., a contradiction. If $\beta_M(\sigma_L) < \frac{1}{2}$, then the DM’s s.r. strategy satisfies $(\delta_L, \delta_M, \delta_H) = (0, 0, 1)$, which makes $\sigma_L \in (0, 1)$ not s.r., a contradiction. Suppose now that $\beta_M(\sigma_L) = \frac{1}{2}$, which pins down $\sigma_L = 2\frac{q(1-q)(2p-1)}{x(1-p)^2 - (1-q)^2}$. If either $x(\delta_M - \delta_L) - c > 0$ or $x(\delta_M - \delta_L) - c < 0$, where $\delta_L = 0$, then $\sigma_L \in (0, 1)$ is not s.r. If $x(\delta_M - \delta_L) - c = 0$, then the sender is indifferent between falsifying and not falsifying upon observing $s = L$, so $\sigma_L \in (0, 1)$ is s.r. This uniquely pins down $\delta = \frac{c}{x}$. Thus, for $\frac{p}{1-p} \in (1, \frac{q^2 + 2q(1-q)}{(1-q)^2 + 2q(1-q)}]$ , the unique equilibrium strategies are $\sigma_L^* = 2\frac{q(1-q)(2p-1)}{x(1-p)^2 - (1-q)^2}$ and $(\delta_L^*, \delta_M^*, \delta_H^*) = \left(0, \frac{c}{x}, 1\right)$.

4. Consider parameter values such that $\beta_M(\sigma_L = 1) > \frac{1}{2}$ and $\beta_L < \frac{1}{2}$, which together are equivalent to $\frac{p}{1-p} \in \left(1, \frac{q^2 + 2q(1-q)}{(1-q)^2 + 2q(1-q)}\right]$ . Then, for any $\sigma_L \in [0, 1]$ , the DM’s s.r. strategy satisfies $(\delta_L, \delta_M, \delta_H) = (0, 1, 1)$. Given that $c < x$, the sender’s s.r. strategy satisfies $\sigma_L = 1$. Thus, for $\frac{p}{1-p} \in \left(1, \frac{q^2 + 2q(1-q)}{(1-q)^2 + 2q(1-q)}\right]$ , the unique equilibrium strategies are $\sigma_L^* = 1$ and $(\delta_L^*, \delta_M^*, \delta_H^*) = (0, 1, 1)$.

5. Consider parameter values such that $\beta_L > \frac{1}{2}$, which is equivalent to $\frac{p}{1-p} \in \left(\frac{q^2}{(1-q)^2}, +\infty\right]$. Then, for any $\sigma_L \in [0, 1]$ , the DM’s s.r. strategy satisfies $(\delta_L, \delta_M, \delta_H) = (1, 1, 1)$. Given that $c < x$, the sender’s s.r. strategy satisfies $\sigma_L = 0$. Thus, for $\frac{p}{1-p} \in \left(\frac{q^2}{(1-q)^2}, +\infty\right]$, the unique equilibrium strategies are $\sigma_L^* = 1$ and $(\delta_L^*, \delta_M^*, \delta_H^*) = (1, 1, 1)$.

Proof of Corollary 2.2

Part (i). Note that $\frac{p}{1-p}$ is increasing in $p$, so we can simply examine how $\sigma_L^*$ changes within the following five intervals for $\frac{p}{1-p}$: (0, $\frac{(1-q)^2}{q^2}$], ($\frac{(1-q)^2}{q^2}$, 1], (1, $\frac{q^2 + 2q(1-q)}{(1-q)^2 + 2q(1-q)}$], $(\frac{q^2 + 2q(1-q)}{(1-q)^2 + 2q(1-q)}$, $\frac{q^2}{(1-q)^2}$], and $(\frac{q^2}{(1-q)^2}, +\infty]$ , and on their boundaries. For $\frac{p}{1-p} \in (0, \frac{(1-q)^2}{q^2}]$
and \( \frac{p}{1-p} \in \left( \frac{(1-q)^2}{q^2}, 1 \right) \), we have \( \sigma_L^* = 0 \). For \( \frac{p}{1-p} \in \left( 1, \frac{q^2x+2q(1-q)}{(1-q)^2x+2q(1-q)} \right) \), we have \( \sigma_L^* = \frac{2q(1-q)(2p-1)}{x(q^2(1-p)-(1-q)^2p)} \in [0, 1] \) and \( \frac{\partial \sigma_L^*}{\partial p} = \frac{2(q^2(1-q)-(1-q)^2)}{x(q^2(1-p)-(1-q)^2p)} \) which is strictly positive for \( q \in \left( \frac{1}{2}, 1 \right) \). For \( \frac{p}{1-p} \in \left( q^2x+2q(1-q), \frac{q^2}{(1-q)^2x+2q(1-q)} \right) \), we have \( \sigma_L^* = 1 \). For \( \frac{p}{1-p} \in \left( \frac{q^2}{(1-q)^2}, +\infty \right) \), we have \( \sigma_L^* = 0 \).

**Part (ii).** For \( p \in \left( 0, \frac{1}{2} \right) \), we have \( \sigma_L^* = 0 \). For \( p \in \left( \frac{1}{2}, 1 \right) \), we have three regions to consider. Note that \( \frac{q^2}{(1-q)^2} > \frac{q^2x+2q(1-q)}{(1-q)^2x+2q(1-q)} \) for \( q \in \left( \frac{1}{2}, 1 \right) \). First, for \( q \) such that \( \frac{q^2}{(1-q)^2} < \frac{p}{1-p} \), we have \( \sigma_L^* = 0 \). Second, for \( q \) such that \( \frac{q^2}{(1-q)^2} > \frac{p}{1-p} \) and \( \frac{q^2x+2q(1-q)}{(1-q)^2x+2q(1-q)} < \frac{p}{1-p} \), we have \( \sigma_L^* = \frac{2q(1-q)(2p-1)}{x(q^2(1-p)-(1-q)^2p)} \in [0, 1] \) and \( \frac{\partial \sigma_L^*}{\partial q} = \frac{2(1-2q)(q^2(1-p)-(1-q)^2p)-(2q(1-p)+2(1-q)p)(1-p)}{(q^2(1-p)-(1-q)^2p)^2} < 0 \) since \( \frac{q^2}{(1-q)^2} > \frac{p}{1-p} \) and \( q \in \left( \frac{1}{2}, 1 \right) \).

**Proof of Proposition 2.12**

The proof follows from the main text.
Chapter 3

Policy Making with Interest Groups
Competing for Access

3.1 Introduction

Policy makers often do not have enough expertise to be able to choose policies on their own, so they seek advice from better-informed but self-interested experts (e.g., interest groups, consultants, advisors). However, they usually cannot consult as many experts as they would like, for example, because of time constraints. This means that experts have to compete between themselves for the opportunity to talk to policy makers. Empirical studies show that it is very common that interest groups compete for access to policy makers. Herndon (1982), based on interviews with representatives of business interest groups, finds that they overwhelmingly name “access” as the main goal of their contributions. Langbein (1986) provides quantitative evidence supporting the claim that higher contributions from an interest group have a positive influence on the time spent by the politician with its representatives. Hall and Wayman (1990), using data from committees in the U.S. House of Representatives, find that contributions lead committee members to spend more time and effort on policy issues which are important to the contributors.

In this paper, I study a model of competition between experts for access to a policy maker. The main objective of the paper is to analyse the benefits which competition for access brings to the policy maker. I also analyse whether there are circumstances under which a policy maker is better off hiring an expert in advance than having the experts compete for access to her. In the real world, on some occasions, policy makers
leave it to interest groups to compete for access to them, but on other occasions they hire advisors or consultants, whom they choose themselves. This paper aims to shed light on why a policy maker may choose the former or the latter procedure.

In the model, there is a policy maker and two experts. The two experts are biased in opposite directions regarding the policy. For example, we can imagine that there is a policy maker who needs to decide on the environmental policy and that there are two interest groups with opposing goals: one represents the industry and supports a lenient policy, while the other represents environmentalists and supports a strict policy. Each expert privately observes the strength of evidence in favour of his preferred direction of policy. The policy maker would like to choose a policy that reflects the relative strengths of evidence of the two experts. After observing their own evidence, the experts compete for access to the policy maker by exerting effort, which is costly to them. The expert who exerts more effort wins access and gets an opportunity to send a message to the policy maker. The policy maker then chooses a policy.

I first consider a benchmark model of perfect communication. The benchmark model assumes that the communication between an expert and the policy maker is always successful, i.e. the policy maker always understands the message sent by an expert. The analysis of this benchmark model reveals that, in equilibrium, the expert who has stronger evidence has an incentive to exert more effort and thus wins access to the policy maker. In equilibrium, he truthfully reveals the strength of evidence in his favour. Although the policy maker does not receive any message from the losing expert, she can make an inference about the strength of his evidence: she realises that his evidence must have been weaker than the winning expert’s evidence. This illustrates an important benefit of the policy maker from the competition for access to her: it allows her to obtain some information about the losing expert’s private information.

In the main model, I assume that the communication between an expert and the policy maker can be unsuccessful, in which case the policy expert receives no information and cannot make any inference about the experts’ strengths of evidence. The experts are allowed to make an investment in their communication skills, which increases the probability that they communicate successfully with the policy maker. Thus, the probability of successful communication is endogenous.

The analysis of the main model shows that, again, the expert with stronger evidence wins access to the policy maker. The experts invest a positive amount in their communication skills and the investment is increasing in how much value they attach to the policy. I compare this model with a scenario in which the policy maker hires a
randomly chosen expert in advance. It turns out that, in that scenario, the hired expert has no incentive to invest in his communication skills. Thus, this comparison illustrates another benefit to the policy maker from the competition for access to her: it provides an incentive for the experts to invest in their communication skills. The reason for this contrast is that a message from an expert who is hired in advance is discounted more severely by the policy maker than a message from the winner of competition for access.

I then consider an extension of the main model, in which I allow the policy maker to offer the experts a monetary reward for successful communication. In the real world, experts such as consultants and advisers are usually paid for their advice regardless of its content as long as it is sufficiently well researched and prepared. In the model, it turns out that such a reward is ineffective when the experts need to compete for access to the policy maker. It is ineffective because an expert’s monetary benefit from investing in his communication skills is completely offset by the consequent increased effort in the competition for access. In contrast, a reward is effective when the policy maker hires an expert in advance. Therefore, hiring an expert in advance becomes attractive to the policy maker as it allows her to use a reward to incentivise them to invest more in their communication skills.

Thus, overall, both approaches, i.e. competition for access between experts and hiring an expert in advance, have their advantages. On the one hand, competition for access reveals some information about the losing expert’s evidence and also provides an incentive for the experts to invest in their communication skills. On the other hand, hiring an expert allows the policy maker to incentivise the expert to invest more by offering him a monetary reward.

I show that if the experts attach a sufficiently high value to the implemented policy, then the policy maker is better off having them compete for access to her than hiring an expert in advance. Otherwise, the policy maker is better off taking the latter approach. The implication of this analysis for policy makers is that their choice of procedure of collecting information should depend on the characteristics of the issue at hand, in particular, on how much those who have policy-relevant information are interested in the policy implemented by the policy maker.

The remainder of the paper is organised as follows. The rest of this section reviews the related literature. Section 3.2 analyses the benchmark model of perfect communication. Section 3.3 studies the main model of imperfect communication. Section 3.4 extends the main model by allowing the policy maker to offer a reward for successful communication. Section 3.5 discusses an alternative assumption about the observability
of the experts’ efforts in the competition, and its implications for the results. Section 3.6 concludes.

**Related Literature**

This paper is related to the literature on strategic communication. This literature is very rich but, in essence, communication models fall into two classes: cheap talk and verifiable disclosure. In cheap talk models, pioneered by Crawford and Sobel (1982), experts can costlessly send messages which contain information that is not verifiable by receivers. Verifiable disclosure models, first studied by Milgrom (1981) and Grossman (1981), lie on the other side of the spectrum. In these models, it is assumed that the content of messages is verifiable, so experts cannot send false information, but they have the option of suppressing information. In my paper, communication between experts and the policy maker is by verifiable disclosure. This modelling approach is motivated by the fact that potential verification of evidence by policy makers is an important aspect of lobbying and, furthermore, policy makers may simply require hard evidence for their policy decisions. Within the literature on strategic communication, my paper is also related to the research on eliciting information from self-interested parties with opposing incentives. Examples of papers on this topic (in a wide variety of model settings) include Shin (1994), Dewatripont and Tirole (1999), Krishna and Morgan (2001), and Bhattacharya and Mukherjee (2013).

The main contribution of this paper is to the literature on the theory of interest groups and lobbying. The two closest papers to mine are Cotton (2009) and Cotton (2016), who study competition for access to the policy maker. Cotton (2009) analyses a stylised model which is similar to the benchmark in my paper: there is an imperfectly informed policy maker and two interest groups with extreme preferences, who compete by making contributions in an all-pay auction. Cotton (2016) develops a model with a similar structure but makes more general assumptions about the number of interest groups, the distribution of their private information, and the players’ payoff functions. In both of these papers, unlike in my model, the policy maker observes the contributions and directly benefits from them. In addition, unlike in my model, communication between interest groups and the policy maker is always successful.

The research questions considered by Cotton (2009) and Cotton (2016) are different to mine. Cotton (2009) concentrates on the question of whether the policy maker should sell access (i.e. allow the highest bidder to present his evidence) or sell policy
favours (i.e. choose the policy preferred by the highest bidder) in exchange for the contributions, whereas the focus of Cotton (2016) is on showing that selling access can make the policy maker fully informed and thus lead to the first best policy. In contrast, I consider whether the policy maker should have the interest groups compete for access or hire an expert in advance. Due to unobservable contributions, the informational role of contributions is minimised in my paper; however, the policy maker can still learn something about the losing interest group’s private information. Furthermore, I analyse the role of imperfect communication between the interest groups and the policy maker, which is absent in both Cotton (2009) and Cotton (2016).

Austen-Smith (1998) and Cotton (2012) study models with access fees, i.e. where the policy maker sets a fee which the interest groups need to pay in exchange for access. In Austen-Smith (1998), paying the fee is not a sufficient condition since the policy maker chooses only one interest group from those who have done so, whereas in Cotton (2012), all interest groups who pay the fee can reveal their information to the policy maker. The general focus of these papers is on how the policy maker can benefit by restricting access through access fees—and hence without direct competition between interest groups—while my paper investigates how she can benefit from direct competition.

Lohmann (1995) and Austen-Smith (1995) analyse models of access where information is unverifiable. Lohmann (1995) studies a model with multiple interest groups which choose whether to buy access and anyone who buys access can send an unverifiable message to a policy maker. Austen-Smith (1995) develops a model where an interest group chooses a contribution, then nature decides whether the issue actually arises on the legislative agenda, after which the policy maker chooses whether to grant access. In both papers, the information about contributions allows the policy maker to form an inference about the credibility of the messages which she receives. This contrasts with my paper, in which information is verifiable and thus its credibility is not an issue for the policy maker.

There is also a well-developed strand of literature which studies competition between interest groups where the prize is not access but a policy favour, i.e. the policy maker simply implements the winner’s policy. It includes papers in which interest groups compete in an all-pay auction (Hillman and Riley 1989; Baye, Kovenock and de Vries 1993; Che and Gale 1998) and in a menu auction (Grossman and Helpman 1994; Grossman and Helpman 1996). Finally, Levy and Razin (2013) study a model of competition where the prize is yet something else: the interest group which wins an
all-pay auction has its bill put on the legislative agenda, which means that it is voted versus the status quo by the policy maker. The focus of their paper is on the dynamics of policy changes, in particular on the convergence of policy and the speed of changes.

3.2 Perfect Communication Benchmark

In this section, I analyse a benchmark model of competition between experts for access to a policy maker. In this benchmark, I assume that the communication between an expert and the policy maker is perfect in the sense that the policy maker receives and understands the message sent by an expert with probability 1.

3.2.1 Model Setup

There are three players: two experts (L and R) and a policy maker (PM).

First, each expert $i \in \{L, R\}$ privately observes the strength of evidence in his favour, which is denoted by $\alpha_i$. Each $\alpha_i$ is a realisation of a random variable uniformly distributed on $[0, 1]$. The experts’ $\alpha_i$’s are independent. After observing their $\alpha_i$’s, the two experts compete for access to the policy maker. They simultaneously and independently decide how much effort to exert in the competition. The effort exerted by expert $i$ (e.g., the amount of money spent) is denoted by $e_i \in \mathbb{R}_+$. The expert who exerts more effort wins the competition and gains access to the policy maker. The winner’s identity is denoted by $w \in \{L, R\}$.

The winner of the competition sends a private, costless, and verifiable message $m_w \in [0, 1]$ to the policy maker. Since the message is verifiable, the expert cannot exaggerate the evidence that he possesses, but he can understate it. Upon receiving the winner’s message and his identity, the policy maker forms a Bayesian inference about the experts’ $\alpha_i$’s and chooses a policy $p \in [-1, 1]$. The policy maker does not observe anything else; in particular, she does not observe how much effort both experts have exerted in the competition.\footnote{In Cotton (2009), the policy maker does observe the experts’ efforts, which are referred to as “contributions”. In Section 3.5 I analyse the implications of observability of efforts for the results of this paper.} Thus, the signalling role of effort is minimised in this setup: the experts cannot use their efforts as a signal of their private information and the only way an expert can signal anything is by winning or losing the competition for access.
The payoff functions of the players are

\[ \begin{align*}
\Pi_L & = -vp - e_L, \\
\Pi_R & = vp - e_R, \\
\Pi_{PM} & = \alpha_R p + \alpha_L (-p) - \frac{1}{2}p^2,
\end{align*} \]

(3.1) \hspace{1cm} (3.2) \hspace{1cm} (3.3)

where \( \Pi_L, \Pi_R, \) and \( \Pi_{PM} \) denote the payoffs to experts \( L \) and \( R \) and to the policy maker.

The experts are biased as far as the policy maker’s policy is concerned. Their biases are extreme and in opposite directions: expert \( L \) prefers the policy to be as low as possible whereas expert \( R \) would like it to be as high as possible. Their payoffs are also decreasing in the effort exerted in the competition. Parameter \( v > 0 \) tells us how much the experts value the implemented policy relative to their effort. If \( v \) is low, then the experts are relatively neutral regarding the policy, whereas if it is high, then they are less neutral because they benefit a lot if the policy is swayed in their preferred direction.

The policy maker’s payoff function is equivalent\(^2\) to \( \Pi'_{PM} = - (\alpha_R - \alpha_L - p)^2 \). The function is maximised at \( p = \alpha_R - \alpha_L \), which means that the policy maker would ideally want to implement policy \( \alpha_R - \alpha_L \). Thus, the policy maker wants the policy to reflect the relative strengths of evidence in favour of expert \( R \) and \( L \).

Each expert has a competition strategy and a communication strategy. An agent’s competition strategy describes what effort, \( e_i \in \mathbb{R}_+ \), he exerts in the competition upon observing the strength of his evidence, \( \alpha_i \in [0, 1] \). Formally, a competition strategy for expert \( i \) is a map \( \sigma^e_i : [0, 1] \rightarrow \mathbb{R}_+ \). An agent’s communication strategy describes what message, \( m_i \in [0, 1] \), he sends to the policy maker upon observing \( \alpha_i \in [0, 1] \) in case he wins the competition for access. A communication strategy for expert \( i \) is a map \( \sigma^m_i : [0, 1] \rightarrow [0, 1] \).

The policy maker’s strategy describes the policy that she chooses upon observing the winner’s identity and message, \( (w, m_w) \). Formally, the policy maker’s strategy is a map \( \sigma^p : \{L, R\} \times [0, 1] \rightarrow [-1, 1] \).

\(^2\)The payoff functions in this benchmark model are similar to those in Cotton (2009), with one important difference: here, the policy maker does not directly benefit from the efforts of the experts. \(^3\)This is also pointed out by Cotton (2009). \(^4\)The term \( \alpha_R - \alpha_L \) could be interpreted as the state of the world, \( \theta \), and the policy maker would then want her policy to be as close as possible to the true state of the world. The distribution of \( \alpha_R - \alpha_L \) is triangular, with the highest density at \( \theta = 0 \) and lowest at the ends of the interval, \( \theta = -1 \) and \( \theta = 1 \). This interpretation would then imply that each expert does not have full information about the true state of the world, but only partial.
3.2.2 Equilibrium

The solution concept is the perfect Bayesian equilibrium, which is defined in the usual way. A quintuple \((\sigma^e_L, \sigma^e_R, \sigma^m_L, \sigma^m_R, \sigma^p)\) is a perfect Bayesian equilibrium if (i) \(\sigma^e_i\) maximises the expected value of \(\Pi_i\) at the competition stage given the strategies \(\sigma^e_j, \sigma^m_i, \sigma^m_j,\) and \(\sigma^p,\) for \(i \neq j\) and \(i, j \in \{L, R\},\) (ii) \(\sigma^m_i\) maximises the expected value of \(\Pi^m_i\) at the communication stage given the strategy \(\sigma^p,\) for \(i \in \{L, R\},\) (iii) \(\sigma^p\) maximises the expected value of \(\Pi_{PM}\) given the strategies \(\sigma^e_i, \sigma^e_j, \sigma^m_i,\) and \(\sigma^m_j,\) where \(i \neq j\) and \(i, j \in \{L, R\},\) and (iv) the players’ beliefs are derived using Bayes’ rule whenever it is possible. The unique equilibrium is described in the following proposition.\(^5\)

**Proposition 3.1.** The game has a unique perfect Bayesian equilibrium. The experts’ and the policy maker’s equilibrium strategies are

\[
\begin{align*}
\sigma^e_i (\alpha_i) & = \frac{v}{2} \alpha_i^2, \\
\sigma^m_i (\alpha_i) & = \alpha_i, \\
\sigma^p (w, m_w) & = \begin{cases} 
-\frac{m_w}{2} & \text{if } w = L \\
\frac{m_w}{2} & \text{if } w = R
\end{cases}
\end{align*}
\]

where \(i \in \{L, R\}.\)

The equilibrium competition strategy \(\sigma^e_i (\alpha_i)\) of each expert \(i\) is increasing in the strength of his evidence, \(\alpha_i.\) This means that the expert with stronger evidence in his favour exerts a higher effort and wins the competition for access to the policy maker.

At the communication stage, no expert is willing to understate his message by reporting \(m_i < \alpha_i,\) because then he would be able to profitably deviate by sending a higher report. Thus, upon winning access to the policy maker, the winning expert sends a message \(m_i = \alpha_i,\) i.e. the equilibrium communication strategy is \(\sigma^m_i (\alpha_i) = \alpha_i.\)

The policy maker’s optimal policy is simply the expected value of \(\alpha_R - \alpha_L\) given the strategies of the experts. Let us suppose without loss of generality that \(\alpha_L > \alpha_R,\) i.e. that expert \(L\) receives stronger evidence and therefore wins access to the policy maker. Expert \(L\) then sends a message \(m_L = \alpha_L.\) The policy maker infers that, since expert \(R\) did not gain access to her, \(\alpha_R\) must have been lower than \(\alpha_L.\) She forms a belief that \(\alpha_R\) is uniformly distributed on \([0, m_L],\) and thus the expected value of \(\alpha_R\) given her information is equal to \(\frac{m_L}{2}\). The policy maker’s optimal policy upon receiving

\(^5\)Throughout the paper, I restrict the analysis to symmetric equilibria.
a message \( m_L \) is then \( \frac{m_f}{2} - m_L = -\frac{m_L}{2} \). Conversely, upon receiving a message \( m_R \), her optimal policy is \( \frac{m_R}{2} \).

We can further note that the equilibrium competition strategy \( \sigma^e_i (\alpha_i) \) is a convex function of \( \alpha_i \). The intuition for this is as follows. When \( \alpha_i \) is low, expert \( i \)'s marginal benefit from increasing effort is relatively low because the difference between the policy upon winning and upon losing is small: if the expert wins, the policy is still very close to zero (in the direction which is desirable for him), and if he loses, the expected policy is also relatively close to zero (in the direction which is undesirable for him) because almost any opponent can beat him. In order for the marginal cost of effort to equal the marginal benefit, the slope of \( \sigma^e_i (\alpha_i) \) needs to be low. When \( \alpha_i \) is high, the situation is different: the marginal benefit from increasing effort is relatively high. Now the difference between the policy upon winning and upon losing is large: if the expert wins, the policy is far from zero (in the direction which is desirable for him) because his \( \alpha_i \) is high, but if he loses, the expected policy is also far from zero (in the direction which is undesirable for him) because only an opponent with an even higher \( \alpha_i \) can beat him. Therefore, for the marginal cost of effort to equal the marginal benefit, the slope of \( \sigma^e_i (\alpha_i) \) needs to be high for high \( \alpha_i \)'s, which means that the competition function needs to become steeper as \( \alpha_i \) increases.

It is intuitive that the equilibrium competition strategy is increasing in \( v \). As the policy becomes more important for the experts relative to their effort, each expert is willing to exert more effort to increase the probability of winning and, hence, of a favourable policy.

### 3.2.3 Welfare Analysis

In this section, I analyse the players’ welfare, i.e. their ex-ante expected payoffs in equilibrium.

First, I look at the experts’ welfare. To begin with, the interim expected payoff to expert \( i \) (i.e. after observing the strength of evidence, \( \alpha_i \), but before choosing how much effort, \( e_i \), to exert in the competition for access) is

\[
E^{PC}_i [\Pi_i | \alpha_i] = v \left[ \int_0^{\alpha_i} \left( \frac{\alpha_i}{2} \right) d\alpha_j + \int_{\alpha_i}^{1} \left( -\frac{\alpha_j}{2} \right) d\alpha_j \right] - \frac{v}{2} \alpha_i^2 \\
= \frac{v}{4} \left( \alpha_i^2 - 1 \right).
\]
Therefore, the interim expected payoff upon observing \( \alpha_i \) is increasing in \( \alpha_i \) and weakly negative (equal to zero when \( \alpha_i = 1 \)). An expert’s welfare is then obtained by integrating the expression in (3.7) with respect to \( \alpha_i \) on the interval \([0, 1]\):

\[
E^{PC}[\Pi_i] = \int_0^1 \frac{v}{4} (\alpha_i^2 - 1) \, d\alpha_i = -\frac{v}{6}.
\]

(3.8)

Hence, an expert’s welfare is negative. The reason for this is that, ex ante, both experts have an equal chance of swaying the policy in their preferred direction, and thus the ex-ante expected policy is zero; however, the ex-ante expected effort is positive for each expert. An expert’s welfare is also decreasing in \( v \). Intuitively, a higher \( v \) has two effects: on the one hand, it means that an expert attaches more value to the policy, but on the other hand, it implies that he exerts more effort in equilibrium. The former effect has no impact on his welfare, as the ex-ante expected policy is zero, but the latter effect increases his ex-ante expected effort, and thus a higher \( v \) reduces his welfare.

The policy maker’s welfare is

\[
E^{PC}[\Pi_{PM}] = \int_0^1 \int_0^1 \left[ \alpha_R \left( -\frac{\alpha_L}{2} \right) + \alpha_L \left( \frac{\alpha_L}{2} \right) - \left( -\frac{\alpha_L}{2} \right)^2 \right] \, d\alpha_L \, d\alpha_R + \\
\int_0^1 \int_0^1 \left[ \alpha_R \left( \frac{\alpha_R}{2} \right) + \alpha_L \left( -\frac{\alpha_R}{2} \right) - \left( \frac{\alpha_R}{2} \right)^2 \right] \, d\alpha_R \, d\alpha_L = \frac{1}{16}.
\]

(3.9)

The first double integral in (3.9) represents the policy maker’s expected payoff when expert \( L \) wins access, and the second double integral represents her expected payoff when expert \( R \) wins. The policy maker’s welfare does not depend on \( v \) because her choice of policy depends only on the winning expert’s message (which is not a function of \( v \)) and she does not benefit directly from the experts’ efforts (which are a function of \( v \)).

The policy maker’s welfare is positive, which is not surprising given that the game allows the policy maker to obtain some information about the experts’ strength of evidence and thus to make a better choice of the policy. However, since she effectively observes only the winner’s strength evidence and cannot precisely infer the loser’s strength of evidence, her welfare is lower than in the first best. In the first best, i.e. if she knew
the values of \( \alpha_R \) and \( \alpha_L \), her welfare would be

\[
\mathbb{E}^{FB}[\Pi_{PM}] = \int_0^1 \int_0^1 \left[ \alpha_R (\alpha_R - \alpha_L) + \alpha_L (\alpha_L - \alpha_R) - \frac{(\alpha_R - \alpha_L)^2}{2} \right] d\alpha_L d\alpha_R \\
= \frac{1}{12}.
\] (3.10)

Finally, since the experts’ welfare is decreasing in \( v \) and the policy maker’s welfare does not depend on \( v \), the total welfare is also decreasing in \( v \). Intuitively, an increase in \( v \) does not help the policy maker make a better policy choice, but it increases the ex-ante expected effort of the experts in the competition for access.

### 3.2.4 Competition for Access versus Hiring in Advance

It is instructive to compare two scenarios: (i) one in which two experts compete for access to the policy maker, as analysed above, and (ii) one in which the policy maker randomly chooses one of the experts in advance. The aim of this comparison is to provide us insight into the policy maker’s benefit from the competition for access to her.

If the policy maker randomly chooses one of two experts in advance, then she receives a message from the chosen expert, but she cannot make any further inference about the other expert’s strength of evidence. Therefore, her belief remains the same as before the game: she believes that the other expert’s \( \alpha_i \) is a random variable uniformly distributed on \([0, 1]\) and thus the expected value of that expert’s \( \alpha_i \) is equal to \( \frac{1}{2} \). The policy maker’s optimal policy upon receiving a message \( m_L \) is then given by \( \frac{1}{2} - m_L \). Conversely, the optimal policy upon receiving \( m_R \) is \( m_R - \frac{1}{2} \). In the perfect Bayesian equilibrium, the players’ strategies are given by

\[
\sigma_i^m = \alpha_i, \quad \sigma_i^p = \begin{cases} \frac{1}{2} - m_i & \text{if } i = L \\ m_i - \frac{1}{2} & \text{if } i = R \end{cases}.
\] (3.11) (3.12)
The policy maker’s welfare in this scenario is

\[ \mathbb{E}_{RC}^{PC} [\Pi_{PM}] = \int_0^1 \int_0^1 \left[ \alpha_R \left( \frac{1}{2} - \alpha_L \right) + \alpha_L \left( \alpha_R - \frac{1}{2} \right) - \frac{(\frac{1}{2} - \alpha_L)^2}{2} \right] d\alpha_L d\alpha_R \]

\[ = \int_0^1 \int_0^1 \left[ \alpha_R \left( \alpha_R - \frac{1}{2} \right) + \alpha_L \left( \frac{1}{2} - \alpha_R \right) - \frac{(\alpha_R - \frac{1}{2})^2}{2} \right] d\alpha_L d\alpha_R \]

\[ = \frac{1}{24}. \] (3.13)

Naturally, regardless of whether the policy maker randomly chooses expert \( L \) or expert \( R \) in advance, her ex-ante expected payoff is the same. Since (3.9) is greater than (3.13), we conclude that the policy maker’s welfare is lower when an expert is chosen in advance than when the experts compete for access to the policy maker. The reason for this is that competition for access provides her some information about the losing expert’s strength of evidence, while choosing an expert in advance does not provide any information about the strength of evidence of the expert who is not chosen.

### 3.3 Model with Imperfect Communication

In this section, I introduce imperfect communication. More precisely, with some probability the communication between an expert and the policy maker is unsuccessful, i.e. the policy maker does not receive or understand the message sent by an expert. I allow the experts to make costly investments at the beginning of the game which increase the probability of successful communication. Thus, the probability of successful communication is endogenous.

#### 3.3.1 Model Setup

The game now proceeds as follows. At the beginning of the game, each expert \( i \) decides on the level of his investment, \( g_i \), in his ability to communicate successfully. Each expert then observes the strength of his evidence, given by \( \alpha_i \), and competes for access to the policy maker by exerting effort, \( e_i \). The expert who exerts more effort wins access to the policy maker and gets an opportunity to send a message to the policy maker. The winner’s investment, \( g_i \), determines the probability of successful communication, \( \pi(g_i) \). With probability \( \pi(g_i) \), the communication is successful, i.e. the policy maker receives and understands the message of the expert. However, with probability \( 1 - \pi(g_i) \),
the communication is unsuccessful, in which case the policy maker observes an empty information set. Whether the communication is successful is known only after the expert sends his message to the policy maker. I assume that \( \pi(g_i) = \min\{f(g_i), 1\} \), where \( f(0) = 0 \), \( f'(g_i) > 0 \), \( f''(g_i) < 0 \), and \( \lim_{g_i \to 0} f'(g_i) = +\infty \). This function captures the fact that investment improves the probability of successful communication, but the marginal impact decreases as investment increases.

The experts’ payoff functions only change in that they now include the experts’ cost of investments in the communication skills, \( g_i \):

\[
\Pi_L = -vp - e_L - g_L, \quad (3.14)
\]
\[
\Pi_R = vp - e_R - g_R. \quad (3.15)
\]

The policy maker’s payoff function remains the same, i.e. it is given by (3.3).

Apart from the competition and the communication strategy, each expert \( i \) now also has an investment strategy, \( \sigma_i^g \), which describes the level of his investment, \( g_i \in [0, 1] \), in his ability to communicate successfully. The policy maker now chooses a policy upon observing an information set \( (I_m, w, m_w) \), where \( I_m \in \{0, 1\} \) indicates whether the communication between the winning expert and the policy maker is successful or not. The term \( I_m = 0 \) denotes unsuccessful communication, in which case the policy maker does not observe the winner’s identity nor the winner’s message, i.e. \( w = \emptyset \) and \( m_w = \emptyset \). The policy maker’s strategy, \( \sigma^p \), is thus a map \( \sigma^p : \{0, 1\} \times \{L, R, \emptyset\} \times ([0, 1] \cup \{\emptyset\}) \to [-1, 1] \).

### 3.3.2 Equilibrium

The equilibrium concept is, like in the benchmark model, the perfect Bayesian equilibrium. The main differences are that now we also need to specify the experts’ investment strategies and that the policy maker needs to make a Bayesian inference about the experts’ strengths of evidence, \( \alpha_i \), in the case of communication being unsuccessful. The following proposition describes the unique equilibrium of the game with imperfect communication.

**Proposition 3.2.** The game has a unique perfect Bayesian equilibrium. The experts’

\(^6\)The assumption that the policy maker observes an empty information set ensures that the experts have no interest in reducing the probability of successful communication upon observing their strength of evidence.
and the policy maker’s equilibrium strategies are

\[ \sigma^g_i = \begin{cases} g_i^\dagger \text{ such that } f'(g_i^\dagger) = \frac{12}{v} & \text{if } f(g_i^\dagger) \leq 1 \\ \arg_{g_i} f(g_i) = 1 & \text{otherwise} \end{cases} \quad (3.16) \]

\[ \sigma^e_i(\alpha_i) = \frac{v}{2} (\pi(g_i^*)) \alpha_i^2, \quad (3.17) \]

\[ \sigma^m_i(\alpha_i) = \alpha_i, \quad (3.18) \]

\[ \sigma^p(I_m, w, m_w) = \begin{cases} -\frac{m_w}{2} & \text{if } w = L \text{ and } I_m = 1, \\ \frac{m_w}{2} & \text{if } w = R \text{ and } I_m = 1, \\ 0 & \text{if } I_m = 0, \end{cases} \quad (3.19) \]

where \( i \in \{L, R\} \) and \( g_i^* \) is the equilibrium investment of expert \( i \).

In the interior solution for the equilibrium investment strategy \( \sigma^g_i \), the marginal impact of investment on the probability of successful communication equals \( \frac{12}{v} \). Since \( f''(g_i) < 0 \) and \( \lim_{g_i \to 0} f'(g_i) = +\infty \), it follows that the equilibrium investment in the interior solution is positive and increasing in \( v \). If the value of \( f(g_i^\dagger) \) such that \( f'(g_i^\dagger) = \frac{12}{v} \) is above 1, then the equilibrium investment is \( \arg_{g_i} f(g_i) = 1 \), i.e. the lowest value of investment that guarantees that communication is successful with probability 1, \( \pi(g_i^*) = 1 \).

The equilibrium competition strategy \( \sigma^e_i(\alpha_i) \) of each expert \( i \) is increasing in the strength of his evidence, \( \alpha_i \), and is in fact a convex function of \( \alpha_i \), like in the benchmark model. The effort exerted in equilibrium is weakly lower than in the perfect communication benchmark, as (3.17) is weakly lower than (3.4) (they are only equal if the equilibrium investment is such that \( \pi(g_i^*) = 1 \)). The intuition is that, due to the possibility of unsuccessful communication, the expected prize for the winner of the competition is smaller and thus the experts have a smaller incentive to exert effort in the competition for access.

At the communication stage, no expert is willing to understate his message by reporting \( m_i < \alpha_i \), and thus the communication strategy is to report truthfully one’s strength of evidence, \( \sigma^m_i(\alpha_i) = \alpha_i \), which is the same as in the perfect communication benchmark.

The policy maker’s strategy now takes into account the possibility that she may receive no message due to the communication being unsuccessful. Like in the benchmark model, upon receiving a message \( m_i = \alpha_i \) from the winning expert, the policy maker
can infer that the other expert’s $\alpha_i$ must have been lower. This leads her to choose a policy $-\frac{m_i}{2}$ if she receives a message from expert $L$ and a policy $\frac{m_i}{2}$ if she receives a message from expert $R$. If the communication is unsuccessful, the policy maker cannot make any inference about the experts’ evidence and hence chooses a policy equal to 0.

For illustration, I now consider an example of a specific functional form of $f(g_i) = \sqrt{g_i}$, which means that the probability of successful communication is given by $\pi(g_i) = \min\{\sqrt{g_i}, 1\}$. This allows me to pin down the equilibrium investment and competition strategies of the experts.

**Example 3.1.** Suppose that $\pi(g_i) = \min\{\sqrt{g_i}, 1\}$. Then, in the unique symmetric perfect Bayesian equilibrium, the experts’ equilibrium investment and competition strategies are

$$
\sigma^g_i = \min\left\{ \left(\frac{v}{24}\right)^2, 1 \right\}, \quad (3.20)
$$

$$
\sigma^e_i(\alpha_i) = \min\left\{ \frac{v^2}{48\alpha_i^2}, \frac{v}{2\alpha_i^2} \right\}, \quad (3.21)
$$

where $i \in \{L, R\}$, and the experts’ communication strategy, $\sigma^m_i(\alpha_i)$, and the policy maker’s strategy, $\sigma^p(w, I_m, m_w)$, are as in Proposition 3.2.

Thus, in the above example, the equilibrium investment is $\left(\frac{v}{24}\right)^2$ if $v \leq 24$ and 1 otherwise. Therefore, it is increasing in $v$ for $v \leq 24$ and constant in $v$ otherwise. The equilibrium effort in the competition is $\frac{v^2}{48\alpha_i^2}$ for $v \leq 24$ and $\frac{v}{2\alpha_i^2}$ otherwise. Hence, it is increasing in $v$ and also increasing in and convex in $\alpha_i$; however, it is weakly lower than the equilibrium effort in the perfect communication benchmark, which equals $\frac{v}{2}\alpha_i^2$ for all $v$.

### 3.3.3 Welfare Analysis

I now analyse the players’ welfare in the game with imperfect communication.

The interim expected payoff to expert $i$ upon observing $\alpha_i$ and for a given equilibrium
investment $g_i^*$ is
\[
\mathbb{E}^{IC}[\Pi_i | \alpha_i, g_i^*] = v \left[ \pi(g_i^*) \int_0^{\alpha_i} \frac{\alpha_i}{2} d\alpha_j + \pi(g_j^*) \int_{\alpha_i}^1 \left( -\frac{\alpha_j}{2} \right) d\alpha_j \right] - \frac{v}{2} (\pi(g_i^*)) \alpha_i^2 - g_i^* \\
= \frac{v}{4} \pi(g_i^*) (\alpha_i^2 - 1) - g_i^*,
\]
where the last equality uses the fact that $g_i^* = g_j^*$. By integrating the expression in (3.22) with respect to $\alpha_i$ on the interval $[0, 1]$, we obtain an expert $i$'s welfare:
\[
\mathbb{E}^{IC}[\Pi_i] = -\frac{v}{6} \pi(g_i^*) - g_i^*.
\]
Since $g_i^* > 0$ and $\pi(g_i^*) > 0$, an expert’s welfare is negative. The intuition for his welfare being negative is similar to the one in the benchmark model: both experts have an equal chance of swaying the policy in their preferred direction, and thus the ex-ante expected policy is zero, while the ex-ante expected investment and effort are positive for each expert. Furthermore, given that both $\pi(g_i^*)$ and $g_i^*$ are increasing in $v$, an expert’s welfare is decreasing in $v$. Intuitively, a higher $v$ means that, on the one hand, an expert attaches more value to the policy but, on the other hand, he invests more in his communication skills and exerts more effort in competition. The former has no impact on his welfare (because the ex-ante expected policy is zero) while the latter decreases it, so a higher $v$ reduces his welfare.

The policy maker’s welfare is
\[
\mathbb{E}^{IC}[\Pi_{PM}] = \int_0^1 \int_0^1 \left[ \pi(g_L^*) \left( \alpha_R \left( -\frac{\alpha_L}{2} \right) + \alpha_L \left( \frac{\alpha_R}{2} \right) - \frac{1}{4} \right) \right] d\alpha_L d\alpha_R \\
+ \int_0^1 \int_0^1 \left[ \pi(g_R^*) \left( \alpha_R \left( \frac{\alpha_L}{2} \right) + \alpha_L \left( -\frac{\alpha_R}{2} \right) - \frac{1}{4} \right) \right] d\alpha_R d\alpha_L \\
= \frac{1}{16} \pi(g_i^*).
\]
The first double integral in (3.24) represents the policy maker’s expected payoff when expert $L$ wins access while accounting for the possibility that communication is successful only with probability $\pi(g_i^*)$. Similarly, the second double integral represents the policy maker’s expected payoff when expert $R$ wins access.
The policy maker’s welfare is positive, since $\pi(g_i^*) > 0$. However, it is weakly lower than in the game with perfect communication, as $3.24$ is lower than $3.9$ for $\pi(g_i^*) < 1$ and they are equal for $\pi(g_i^*) = 1$.

Furthermore, the policy maker’s welfare is increasing in $v$, which follows from the equilibrium investment, $g_i^*$, being increasing in $v$. Intuitively, a higher $v$ makes the experts invest more in their communication skills and hence there is a higher probability that the policy maker obtains information about the strength of their evidence, which helps her choose a better policy. This contrasts with the perfect communication model, where the policy maker’s welfare does not depend on $v$.

### 3.3.4 Competition for Access versus Hiring in Advance

I now compare the game with imperfect communication under two scenarios: (i) one in which two experts compete for access to the policy maker, and (ii) one in which the policy maker randomly chooses one of the experts in advance.

Consider the scenario in which the policy maker randomly chooses one of the experts in advance. The perfect Bayesian equilibrium in this setup is given by

$$\sigma_i^g = 0,$$  \hspace{1cm} (3.25)

$$\sigma_i^m(\alpha_i) = \alpha_i,$$  \hspace{1cm} (3.26)

$$\sigma^p(I_m, w, m_w) = \begin{cases} 
\frac{1}{2} - m_w & \text{if } w = L \text{ and } I_m = 1, \\
0 & \text{if } I_m = 0, \\
m_w - \frac{1}{2} & \text{if } w = R \text{ and } I_m = 1.
\end{cases}$$  \hspace{1cm} (3.27)

where $i \in \{L, R\}$.

Like in the benchmark model, the expected value of the strength of evidence of the expert who has not been chosen is equal to $\frac{1}{2}$. Thus, the optimal policy for the policy maker upon receiving a message $m_L$ is given by $\frac{1}{2} - m_L$ and her optimal policy upon receiving a message $m_R$ is $m_R - \frac{1}{2}$. If the chosen expert communicates unsuccessfully, then the policy maker does not learn anything about the values of $\alpha_L$ and $\alpha_R$, and so her optimal policy is $0$.

It turns out that the experts’ investment strategy in this setup is to invest zero in their ability to communicate successfully. To see why it is so, note that the chosen expert’s (suppose it is expert $L$) expected payoff for a given $g_L$ but before observing $\alpha_L$
can be expressed as
\[
\mathbb{E}_{RC}^{IC} [\Pi_L \mid g_L] = \int_0^1 \left[ \pi(g_L) \times \left( \frac{1}{2} - \alpha_L \right) - g_L \right] d\alpha_L = -g_L. \tag{3.28}
\]
Hence, the optimal investment of expert \( L \) is \( g^*_L = 0 \). Similarly, if expert \( R \) were chosen in advance, his expected payoff for a given \( g_R \) would be equal to \( -g_R \), and thus the optimal investment would be \( g^*_R = 0 \).

Given that the equilibrium investment is equal to zero, the probability of successful communication is zero\(^7\) and thus the policy maker’s welfare in this scenario is also zero:
\[
\mathbb{E}_{RC}^{IC} [\Pi_{PM}] = 0. \tag{3.29}
\]
Therefore, the policy maker’s welfare is lower here than in the scenario of competition for access, where it is positive. In other words, the policy maker is better off having the experts compete for access to her than hiring an expert in advance.

The above comparison shows that competition for access is beneficial to the policy maker not only because it provides information about the losing expert’s strength of evidence, but also because it provides an incentive for the experts to invest in their ability to communicate successfully. An expert chosen in advance has no incentive to make any investment because of the way the information from him is used in order to make a decision on the policy. His information is “discounted” by the policy maker more severely when he is chosen in advance than when he wins competition for access. In the former case, the policy maker expects that the other agent’s strength of evidence is equal to \( \frac{1}{2} \). In the latter case, she expects it to be half of the winner’s strength of evidence, which means that she expects it to be \( \frac{1}{2} \) only if the winner’s strength evidence is \( \alpha_w = 1 \) and lower than \( \frac{1}{2} \) whenever \( \alpha_w < 1 \).

The following table summarises the results on the policy maker’s welfare from Sections 3.2 and 3.3.

\(^7\)Therefore, all the information sets with the property that the message reaches the policy maker are off the equilibrium path. Technically, in a perfect Bayesian equilibrium, the policy maker’s beliefs can be arbitrary for these information sets. I assume that, in these information sets, the belief of the policy maker is the same as her belief conditional on receiving a message, given the strategies of the experts.
<table>
<thead>
<tr>
<th></th>
<th>perfect communication</th>
<th>imperfect communication</th>
</tr>
</thead>
<tbody>
<tr>
<td>competition for access</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{16} \pi(g^*)$</td>
</tr>
<tr>
<td>hiring an expert in advance</td>
<td>$\frac{1}{24}$</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 3.1: The policy maker’s welfare from the two procedures, i.e. (i) the competition between experts for access and (ii) hiring an expert in advance, in the perfect communication benchmark and the model with imperfect communication.

### 3.4 Providing Incentives for Investment in Communication Skills

In this section, I extend the model by allowing the policy maker to incentivise the experts to invest in their ability to communicate successfully by offering them a reward for successful communication. The reward is best understood as a monetary payment. For example, consultants, advisers, etc., are usually paid for their advice regardless of its content as long as it is sufficiently well researched and prepared.

Now, at the beginning of the game, the policy maker can announce a reward, $r \geq 0$, which is paid by her to an expert on the condition that he successfully communicates with her. Such a reward is naturally a cost to the policy maker, but if it provides a sufficient incentive to the experts to invest in their ability to communicate successfully, it could potentially make the policy maker better off. The reward takes the form of a lump-sum payment, i.e. it does not depend on the level of evidence $\alpha_i$. There is no need to incentivise any expert to truthfully reveal $\alpha_i$ because he has an incentive to do so due to his preferences and the verifiability of information alone.

The following proposition describes the effect of the possibility of offering a reward on the experts’ investment strategy when they compete for access. It also describes the optimal reward from the perspective of the policy maker.

**Proposition 3.3.** Suppose that there is a reward $r \geq 0$ for successful communication. If two experts compete for access to a policy maker, then the experts’ investment strategy is

$$\sigma_i^g = \begin{cases} 
  g_i^\dagger \text{ such that } f'(g_i^\dagger) = \frac{12}{v} & \text{if } f(g_i^\dagger) \leq 1 \\
  \arg_{g_i} f(g_i) = 1 & \text{otherwise}
\end{cases},$$

(3.30)

and thus it does not depend on $r$. The policy maker’s optimal reward is $r^* = 0$.

Proposition 3.3 tells us that a reward $r \geq 0$ has no effect on the experts’ investment
strategy. The reasoning behind this is as follows. On the one hand, a reward provides an incentive for the experts to invest in their communication skills because if an expert wins access and communication turns out to be successful, he benefits from the reward. However, the reward also means that, for a given investment, the expert exerts more effort in the competition for access to the policy maker. This increased effort completely offsets the expected benefit from the reward, and thus the expert’s investment does not depend on the reward. Since any positive reward, \( r > 0 \), would be a cost for the policy maker, it is optimal for her to offer no reward, i.e. \( r^* = 0 \).

I now turn to the scenario where the policy maker randomly chooses an expert in advance.

**Proposition 3.4.** Suppose that there is a reward \( r \geq 0 \) for successful communication. If the policy maker randomly chooses one expert in advance, then the expert’s investment strategy is

\[
\sigma^*_g = \begin{cases} 
  g^*_i & \text{such that } f'(g^*_i) = \frac{1}{r} \text{ if } f(g^*_i) \leq 1 \\
  \arg_{g_i} f(g_i) = 1 & \text{otherwise}
\end{cases}
\]

and thus the investment is increasing in \( r \). The policy maker’s optimal reward is positive, \( r^* > 0 \).

When an expert is hired in advance, then the expert’s investment strategy does depend on the reward. Here, a reward provides an incentive for the chosen expert to invest because if communication turns out to be successful, he receives the reward; however, the effect of the reward on the expert’s effort in competition is now absent. Hence, the chosen expert’s investment strategy is increasing in the reward, and it is now possible for the policy maker to incentivise him to invest more. The optimal reward is positive. If the policy maker offered no reward, then the marginal benefit to her from increasing the reward would be infinitely large and thus it would exceed the marginal cost.

For illustration, consider the example of \( f(g_i) = \sqrt{g_i} \), which implies that \( \pi(g_i) = \min\{\sqrt{g_i}, 1\} \). This allows me to pin down the experts’ investment strategy as a function of \( r \), \( \sigma^*_g(r) \), and the policy maker’s optimal reward, \( r^* \).

**Example 3.2.** Suppose that \( \pi(g_i) = \min\{\sqrt{g_i}, 1\} \) and that the policy maker randomly chooses one expert in advance. Then, the expert’s investment strategy as a function of \( r \) is \( \sigma^*_g(r) = \min\{\frac{1}{4}r^2, 1\} \) and the policy maker’s optimal reward is \( r^* = \frac{1}{48} \).
Clearly, in the above example, \( \sigma^g(r) \) is increasing in \( r \) for \( r \leq 2 \) and is equal to 1 otherwise. Furthermore, \( r^* \) is positive.

The main message behind Propositions 3.3 and 3.4 is that, from the perspective of the policy maker, hiring an expert in advance has one advantage over having the experts compete for access: the former allows the policy maker to provide incentives for investment in communication skills by a reward, whereas the latter does not. The policy maker therefore faces a dilemma between hiring an expert in advance and having the experts compete for access as each of these two approaches has its advantages and disadvantages. Competition for access is beneficial to the policy maker in that it reveals information about the losing expert’s strength of evidence and provides intrinsic incentives for the experts to invest in their communication skills. Hiring an expert in advance is beneficial in that it allows the policy maker to incentivise the chosen expert to invest more by offering him a reward.

The above analysis leads us to the following question: which of the two procedures does the policy maker find optimal?

**Proposition 3.5.** The policy maker is better off having the experts compete for access if and only if the experts attach a sufficiently high value to the policy, i.e. \( v \) is sufficiently high. Otherwise, the policy maker is better off randomly choosing one of the experts in advance.

The intuition for the result in Proposition 3.5 is as follows. In the competition for access scenario, the experts’ investments in communication skills are increasing in how much value they attach to the policy maker’s policy, which is measured by parameter \( v \). Then, for low values of \( v \), hiring an expert in advance is an attractive option because the experts’ investments in the scenario of competition for access are low, while hiring an expert in advance offers the possibility of incentivising the expert by a reward. As \( v \) increases, competition for access becomes more attractive because then the experts’ intrinsic incentives to invest increase. Once \( v \) is high enough, the investments are so high that it is optimal for the policy maker to have the experts compete for access to her.

I now consider the example of a specific functional form of \( f(g_i) = \sqrt{g_i} \), which implies that \( \pi(g_i) = \min\{\sqrt{g_i}, 1\} \). The objective is to pin down the threshold value of \( v \), above which it is optimal to have the experts compete for access.

**Example 3.3.** Suppose that \( \pi(g_i) = \min\{\sqrt{g_i}, 1\} \). If the experts compete for access, then the policy maker’s ex-ante expected payoff is \( \min\{\frac{v}{15}, \frac{1}{16}\} \). If the policy maker
randomly chooses an expert in advance, then her policy maker’s ex-ante expected payoff achieves its maximum at $\frac{1}{4608}$ with a reward of $r^* = \frac{1}{48}$. Thus, for $\frac{v}{384} > \frac{1}{4608}$, which can be rearranged to $v > \frac{1}{12}$, the policy maker is better off having the experts compete for access. If $v < \frac{1}{12}$, she is better off randomly choosing one of the experts in advance.

The lesson from Proposition 3.5 is that whether competition for access or random choice of an expert is more beneficial to the policy maker depends on how much value the experts attach to the policy relative to their efforts, investments and the reward. Hiring an expert in advance is more beneficial as long as the expert is sufficiently neutral (in the sense that he attaches relatively little value to the policy implemented by the policy maker) and it must be accompanied by a promise of a reward for successful communication. If the experts are less neutral in the sense that they attach a greater value to the policy, then the policy maker is better off if they compete for access to her by exerting effort.

3.5 What if Efforts in the Competition for Access are Observable?

In my paper, I assume that the experts’ efforts are unobservable to the policy maker so as to minimise the role of signalling one’s strength of evidence by effort, and to look at other advantages of the competition between experts for access to a policy maker. In this section, I make an alternative assumption: the policy maker observes the efforts of experts.

Perfect communication benchmark. To start with, the following proposition describes the equilibrium in the perfect communication benchmark. Cotton (2009) studies a very similar setup with one important difference: the policy maker also directly benefits from the experts’ efforts (which are referred to as “contributions” in his paper).

Proposition 3.6. In the perfect communication benchmark, if the policy maker observes the efforts of both experts, then the players’ strategies in the unique equilibrium
are

\[ \sigma^e_i (\alpha_i) = v\alpha_i \left(1 - \frac{1}{2}\alpha_i \right), \quad (3.32) \]
\[ \sigma^m_i (\alpha_i) = \alpha_i, \quad (3.33) \]
\[ \sigma^p (w, m, \epsilon, e_L, e_R) = p^*, \quad (3.34) \]

where \( i \in \{L, R\} \) and \( p^* = \alpha_R - \alpha_L \).

There are several differences between the equilibrium in this setup and the setup with unobservable efforts.

First, with observable efforts, a one-to-one mapping between an expert’s strength of evidence and his effort allows the policy maker to become fully informed about the strength of evidence of both experts. This implies that the policy maker’s welfare is the same as in the first best, i.e. as in (3.10).

Second, the slope of the experts’ competition strategy is concave in the strength of evidence, \( \alpha_i \), rather than convex like in the setup with unobservable efforts. Intuitively, when \( \alpha_i \) is low, the marginal benefit from increasing effort is relatively high because—by exerting more effort—an expert can signal a higher strength of evidence and, with high probability, the policy maker will not be able to verify this because an expert with a low \( \alpha_i \) is unlikely to win access. On the other hand, when \( \alpha_i \) is high, the marginal benefit from increasing effort is relatively low. An expert with a high \( \alpha_i \) most likely gains access and the policy maker effectively gets to know his true \( \alpha_i \), so overspending on the competition does not bring much benefit. Therefore, in equilibrium, the slope of the competition strategy needs to be concave in the strength of evidence.

Third, the expected efforts of experts are lower when the efforts are unobservable than when they are observable. In the former case, the expected effort of expert \( i \) is

\[ \mathbb{E} [e_i] = \int_0^1 \left[v\alpha_i^2 \right] d\alpha_i = \frac{v}{6}, \quad (3.35) \]

whereas in the latter case it is

\[ \mathbb{E} [e_i] = \int_0^1 \left[v\alpha_i \left(1 - \frac{1}{2}\alpha_i \right) \right] d\alpha_i = \frac{v}{3}. \quad (3.36) \]

The difference in the ex-ante expected efforts of the experts in these two cases has

\[ ^{8} \text{The intuition for the concavity is essentially the same as in Cotton (2009).} \]
implications for how lobbying should be organised by a policy maker. If the primary aim of the policy maker is to maximise the efforts (e.g., because they are financial contributions and the policy maker benefits significantly from them), then lobbying should be organised in a way such that the lobbyists’ efforts are observable. However, if the objective is to minimise the efforts (e.g., because they are significant losses incurred by the experts and the policy maker does not directly benefit from them), then the policy maker should not be able to observe the lobbyists’ efforts.

Model with imperfect communication. I now turn to the setup with imperfect communication. I assume that, if communication is unsuccessful, then the policy maker observes an empty information set. The following proposition describes the equilibrium in this scenario.

**Proposition 3.7.** In the model with imperfect communication, if the policy maker observes the efforts of the experts, then the players’ strategies in the unique equilibrium are

\[
\sigma_i^g = \begin{cases} 
  g_i^\dagger \text{ such that } f'(g_i^\dagger) = \frac{6}{v} & \text{if } f(g_i^\dagger) \leq 1 \\
  \arg_{g_i} f(g_i) = 1 & \text{otherwise}
\end{cases} 
\]  

(3.37)

\[
\sigma_i^e (\alpha_i) = v \alpha_i \left( 1 - \frac{1}{2} \alpha_i \right) \pi(g_i^*) 
\]  

(3.38)

\[
\sigma_i^m (\alpha_i) = \alpha_i 
\]  

(3.39)

\[
\sigma^p (I_m, w, m_w, e_L, e_R) = \begin{cases} 
  p^* & \text{if } I_m = 1, \\
  0 & \text{if } I_m = \emptyset
\end{cases} 
\]  

(3.40)

where \( i \in \{L, R\} \) and \( p^* = \alpha_R - \alpha_L \).

There are a few differences between the equilibrium in this setup and in the setup with unobservable efforts. Like under unobservable efforts, the equilibrium investment in the interior solution is positive and increasing in \( v \). However, it is weakly greater here than under unobservable efforts, which follows from \( \frac{6}{v} < \frac{12}{v} \) and \( f''(g_i) < 0 \). The competition strategy is different in that it is now concave in \( \alpha_i \), like in the benchmark for the setup with observable efforts.

---

9If the policy maker observes the levels of effort of the experts but does not observe the winner’s identity and message, then she cannot distinguish which expert exerted which effort. Then, she cannot make any inference, i.e. it is as if she observed an empty information set.
Impact of a reward for successful communication. Finally, I consider whether a reward can provide an incentive for the experts to invest more in their ability to communicate successfully.

**Proposition 3.8.** Suppose that there is a reward $r \geq 0$ for successful communication. In the model with imperfect communication with observable efforts, if two experts compete for access to a policy maker, then the experts’ investment strategy is

$$
\sigma_i^g = \begin{cases} 
    g_i^\dagger & \text{such that } f'(g_i^\dagger) = \frac{g_i^\dagger}{v} \quad \text{if } f(g_i^\dagger) \leq 1 \\
    \arg_{g_i} f(g_i) = 1 & \text{otherwise}
\end{cases},
$$

and thus it does not depend on the reward for successful communication, $r$. The policy maker’s optimal reward is $r^* = 0$.

Thus, like in the setup with unobservable efforts, a reward is ineffective in incentivising the experts to invest more. The reason is that, again, the benefit from increasing investment is completely offset by the consequent increased effort in the competition for access. Since a reward would be a cost to the policy maker and does not bring any benefits to her, the optimal reward is zero.

The above results imply that the welfare of the policy maker is increasing in $v$. Hence, the results on the optimal choice between competition for access and hiring a randomly chosen expert would be qualitatively the same as those for the case of unobservable efforts.

### 3.6 Conclusion

This paper studies the benefits and drawbacks which competition between experts for access brings to a policy maker, as opposed to hiring an expert in advance. The model features two experts with opposite biases who compete for access to a policy maker by exerting effort. The prize of the competition is an opportunity to communicate with the policy maker. The model assumes that the communication between an expert and the policy maker is not necessarily successful, but the experts can invest in their communication skills in order to increase the probability that it is successful. Furthermore, the policy maker can offer a monetary reward for successful communication. The main aim of the paper is to investigate how policy makers benefit from having experts compete for access to them.
Several important findings emerge in this paper. First, the policy maker benefits from the fact that if the experts compete for access to her, then she obtains information not only about the private evidence of the expert who wins access, but also about the loser’s private evidence. Second, more interestingly, competition for access provides an incentive for the experts to invest in their communication skills, while there is no such incentive if an expert is hired in advance. The reason for this is that the policy maker “discounts” a message from a randomly chosen expert more severely than a message from the winner of competition for access. Third, hiring an expert in advance has one advantage: the policy maker can use a monetary reward for successful communication to incentivise the experts to invest in their communication skills. When the experts compete for access, a reward is ineffective because it leads the experts to exert more effort in the competition, which completely offsets their expected benefit from the reward. A comparison of the two procedures reveals that if the experts attach a high value to the implemented policy, i.e. they are heavily biased in the sense that they benefit a lot if the policy is swayed in their desired direction, then competition for access is preferred by the policy maker. Otherwise, she is better off by hiring an expert in advance.

The model is stylised and naturally does not capture all elements of lobbying in the real world, but it still provides interesting insights into how policy makers benefit from limiting access to them and having experts compete for access. As an avenue for future research, it would be useful to further explore the potential advantages and disadvantages of the competition between experts for access, beyond those analysed in this paper. For example, it would be interesting to consider how competition for access affects the experts’ incentives to invest in the collection of evidence or to obfuscate their information when communicating with a policy maker. Furthermore, it would be worth analysing how these advantages and disadvantages depend on the communication protocol; in particular, one could ask how they would change if communication was by cheap talk or if reporting false information was possible but costly. For an even more ambitious research agenda, it would be particularly interesting to analyse a broader class of mechanisms than only either allowing competition for access in an all-pay auction or hiring a randomly chosen expert.
Appendix to Chapter 3

C.1 Proofs

Proof of Proposition 3.1

The proof partly follows Cotton (2009). We begin with two observations. First, each expert’s effort function, $e(\alpha_i)$, must be strictly increasing in $\alpha_i$. Since $e(\alpha_i)$ is strictly increasing in $\alpha_i$, $e(\alpha_i)$ is invertible, and so we can write $\alpha(e) = e^{-1}(\alpha)$. This also implies that there is a one-to-one mapping between $e_i$ and $\alpha_i$. Second, no expert is willing to report $m_i < \alpha_i$, because then he would be able to profitably deviate by sending a higher $m_i$. Thus, the winning expert sends $m_i = \alpha_i$.

Consider the policymaker’s policy strategy. The policymaker’s sequentially rational policy is equal to the expected value of $\alpha_R - \alpha_L$ given $\{w, m_w\}$ and given the strategies of experts $L$ and $R$. Suppose without loss of generality that $\alpha_L > \alpha_R$. Expert $L$ then wins access to the policy maker and sends $m_L = \alpha_L$. The policy maker forms a belief that $\alpha_R$ is uniformly distributed on $[0, \alpha_L]$, so the expected value of $\alpha_R$ given her information is $\frac{\alpha_L}{2}$. The policy maker’s sequentially rational policy is then equal to $\frac{\alpha_L}{2} - \alpha_R = -\frac{\alpha_L}{2}$. Conversely, upon receiving a message $m_R = \alpha_R$, the policy maker’s sequentially rational policy is $\frac{\alpha_R}{2}$.

We can now derive the experts’ competition strategy. Upon observing $\alpha_i$, expert $i$ chooses $e_i$ to maximise his expected payoff:

$$v \int_0^{\alpha(e_i)} \left[ \frac{\alpha}{2} \right] d\alpha_j + v \int_{\alpha(e_i)}^{1} \left[ -\frac{\alpha_j}{2} \right] d\alpha_j - e_i.$$  \hspace{1cm} (3.42)

By differentiating (3.42) with respect to $e_i$, we obtain the first order condition:

$$v \frac{\partial \alpha(e_i)}{\partial e_i} \frac{\alpha_i}{2} - v \frac{\partial \alpha(e_i)}{\partial e_i} \left( -\frac{\alpha(e_i)}{2} \right) - 1 = 0.$$  \hspace{1cm} (3.43)

In equilibrium, it must be that $\alpha(e_i) = \alpha_i$. Strict monotonicity implies $\left( \frac{\partial \alpha(e_i)}{\partial e_i} \right)^{-1} = e'(\alpha_i)$. Therefore, (3.43) simplifies to

$$e'(\alpha_i) = v \frac{\alpha_i}{2} + v \frac{\alpha_i}{2} = v\alpha_i.$$  \hspace{1cm} (3.44)
By integrating with respect to $\alpha_i$ we obtain

$$e(\alpha_i) = \frac{v}{2} \alpha_i^2 + C, \quad (3.45)$$

where $C$ is a constant. If we further assume that $e(0) = 0$, i.e. an expert with evidence $\alpha_i = 0$ exerts no effort, then $C = 0$. Then, the sequentially rational effort exerted by expert $i$ upon observing $\alpha_i$ is given by

$$e(\alpha_i) = \frac{v}{2} \alpha_i^2, \quad (3.46)$$

which gives us the equilibrium competition strategy.

**Proof of Proposition 3.2**

First, each expert’s effort function, $e(\alpha_i)$, must be strictly increasing in $\alpha_i$. Second, no expert is willing to report $m_i < \alpha_i$ and thus reports $m_i = \alpha_i$.

The policy maker’s sequentially rational policy is then $-\frac{\alpha_i}{2}$ if she receives a message from expert $L$ and $\frac{\alpha_i}{2}$ if she receives a message from expert $R$. If communication is unsuccessful, then her sequentially rational policy is 0.

The expected payoff to expert $i$ for given $\alpha_i$, $g_i$ and $g_j$ is therefore:

$$v \int_{0}^{\alpha(e_i)} \left[ \pi(g_i) \left( \frac{\alpha_i}{2} \right)^2 \right] d\alpha_j + v \int_{\alpha(e_i)}^{1} \left[ \pi(g_j) \left( -\frac{\alpha_j}{2} \right) \right] d\alpha_j - e_i - g_i. \quad (3.47)$$

We can now derive the sequentially rational choice of effort, $e_i$, by differentiating (3.47) with respect to $e_i$. This gives us the following first order condition:

$$v \frac{\partial \alpha(e_i)}{\partial e_i} \left[ \pi(g_i) \left( \frac{\alpha_i}{2} \right) \right] - v \frac{\partial \alpha(e_i)}{\partial e_i} \left[ \pi(g_j) \left( -\frac{\alpha_j}{2} \right) \right] = 1 = 0. \quad (3.48)$$

In equilibrium, it must be that $\alpha(e_i) = \alpha_i$. Strict monotonicity implies $\left[ \frac{\partial \alpha(e_i)}{\partial e_i} \right]^{-1} = e'(\alpha_i)$. Then, from (3.48) we obtain

$$e'(\alpha_i) = \frac{v}{2} \left( \pi(g_i) + \pi(g_j) \right) \alpha_i. \quad (3.49)$$

Integration with respect to $\alpha_i$ gives us

$$e(\alpha_i) = \frac{v}{4} \left( \pi(g_i) + \pi(g_j) \right) \alpha_i^2 + C. \quad (3.50)$$
If we further assume that $e(0) = 0$, then $C = 0$. Then, the sequentially rational effort exerted by expert $i$ upon observing $\alpha_i$ is given by

$$e(\alpha_i) = \frac{v}{4} (\pi(g_i) + \pi(g_j)) \alpha_i^2,$$  \hspace{1cm} (3.51)

which gives us the equilibrium competition strategy as a function of $g_i$ and $g_j$.

Given the above communication strategy and competition strategy (with $C = 0$), the expected payoff to expert $i$ for given $\alpha_i$ and $\pi(g_i)$ is

$$v \int_0^{\alpha(e_i)} \left[ \pi(g_i) \frac{\alpha_i}{2} \right] d\alpha_j + v \int_{\alpha(e_i)}^{1} \left[ \pi(g_j) \left( -\frac{\alpha_j}{2} \right) \right] d\alpha_j - \frac{v}{4} (\pi(g_i) + \pi(g_j)) \alpha_i^2 - g_i, \hspace{1cm} (3.52)$$

which can be rearranged to

$$\frac{v}{4} \left( \pi(g_i) \alpha_i^2 - \pi(g_j) \right) - g_i. \hspace{1cm} (3.53)$$

By integrating the expression in (3.53) with respect to $\alpha_i$ on the interval $[0, 1]$, we obtain the expected payoff to expert $i$ for a given $\pi(g_i)$ but before observing $\alpha_i$:

$$\int_0^{1} \left[ \frac{v}{4} \left( \pi(g_i) \alpha_i^2 - \pi(g_j) \right) - g_i \right] d\alpha_i = \frac{v}{4} \left( \frac{1}{3} \pi(g_i) - \pi(g_j) \right) - g_i. \hspace{1cm} (3.54)$$

Then, (3.54) is maximised at $g_i^\dagger$ such that

$$f'(g_i^\dagger) = \frac{12}{v}, \hspace{1cm} (3.55)$$

if $f(g_i^\dagger) \leq 1$, and at $\arg_{g_i} f(g_i) = 1$ otherwise, which yields the equilibrium investment strategy. Then, given that the equilibrium investment is the same for both experts, $g_i^* = g_j^*$, and given (3.51), the equilibrium competition strategy of expert $i$ is

$$\frac{v}{2} (\pi(g_i^*)) \alpha_i^2. \hspace{1cm} (3.56)$$

**Proof of Proposition 3.3**

The proof starts the same as Proof of Proposition 3.2.
The expected payoff to expert $i$ for given $\alpha_i, g_i, g_j$ and $r$ is now

$$v \int_0^{\alpha(e_i)} \left[ \pi(g_i) \frac{\alpha_i}{2} \right] d\alpha_j + v \int_0^1 \left[ \pi(g_j) \left( -\frac{\alpha_j}{2} \right) \right] d\alpha_j + \int_0^{\alpha(e_i)} \pi(g_i) r d\alpha_j - e_i - g_i. \tag{3.57}$$

Following the same steps as in Proof of Proposition 2, we derive the equilibrium competition strategy:

$$e(\alpha_i) = v \left( \frac{\pi(g_i)}{4} + \pi(g_i) \alpha_i^2 + \pi(g_i) r \alpha_i \right). \tag{3.58}$$

By substituting (3.58) into (3.57) and integrating with respect to $\alpha_i$ on the interval $[0, 1]$, we obtain the expected payoff to expert $i$ for a given $g_i$ but before observing $\alpha_i$:

$$v \left( \frac{1}{4} \pi(g_i) - \pi(g_j) \right) - g_i, \tag{3.59}$$

which is the same as (3.54). Thus, the experts’ investment strategy is not a function of $r$. Since a reward is costly for the policy maker and does not influence the equilibrium strategies of the experts, the policy maker’s optimal reward is $r^* = 0$.

**Proof of Proposition 3.4**

The policy maker’s sequentially rational policy is then $\frac{1}{2} - \alpha_L$ if she receives a message from expert $L$ and $\alpha_R - \frac{1}{2}$ if she receives a message from expert $R$. If communication is unsuccessful, then her sequentially rational policy is 0.

The expected payoff to expert $i$ for a given $g_i$ and $r$ is

$$v \int_0^1 \left[ \pi(g_i) \left( \alpha_i - \frac{1}{2} \right) \right] d\alpha_i + \pi(g_i) r - g_i = \pi(g_i) r - g_i. \tag{3.60}$$

Then, (3.60) is maximised at $g_i^\dagger$ such that

$$f'(g_i^\dagger) = \frac{1}{r} \tag{3.61}$$

if $f(g_i^\dagger) \leq 1$, and at $\arg_{g_i} f(g_i) = 1$ otherwise. Given the assumptions about $f(g_i)$, $g_i^\dagger$ is positive and increasing in $r$. By $g_i^\dagger(r)$ we now denote the equilibrium investment in communication skills as a function of the reward.
The expected payoff to the policy maker for a given \( r \) is then

\[
\pi(g^*_i(r)) \int_0^1 \int_0^1 \left[ \alpha_R \left( \frac{1}{2} - \alpha_L \right) + \alpha_L \left( \alpha_L - \frac{1}{2} \right) - \frac{(\frac{1}{2} - \alpha_L)^2}{2} \right] \, d\alpha_L \, d\alpha_R - \pi(g^*_i(r)) \, r,
\]

which can be rearranged to

\[
\pi(g^*_i(r)) \left( \frac{1}{24} - r \right).
\]

(3.63)

To show that \( r^* > 0 \), we consider the interior solution for equilibrium investment, i.e. \( \pi(g^*_i(r)) = f(g^*_i(r)) \). Then, the expected payoff to the policy maker for a given \( r \) is

\[
f(g^*_i(r)) \left( \frac{1}{24} - r \right).
\]

(3.64)

By differentiating the expression in (3.64) with respect to \( r \), we obtain the first order condition:

\[
\frac{\partial g^*_i}{\partial r} \frac{\partial f}{\partial g^*_i} \left( \frac{1}{24} - r \right) - \pi(g^*_i(r)) = 0.
\]

(3.65)

It is straightforward to show that (3.65) cannot hold at \( r = 0 \) because at \( r = 0 \) have \( f'(g^*_i) = +\infty \), while \( \frac{\partial g^*_i}{\partial r} \), \( \left( \frac{1}{24} - r \right) \) and \( \pi(g^*_i(r)) \) are positive and finite. Thus, \( r^* > 0 \).

**Proof of Proposition 3.5**

If the experts, compete for access, then the ex-ante expected payoff to the policy maker is given by

\[
\frac{1}{16} \pi(g^*_i),
\]

(3.66)

where \( g^*_i \) is determined by (3.30). Therefore, the ex-ante expected payoff to the policy maker is increasing in \( v \). If the policy maker randomly chooses an expert in advance, then her ex-ante expected payoff is given by

\[
\pi(g^*_i(r^*)) \left( \frac{1}{24} - r^* \right),
\]

(3.67)

where \( g^*_i(r^*) \) is determined by (3.31). Thus, the ex-ante expected payoff to the policy maker is not a function of \( v \). Hence, the ex-ante expected payoff under competition for access is greater than under random choice of an expert only if \( v \) is high enough. For sufficiency, note that a sufficiently high \( v \) guarantees that the ex-ante expected payoff under competition for access is \( \frac{1}{16} \), whereas the ex-ante expected payoff under random choice of an expert cannot be greater than \( \frac{1}{24} \).
Proof of Proposition 3.6

First, each expert’s effort function, \( e(\alpha_i) \), must be strictly increasing in \( \alpha_i \). Second, no expert is willing to report \( m_i < \alpha_i \) and thus reports \( m_i = \alpha_i \). Since \( e(\alpha_i) \) is strictly increasing in \( \alpha_i \), there is a one-to-one mapping between an expert’s effort and his strength of evidence. By observing the efforts, the policy maker effectively learns \( \alpha_L \) and \( \alpha_R \), and chooses a policy \( p^* \) such that \( p^* = \alpha_R - \alpha_L \).

The expected payoff to expert \( i \) for a given \( \alpha_i \) is

\[
v \int_0^{\alpha(e_i)} (\alpha_i - \alpha_j) \, d\alpha_j + v \int_{\alpha(e_i)}^1 (\alpha_j - \alpha_i) \, d\alpha_j - e_i. \tag{3.68}\]

By differentiating the above expression with respect to \( e_i \), we obtain the first order condition:

\[
v \frac{\partial \alpha(e_i)}{\partial e_i} (\alpha_i - \alpha(e_i)) + \int_{\alpha(e_i)}^1 \left[ \frac{\partial \alpha(e_i)}{\partial e_i} \right] \, d\alpha_j - 1 = 0, \tag{3.69}\]

which can be rearranged to

\[
v \frac{\partial \alpha(e_i)}{\partial e_i} (1 - \alpha_i) - 1 = 0. \tag{3.70}\]

In equilibrium, it must be that \( \alpha(e_i) = \alpha_i \). Strict monotonicity implies \( \left( \frac{\partial \alpha(e_i)}{\partial e_i} \right)^{-1} = e'(\alpha_i) \). Therefore, (3.70) simplifies to:

\[e'(\alpha_i) = v (1 - \alpha_i). \tag{3.71}\]

By integrating with respect to \( \alpha_i \) we obtain

\[e(\alpha_i) = v \alpha_i \left( 1 - \frac{1}{2} \alpha_i \right) + C, \tag{3.72}\]

where \( C \) is a constant. If we further assume that \( e(0) = 0 \), then \( C = 0 \). Then, the sequentially rational effort exerted by expert \( i \) upon observing \( \alpha_i \) is given by

\[e(\alpha_i) = v \alpha_i \left( 1 - \frac{1}{2} \alpha_i \right). \tag{3.73}\]

which gives us the equilibrium competition strategy.
Proof of Proposition 3.7

The proof starts the same as Proof of Proposition 3.6. If communication is successful, by observing the efforts, the policy maker effectively learns $\alpha_L$ and $\alpha_R$, and chooses a policy $p^*$ such that $p^* = \alpha_R - \alpha_L$. If communication is unsuccessful, then the policy maker cannot make any inference and chooses a policy of zero.

The expected payoff to expert $i$ for given $g_i$, $g_j$ and $r$ is

$$v \int_0^{\alpha(e_i)} \pi(g_i) \alpha_i d\alpha_j + v \int_{\alpha(e_i)}^1 \pi(g_j) \alpha_j d\alpha_j - e_i - g_i. \quad (3.74)$$

By following the same steps as in Proof of Proposition 3.6, we obtain the equilibrium competition strategy:

$$e(\alpha_i) = v\alpha_i \left(1 - \frac{1}{2}\alpha_i \right) \pi(g_i). \quad (3.75)$$

Given the above communication strategy and competition strategy (with $C = 0$), the expected payoff to expert $i$ for given $\alpha_i$, $g_i$ and $g_j$ is

$$v \int_0^{\alpha(e_i)} \pi(g_i) \alpha_i - \alpha_j d\alpha_j + v \int_{\alpha(e_i)}^1 \pi(g_j) \alpha_j - \alpha_i d\alpha_j - v\alpha_i \left(1 - \frac{1}{2}\alpha_i \right) \pi(g_i) - g_i, \quad (3.76)$$

which can be rearranged to

$$\frac{v}{2} \left(\pi(g_i)\alpha_i^2 + \pi(g_j)\right) - g_i. \quad (3.77)$$

By integrating the expression in (3.77) with respect to $\alpha_i$ on the interval $[0, 1]$, we obtain the expected payoff to expert $i$ for given $g_i$ and $g_j$ but before observing $\alpha_i$:

$$\int_0^1 \left[ \frac{v}{2} \left(\pi(g_i)\alpha_i^2 + \pi(g_j)\right) - g_i \right] d\alpha_i = \frac{v}{6} (\pi(g_i) + \pi(g_j)) - g_i. \quad (3.78)$$

Then, (3.78) is maximised at $g_i^\dagger$ such that

$$f'(g_i^\dagger) = \frac{6}{v} \quad (3.79)$$

if $f(g_i^\dagger) \leq 1$, and at $\arg_{g_i} f(g_i) = 1$ otherwise, which yields the equilibrium investment strategy. Then, given that the equilibrium investment is the same for both experts,
\( g_i^* = g_j^* \), and given (3.75), the equilibrium competition strategy of expert \( i \) is
\[
v \alpha_i \left( 1 - \frac{1}{2} \alpha_i \right) \pi(g_i^*). \tag{3.80}
\]

**Proof of Proposition 3.8**

The proof proceeds similarly to Proof of Proposition 3.7.

The expected payoff to expert \( i \) for given \( \alpha_i, g_i, g_j \) and \( r \) is
\[
v \int_0^{\alpha(e_i)} \pi(g_i) [\alpha_i - \alpha_j] \ d\alpha_j + v \int_{\alpha(e_i)}^1 \pi(g_j) [\alpha_j - \alpha_i] \ d\alpha_j + \int_0^{\alpha(e_i)} \pi(g_i) r \ d\alpha_j - e_i - g_i. \tag{3.81}\]

By following the same steps, we obtain
\[
e(\alpha_i) = v \alpha_i \left( 1 - \frac{1}{2} \alpha_i \right) \pi(g_i) + \pi(g_i) r \alpha_i. \tag{3.82}\]

Substituting (3.82) into (3.81), we obtain the expected payoff to expert \( i \) for given \( g_i \) and \( g_j \) but before observing \( \alpha_i \):
\[
\frac{v}{6} (\pi(g_i) + \pi(g_j)) - g_i, \tag{3.83}\]

which is the same as (3.78). Thus, the experts’ investment strategy is not a function of \( r \). Since a reward is costly for the policy maker and does not influence the equilibrium strategies of the experts, the policy maker’s optimal reward is \( r^* = 0 \).
References


