Essays in Political Economy

by

Federico Trombetta

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy in Economics

Department of Economics

October 2019

A Michela e Beatrice

Contents

List of Figures

List of Tables

Acknowledgments

I am deeply indebted to my supervisors, Ben Lockwood and Helios Herrera, for their constant advice, help and support. They taught me how to think about economic problems and how to approach research in the best possible way.

I thank the Economic and Social Research Council [grant number DTC-ES/J500203/1]] and the Department of Economics for generously funding my studies (and acquiring part of the data used for Chapter 2).

Together with my supervisors, I met many amazing scholars that - through teaching or conversation - were important for the development of this thesis. In particular, I am grateful to Robby Akerlof, Sascha O. Becker, Carlo Beretta, Dan Bernhardt, Peter Buisseret, Costas Cavounidis, Bhaskar Dutta, James Fenske, Larry Iannaccone, Mario Maggioni, Guido Merzoni, Abhinay Muthoo, Motty Perry, Kirill Pogorelskiy, Herakles Polemarchakis, Debraj Ray, Jared Rubin, Francesco Squintani, Stephane Wolton.

I received help and comments in various occasions from a number of friends and colleagues: Zeinab Aboutalebi, Rocco D'Este, Yannick Dupraz, Andy Ferrara, Matteo Gamalerio, Chris Koenig, Giulia Lotti, Erik Melander, Andy Menzel, Martina Miotto, Ayush Pant, Domenico Rossignoli, Tilko Swalve, Giulio Trigilia.

Nothing of this long and tough journey would have been possible without friends. So, thanks Marti, Zeinab, Ayush & Akanksha for travelling with me.

I thank my grandparents, my parents and my sister. And, above all, I thank my family, Michela and Beatrice: they bore the burden of an excessively busy husband and dad. Nothing of what I have done or I will be able to do is remotely comparable with their importance in my life.

Finally, I thank God and I pray that, if there is any talent in me, it can be at His service.

Declarations

I declare that this thesis is my own work. It contains work based on collaborative research in Chapter 3 (with Ayush Pant) and 4 (with Domenico Rossignoli). With respect to Chapter 3, I developed mainly the baseline model, while my co-author developed mainly the sections that follow. With respect to Chapter 4, my contribution is mainly in the historic background and in the identification strategy for the difference in difference. The cleaning and coding of the data is mainly due to my co-author, and the analysis was carried out together.

Chapter 2 develops the dissertation I submitted at the University of Warwick for the Master of Research. The basic set up of the model has changed, as well as most of the analysis. The empirical part is entirely new.

This thesis has not been submitted for a degree in another university.

Chapter 1 contains material published by the Journal of Public Economic Theory as: Trombetta, F. (2020) When the light shines too much: Rational inattention and pandering. *Journal of Public Economic Theory*, 22(1): 98–145. DOI: `10.1111/jpet.12402`. A link to the Creative Commons licence is: `https://creativecommons.org/licenses/by/4.0/`. The content is the same, with the exception of: section and equation numbering; maths and referencing style; few typos; that I am keeping consistent with the rest of the thesis. The paper was entirely written during my PhD. Papers based on chapter 2 and chapter 4 have been submitted as well.

Abstract

Chapter one explores the role of endogenous costly attention allocation in politics. I show that voters may choose to pay too much attention to politicians' actions, and this induces too much political pandering. Moreover, when attention to the action and to the state of the world are both endogenous, voters may not pay enough attention the state. This model can be a demand-driven explanation of the under-provision of analytical contents by news channels.

Chapter two looks at the role of the media market structure in keeping outlets free to publish their findings. Building on the literature on media capture, the model highlights that high competition in the media market can make capture easier. Moreover, it highlights conditions on the parameters where the effect of competition on capture is non-monotonic. The model is motivated by empirical analysis of the digitization of terrestrial television in Europe.

Chapter three shows that it is not necessarily true that competition in the modern digital environment is pushing media outlets towards early release of less accurate information. More competitive environments may be more conducive to reputation building. Therefore, it is possible to have better reporting in a more competitive world. However, we show that the audience may be worse-off due to outlets' better initial information. Finally, we show how a source may exploit the speed-accuracy trade-off.

Chapter four studies the effect of institutions on economic outcomes looking at the historical case of Medieval England. The leader of medieval Benedictine monasteries was elected by the monks. At the same time, monasteries were feudal landlords, alongside secular noblemen. This chapter shows that holdings governed by Benedictine monasteries were more productive than those controlled by secular landlords. We explore a range of potential channels, ruling out those that are not related to the different institutional structure.

Abbreviations

AD: Anno Domini

ASO: Analogue switch off

ATT: Average Treatment Effect on Treated

CFE County Fixed Effect

DDB: Domesday Book

DID: Difference-in-difference

DTT: Digital Terrestrial Television

EAO: European Audiovisual Observatory

EC: European Commission

EU: European Union

FE: Fixed Effects

GDP: Gross Domestic Product

HFE: Hundred Fixed Effects

NGO: Non-Governmental Organization

SATT: Sample Average Treatment Effect on Treated

TV: television

UNESCO: United Nations Educational, Scientific and Cultural Organization

Introduction

If a nation expects to be ignorant and free, in a state of civilisation, it expects what never was and never will be. The functionaries of every government have propensities to command at will the liberty and property of their constituents. There is no safe deposit for these but with the people themselves; nor can they be safe with them without information. Where the press is free and every man able to read, all is safe.

Thomas Jefferson to Charles Yancey, 6 January 1816

This thesis collects together four essays linked by their common belonging to the field of political economy. They all make use of economic tools to answer questions on the edge between politics and economics.

The first three chapters are, generally speaking, about information. They contribute to the frontier of political economy and media economics answering to three different and important questions. Are voters allocating attention to politics efficiently? Is competition in the media market reducing the ability of politicians to influence the content published by news outlets? Is competition reducing the quality of journalism?

The importance of voters' information for an effective democracy has been highlighted by many, including Thomas Jefferson in the opening quote. However, it is worth trying to understand deeply what incentives are at play, so that policies can be implemented taking them into account. The main contribution of those three chapters, in particular, is to highlight conditions where the "wisdom of the crowd" (or of the existing literature) does not work, suggesting some alternative channels.

In particular, chapter one qualifies Jefferson's statement showing that, even when the press is free and everyone is able to read, something may go wrong. More in details, it shows that voters may be allocating a sub-optimally high amount of attention to politicians' actions and a sub-optimally low amount of attention to the context in which those actions are taken. This induces politicians to increase their pandering to the voters' opinion, with negative effect on voters' welfare.

Chapter two looks at the role of the media market structure in keeping outlets free to publish their findings and, as a consequence, Governments accountable to their citizens. In particular, it shows that competition in the media market is not always a good deterrent against media capture, i.e. against the ability of bad politicians to influence the content published by media outlets. As highlighted by the existing literature on media capture, competition increases the number of outlets that the politician has to capture. But there is a trade off induced by the fact that competition may reduce the outside option of captured outlets, making it cheaper to buy them out.

Chapter three goes beyond media freedom to focus on a different aspect: the quality of the information provided and, as a consequence, its impact on people's choices. It shows that increased competition in the media market does not necessarily lead to more "speed driven journalism", i.e. earlier release of less accurate information. On the contrary, it may allow high quality outlets to better signal their ability to check facts, hence increasing their willingness to do so, if reputation matters, despite the increased pre-emption concern. On the other hand, however, better "first rumours" may discourage additional investigation, reducing reader's welfare.

The fourth one has a different focus, but still in the realm of political economy, as it uses the historical case of land ownership in Medieval England to study the role of institutions, and in particular of democratic selection, on economic outcomes, even in absence of (or with very reduced) accountability tools.

Chapter 1

When 24 hour news channels and social media provide constant access to political news, what is the role of voters' attention in shaping political decisions? Are voters paying the right amount of attention to politics? And are they paying attention to the right elements?

At first, it seems obvious that voters should pay as much attention as possible to politics, so that they make better choices and elect better politicians. Moreover, tools that make attention “cheaper” should have an unambiguously positive impact. I show in this chapter that reality is more complex: the decision about how much attention to pay to politics and how to allocate this attention is non trivial, with equilibrium incentives that can motivate voters to pay too much attention to what politicians do and too little attention to what politicians *should* do. This is consistent with the media studies literature that looks at the increasingly important news channels and finds that “80-85 per cent of news stories on the news channels contain no context or analytical content at all”.[1]

Voters’ attention equilibrium allocation may motivate politicians to do what voters want, irrespective of what would be more appropriate, and voters may be paying too much attention to some aspects of the political process and too little to others. Given that these allocations might be suboptimal, policies intended to improve citizen’s attention to politics could be making things worse.

This is the first work in political economy to look at the effects of voters’ rational inattention on pandering in a political agency framework. The model builds on the literature on pandering, under which the politician chooses the action that guarantees his re-election, rather than the optimal one, e.g. Fox (2007), Maskin and Tirole (2004). In that context, this paper endogenizes voters’ level of attention to politics, following the recent and growing literature on the political consequences of rational inattention (Matejka and Tabellini 2017, Prato and Wolton 2016 and 2018). On top of that, it takes into account that attention to politics is a manifold concept. Therefore, this paper endogenizes both the overall amount of attention to politics and the way attention is distributed between actions and the state of the world, studying the interaction between the two.

The model shows that voters’ equilibrium allocation of attention is generically suboptimal. First of all, when only endogenous attention to the action is considered, I show that it increases pandering and that the equilibrium level of attention chosen by the voter is higher than the ex ante optimal level. This is because, when choosing the equilibrium level of attention, the voters are concerned about the selection of

[1] Lewis et al. (2005), p. 470.

good politicians for the next period, and observing the action is useful to that purpose. However, they are not taking into account the fact that higher attention increases the incentive to pander: after all, pandering is effective as long as it is observed by voters. This has a negative effect on both period one voters' welfare and the ability to discriminate between types of politicians.

I then look at how voters allocate attention between action and the state of the world, showing that, in equilibrium, we tend to pay too much attention to what politicians are doing and too little to what they should do. In particular I show that, in equilibrium, there is always under-attention to the state of the world and I find sufficient conditions for over-attention to action. When selection is the only thing that matters, knowing the state of the world on top of the action buys nothing to the voter, in expectation. This is because learning the state of the world is never going to strictly change the voter's re-election strategy, and as a consequence there is no incentive to invest any costly attention there. On the other hand, knowing the action is useful for the voting choice.

From an ex ante perspective, however, attention to the state of the world can be useful in reducing pandering, leading to a higher first period utility, but the equilibrium choice ignores the disciplining effect of this type of attention, because the politician's decision is sunk. Moreover, I show that, since attention to the state reduces only the "bad pandering" (i.e. politicians choosing an action different than the state in order to be re-elected) but not the "good pandering" (i.e. a dissonant politician choosing the voter's preferred action in order to be re-elected), while attention to the action increases both, this model exhibits an interesting complementarity between the two types of attention, that is ignored in the equilibrium choice.

As a result of this inefficient equilibrium allocation politicians have incentives to pander too much, reducing the voters' welfare. Importantly, an exogenous reduction in the total cost of attention is likely to make things worse, overall, because it induces an even higher level of over-attention to action and rewards more pandering, without improving the level of attention to the state.

Chapter 2

A free media has been seen as a powerful guarantor of political accountability, both theoretically (e.g. Besley 2006) and empirically (e.g. Ferraz and Finan 2008, Snyder and Stromberg 2010). However, the media may be powerful enough to determine an electoral outcome and to promote a bad candidate, even when voters are fully rational and "Bayesian" (Prat 2014, Anderson and McLaren 2012, Enikolopov et al. 2011). As a consequence, an incumbent politician may be interested in controlling what media outlets report, to present a positive image to voters and to stay in power. This chapter looks at the effect of competition in influencing the incentives towards media capture, with novel results.

The current literature stresses the positive effect of competition on media freedom: increasing the number of outlets means that the bad politician has more publishers to deal with, so capture is more costly. However, there is a more subtle effect, because the competitive pressure decreases profits, and firms with smaller financial margins may be more willing to sacrifice editorial independence for political money[2]. Hence, they are cheaper to capture.

Overall, the direction of this trade off is not trivial: the "positive" view of the role of competition in deterring capture can be questioned if more competition means smaller (and financially weaker) media outlets, less able to inform the public opinion and to resist to political pressure. Media outlets may be more numerous, but do they have more freedom? Under some conditions, more competition can actually harm the media's independence from political influence, as the empirical section of this paper suggests. This highlights the need for a deeper understanding of the forces behind this trade off.

This chapter makes two contributions: on the empirical side, it is the first to point out the existence of a robust negative relationship between competition in the media market and media freedom from political influence. On the theoretical side, the model provides an explanation for the counter-intuitive empirical results, showing that potential risks to media independence from the political power are high not only when competition

[2] Although in a different context (i.e. advertisement driven media bias), this is consistent with the results of Beattie et al. (2017), where they show that online competition for advertisement increases media bias

is too little, but also when it is too much. Moreover, the model highlights that the standard "positive" result of competition and capture relies on restrictive assumptions about voters' behaviour and media outlets' profits. From a policy perspective, both the empirics and the theory stress the risk, in terms of editorial independence, of excessive competition in the media sector.

More specifically, the empirical section of this chapter suggests that high competition in the mass media market reduces media freedom from political influence. I exploit the staggered digitization of terrestrial television in Europe: a technological change that allowed for a more efficient use of the spectrum and, hence, for the entry of new players in the market. In an event study analysis, this chapter shows that the sign of the relationship between digitization (and hence competition) and the freedom of the media from political influence is negative and strongly significant. And this happens precisely in countries with high level of pre-treatment competition, suggesting that media capture is easier in those places.

The theoretical model is motivated by those findings. It is a natural extension of the seminal contribution by Besley and Prat (2006). It shows that, relaxing some of the assumptions in a fairly natural way, high competition can actually be bad for media freedom, as the cost of media capture is driven to zero as the number of outlets goes to infinity. Moreover, the relationship can be non-monotonic, overall, meaning that media capture occurs when competition is either very weak or very strong. Intuitively, increasing the number of media outlets can make capture overall more expensive for the politician, as long as the number of outlets that need to be silenced increases in the same way, and the politician pays monopoly profits to every captured outlet. But competition has a decreasing effect on the influence that each individual outlet has on voters: basically, with more competition every outlet is able to inform a smaller fraction of the electorate. In fact, even established media outlets have limits in their ability to inform voters, as the recent spreading of fake news highlights. Hence, when there are enough outlets, the politician may be willing to allow for some free media outlets, since they will not be powerful enough to change electoral outcomes. This decreases the profits that captured outlet would make by rejecting the bribe from the politician, hence they are cheaper. The overall effect of this trade off induced by competition (more outlets to be silenced, but each of the is cheaper) depends on the relationship between readership and profits. When it is convex (e.g. because the media market is modelled as a two sided one) then high competition makes capture cheaper, overall.

Chapter 3

In the past decade there has been a growing concern that competition induced by lower entry barriers in the media market due to the Internet is eroding the incentives to conduct investigations (e.g. Cairncross 2019), or even to check the facts properly before publication. Conventional wisdom and relevant contributions in the media studies literature suggest that more competition for the same scoop has driven media outlets towards early release of less accurate information. In this chapter we argue that this is not necessarily the case. Moreover, we highlight a counter-intuitive non-monotonic effect of better rumours on readers' welfare, as more reliable rumours may discourage further investigation from high quality outlets.

We argue that in the face of competition the resolution of speed-accuracy tradeoff, or simply "the newsroom dilemma", is determined by two opposing forces. Competing media outlets not only care about being pre-empted on stories they are working on, but also about their reputation. While the risk of pre-emption pushes them towards early release of less developed stories, i.e., towards speed, reputation concerns may allay such fears pushing towards later release of well-researched stories, i.e., towards accuracy. This motivates us to build a model of a pre-emption game with career-concerned players in a natural setting.

Our model makes two fundamental observations. In a model of media competition and career concerns, we identify conditions in which an equilibrium where the higher quality firm investigates the rumours can be more easily sustained in the presence of competition. This counter-intuitive result is consistent with the surprising findings of Knobel (2018), who points out how the space dedicated to "deep accountability journalism" has increased over time in a sample of US based newspapers. Second, we show that an *improvement* in the quality of an initial rumour may *hurt* the readers. This is so because it makes the aforementioned equilibrium less likely as the high quality firm views lower informational gain from conducting the investigation, thereby making pre-emption concerns more salient. Thus, improving the technology through which initial information is generated may make the final information received by the end consumers worse. Finally, we show how parameters of the model affect the behaviour of a strategic source who owns the rumour and has preferences over speed (i.e. that the news goes out as soon as possible) and accuracy (i.e. that the piece of information is carefully verified).

These observations help us disentangle what effects different types of technological advancements may have on the media market. Improvements that lift barriers to entry and induce more competition need not to lead to reduced consumer welfare. Competing firms, driven by reputational concerns, may fuel more investigative journalism and more fact checking. This incentive, as we discuss later, is greater for higher quality firms when faced with more competition. However, any technology driven improvement in the precision of the initial rumours may make readers worse off.

Chapter 4

This chapter uses the historical case of Benedictine Monasteries in Medieval England to study the effect of institutional differences in non-democratic societies. More in details, monasteries where part of the feudal organisation of the society, playing an important role as landlords in a time where ownership and political power were overlapping (Aston 1958). Quite often, monasteries where controlling not only the land where the monastery was, but also other holdings located in different places. Moreover, this role as big landlords, alongside secular noblemen and bishops, remained salient after the Norman conquest, and the revolutionary reshuffling of land possessions that followed (Knowles 1966, Finn 1963).

Benedictine monasteries were following the “Rule” established by Benedict of Nursia in 529 AD. This vast and comprehensive document was (and still is) supposed to regulate monastic life including, of course, its governance. And it did so in a way that scholars defined “monastic democracy” (Moulin 2016): the abbot was an elective office, elected for life by the community of monks, and he had an obligation to consult with the community on important decisions. However, the ultimate decision making power was in his own hands. This was happening at a time where secular landlords, appointed by the King (or heirs of appointed landlords) had full discretionary power on their own land.

Were those institutional differences having an impact on economic performance? Or, in other words, was "monastic democracy" a better way of governance than hereditary secular landownership? The answer to this question is not obvious, *ex ante*. Different contributions highlight a higher (e.g. Rost 2017, Ekelund et al. 1996, Postan 1973) or a lower (e.g. Heldring et al. 2017) level of efficiency in monastic management of lands, with various potential channels. In our analysis we combine data from the Domesday book, the English Monastic Archive and various sources for geographic and historical controls and we compare the productive capacity of holdings controlled by Benedictine monasteries with holdings controlled by secular landlords. To reduce the concerns related with selection into treatment, we exploit the massive re-shuffling of land ownership that followed the Norman conquest.

We obtain a consistently positive and significant effect of being governed by a Benedictine monastery on economic performance. This may be due to several channels, but we argue that the institutional structure, and in particular the possibility of selecting a leader (even though the lifelong nomination reduces its accountability to the electorate) played an important role. Moreover, we rule out channels that are not related with that. In particular, we control for the historicity of the holdings, the presence of economies of scale (the number of holdings of the same landlord and their geographic dispersion), a range of geographic factors (terrain suitability to pasture and agriculture, altitude, distance to markets and boroughs etc). On top of that, we show that bishops (i.e. ecclesiastic figures, usually well educated, that were not constrained by the Rule) are not statistically different from secular landlords. Moreover, monastic holdings are not statistically different from secular holdings in terms of technology or population; the physical distance to a monastery does not explain our effect; we can exclude holdings closer than 5km to a Monastery, so that we compare only holdings where monks were not working directly: the result holds in this specification as well, that excludes an effect coming from the monks' famous "hard work ethic". Finally, we show that Monastic holdings were better managed not only when compared with Norman holdings (i.e. relatively newly arrived landlords), but also when compared with Anglo-Saxon holdings, and that more ancient monasteries were not better than more recent ones, ruling out the effect of expertise. In our interpretation, these results suggest that unaccountable elected leaders, like the abbots, are able to make better decisions than unaccountable, non elected leaders like secular landlords or bishops.

Chapter 1

When the light shines too much. Rational inattention and pandering

1.1 Introduction

Are voters paying attention to the right elements when judging, for example, the political implications of a wall on the USA-Mexico border? Despite concrete doubts over the effectiveness of the wall, both in terms of reducing illegal immigration and fighting drug cartels,[1] President Donald Trump seems to insist on this political choice. One possible explanation is that it strengthens in voters the idea of a President aligned with America's interests, something that the voters would notice and appreciate. At the same time, maybe, those voters are not paying too much attention to the effectiveness of the wall itself.

When 24 hr news channels and social media provide constant access to political news, what is the role of voters' attention in shaping political decisions? Are voters paying the right amount of attention to politics? And are they paying attention to the right elements? What are the consequences of inefficient allocation of attention?

At first, it seems obvious that voters should be motivated to pay as much attention

[1] See, for example, Driver (2018).

as possible to politics, so that they make better choices and elect better politicians. Moreover, tools that make attention "cheaper" should have an unambiguously positive impact. I show that reality is more complex: the decision about how much attention to pay to politics and how to allocate this attention is non trivial, with equilibrium incentives that can motivate voters to pay too much attention to what politicians do (e.g., building a wall on the USA-Mexico border) and too little attention to what politicians *should* do (e.g., given the way illegal immigration and drug dealing work, is a wall the most effective way to achieve the objective?). This is consistent with the media studies literature that looks at the increasingly important news channels and finds that "80-85 per cent of news stories on the news channels contain no context or analytical content at all".[2]

Voters' attention equilibrium allocation may motivate politicians to do what voters want, irrespective of what would be more appropriate, and voters may be paying too much attention to some aspects of the political process and too little to others. Given that these allocations might be suboptimal, policies intended to improve citizen's attention to politics could be making things worse.

This is the first paper in political economy to look at the effects of voters' rational inattention on pandering in a political agency framework. The model builds on the literature on pandering, under which the politician chooses the action that guarantees his re-election, rather than the optimal one, for example, Fox (2007), Maskin and Tirole (2004).[3] My definition of pandering borrows directly from Besley (2006, ch. 3, footnote 35): pandering occurs when a politician "fails to follow information about the optimal action for fear of its electoral consequences". That is, when the politician chooses the action that gives him higher re-election chances over the one that he favours given the state of the world.[4]

In that context, this paper endogenizes voters' level of attention to politics, following the recent and growing literature on the political consequences of rational inattention (Matejka and Tabellini 2017, Prato and Wolton 2016 and 2018). On top of that, it

[2]Lewis et al. (2005), p. 470.

[3]The idea that political delegation may imply suboptimal choices is also, for example, in Laussel and Van Long (2020), although with a different focus (optimality of ex-ante constraints to politicians' actions).

[4]Pandering can be a good theoretical approximation of the basic essence of populism. This interpretation is not new in economics, for example, Frisell (2009).

takes into account that attention to politics is a manifold concept. Therefore, this paper endogenizes both the overall amount of attention to politics and the way attention is distributed between actions and the state of the world, studying the interaction between the two.

I find novel results based on the fact that the inability to precommit to a certain level of attention, together with the direction of the relationship between different types of attention and the politicians' actions, can create important and unexplored trade-offs.

The model shows that voters' equilibrium allocation of attention is generically sub-optimal. First of all, when only endogenous attention to the action is considered, I show that it increases pandering and that the equilibrium level of attention chosen by the voter is higher than the ex ante optimal level. This is because, when choosing the equilibrium level of attention, the voters are concerned about the selection of good politicians for the next period, and observing the action is useful to that purpose. However, they are not taking into account the fact that higher attention increases the incentive to pander, and this has a negative effect on both period one voters' welfare and the ability to discriminate between types of politicians.

I then look at how voters allocate attention between action and the state of the world, showing that, in equilibrium, we tend to pay too much attention to what politicians are doing and too little to what they should do. In particular I show that, in equilibrium, there is always under-attention to the state of the world and I find sufficient conditions for over-attention to action.

In this model, when selection is the only thing that matters, knowing the state of the world on top of the action buys nothing to the voter, in expectation. This is because learning the state of the world is never going to strictly change the voter's re-election strategy, and as a consequence there is no incentive to invest any costly attention there. In contrast, knowing the action is useful for the voting choice.

From an ex ante perspective, however, attention to the state of the world can be useful in reducing pandering, leading to a higher first period utility (I use also the term "policy welfare" for this part of the voter's objective function), but the equilibrium choice ignores the disciplining effect of this type of attention, because the politician's decision is sunk. Moreover, I show that, since attention to the state reduces only the "bad pandering" (i.e., politicians choosing an action different than the state in order to be re-elected) but not the "good pandering" (i.e., a dissonant politician choosing the voter's preferred action in order to be re-elected),[5] while attention to the action increases both, this model exhibits an interesting complementarity between the two types of attention, that is ignored in the equilibrium choice. From an ex ante perspective, more attention to the action makes attention to the state more desirable: reducing bad pandering is more important when pandering is an issue, that is, when attention to the action is high.

Attention to the action works in the opposite way: it makes pandering more desirable for the politician. The whole point of choosing a sub-optimal action is the electoral reward, which only happens if this action is observed. In equilibrium the voter does not take into account this negative effect[6] and as a consequence she tends to allocate too much attention to the action.[7] As a result of this inefficient equilibrium allocation politicians have incentives to pander too much, reducing the voters' welfare. Importantly, an exogenous reduction in the total cost of attention is likely to make things worse, overall, because it induces an even higher level of over-attention to action and rewards more pandering, without improving the level of attention to the state.

Those results help to explain the relationship between social media and populism. Interestingly, politicians advocating populist policies are over-represented on social media;[8] social media usage is highly correlated with political engagement,[9] political attention and consumption of political news,[10] and journalists are asking whether social media is "empowering populism".[11] In the framework of this paper, social media reduces the

[5] "Good" and "bad" are to be interpreted with respect to policy welfare only. This is different from the total welfare of the voter, that comes from the selection part as well, that is, from the ability to select a good politician in period 2.

[6] While the result of no attention to the state is specific of this model, the equilibrium tendency to under-evaluate the disciplining effect of attention to the state and under-consider the downsides of attention to the action is more general, as shown in Appendix F1.

[7] This is always the case, at least weakly, when the voter has to allocate a certain amount of attention between action and state, as in Appendix D1. When the level of the two attentions is endogenous and the cost function is additively separable there is always under-attention to the state. Over-attention to the action depends on the strength of the ex ante complementarity between the two.

[8] Extreme Tweeting, The Economist, November 19, 2015.

[9] Wihbey (2015) and Barberá (2018) for a review, although evidence is somehow mixed.

[10] Both Mosquera et al. (2019) and Allcott et al. (2019) show that, among the causal effects of Facebook deactivation, there is reduced news consumption, reduced ability to recognize politically-skewed stories, reduced news knowledge and attention to politics.

[11] *"Is social media empowering Dutch populism?"*, Financial Times, March 14, 2017.

cost of paying attention to politics and politicians, generating more attention[12] but also inducing more pandering (which has a negative effect on voters' welfare), that can be seen as a good proxy for populism. The overall welfare effect depends on how damaging pandering is, but there are parameters where a decrease in the cost of attention has negative welfare consequences overall. Hence, social media can lead to too much attention to what politicians do, too little analysis of the state of the world, and too much populism.[13]

This paper contributes to the growing political economy literature on the effects of voters' cognitive biases and limitations on decision making and political outcomes. This is the first to look at the effect of rational inattention on pandering. As such, it is at the intercept of several different branches of the political economy literature.

First, it is related to the literature on pandering and political agency (Canes-Wrone et al. 2001, Maskin and Tirole 2004, Fox 2007, Besley 2006, Ashworth and Shotts 2010, Morelli and van Weelden 2013), from which it borrows the basic structure of the model.

Second, it is related to the growing literature on the consequences of behavioural biases and cognitive limitations on political choice (Ashworth and Bueno de Mesquita 2014, Levy and Razin 2015, Ortoleva and Snowberg 2015, Alesina and Passarelli 2019, Glaeser and Ponzetto 2017, Lockwood 2016, Lockwood and Rockey 2018).

Third, it relates to the small but growing literature on rational inattention (Sims 2003) and its interaction with political economy (Matejka and Tabellini 2017). In particular, Prato and Wolton (2016) and (2018) look at rational inattention in a political agency set up.

Of these, the closest is Prato and Wolton (2016) who highlight the importance of voters' attention choice in determining political behaviour in a model of competition and reforms. This paper is different in three important aspects: first of all, their main focus is on the distinction between interest (exogenous) and attention to the action (endoge-

[12] According to the Pew Research Center, the 2016 US presidential campaign was characterized by very high levels of interest, across parties and age groups (Pew Research Center 2016).

[13] As one of the referees correctly pointed out, social media provides political information but also a lot of entertainment unrelated to politics. Hence, if the effect of social media is to increase the cost of attention to politics, rather than to reduce it, then it may increase welfare. In the logic of this model, it all depends on the direction of the relationship between social media penetration and the cost of attention parameter.

nous), and their voter does not choose "over-attention" as an equilibrium behaviour.[14] Here I show that, in models of pandering, the level of attention to the action can exceed the level that is (ex ante) welfare maximising. Second, this paper endogenizes attention and its allocation between actions and the state of the world, with important additional implications in terms of optimality. Finally, theirs is a model of pre-electoral political competition, while this is the first one to look at rational inattention in the framework of standard principal-agent literature on pandering. A consequence of this are some interesting implications for the relationship between attention and populism that I discuss in Section 1.5.

The mechanics of the paper are similar to Lockwood (2016). Both papers look at consequences of cognitive constraints in a set up of political agency and pandering. However, Lockwood (2016) studies confirmation bias and I study rational inattention.

Finally, attention is of course related to transparency (e.g., Prat 2005; recent work by Wolton 2019 and Andreottola 2018 highlights the potentially welfare depressing effect of "too informative" media). My focus is on the attention choice from the voter's side of the problem, hence highlighting a different and unexplored channel.

The remainder of this paper is as follows: section 1.2 introduces the model in its most generic form. Section 1.3 presents the different results endogenizing attention to the action in 1.3.1 and to both action and state in 1.3.2. Section 1.4 explores inefficiencies in the allocation of attention when information about the state and the action is bundled together. Section 1.5 concludes.

1.2 The Model

1.2.1 Set up

I use the models of pandering from Fox (2007) and Besley (2006). The game is a standard two periods political agency model with two players: an incumbent politician P (he) and a representative voter V (she). There is a binary state of the world $s_t = \{A, B\}$, known ex ante by P but not by V (she can decide to pay attention, learning

[14] When the voter would like to pay too much attention, the candidates respond by choosing similar platforms, hence reducing the scope for attention.

it with some probability), with common prior $Pr(s_t = A) = \frac{1}{2}$.[15] The action space for the politician is binary as well, with $x_t \in \{A, B\}$. The action is observed by V with some probability, which I am going to endogenize using the choice of attention.

The politician can be of two types, Consonant (C) and Dissonant (D); formally, $\theta \in \{C, D\}$ with $Pr(\theta = C) = \pi$. The type is private information of the politician, while the prior is common knowledge. In terms of payoff, every type of politician derives some utility E from being in office. Moreover, the Consonant incumbent gains u_t when he matches the action with the state of the world. Formally, when in office, a type C incumbent gets $u_t + E$ if $x_t = s_t$, E if $x_t \neq s_t$. A dissonant incumbent is biased toward a particular action (I assume it is A without loss of generality), irrespective of the state of the world. Formally, when in office, D gets $u_t + E$ if $x_t = A$ and E if $x_t \neq A$. It is assumed that both types get 0 when out of office and the challenger is drawn from the same distribution of the incumbent.[16]

The part of P's utility defined by u_t is private information of the politician. Ex ante, it is distributed according to a cumulative density function F_t with support $[0, \bar{u}_t]$. F_t is continuous and strictly increasing in the whole interval; its probability density function is f_t.

Assumption 1.1 *F_t is uniform, hence $\mathbb{E}[u_t] = \frac{\bar{u}_t}{2}$.*

The uniformity assumption is for convenience[17] as it simplifies the concavity conditions of the ex ante objective function and it does not require to keep track (and impose conditions on) derivatives of the pdf. Moreover, consistent with Fox (2007) and Besley (2006), it is assumed that $\bar{u}_1 > \delta\left(\frac{\bar{u}_2}{2} + E\right)$. Finally, V gets a utility of 1 if $x_t = s_t$ and 0 otherwise, and there is a common discount factor δ.

So far, the model is unchanged with respect to the standard political agency and pan-

[15] The flat prior guarantees that there is not an *ex ante* popular state of the world. Appendix C1 considers the case of a generic prior probability on the state space finding sufficient conditions for the same results to go through.

[16] Potentially, voters can derive information on the incumbent from the challenger's decision on whether to run or not, as shown, for example, in McCannon and Pruitt (2018) in the specific context of prosecutor elections. However, I abstain from this channel to focus on the information deriving from the voter's attention allocation.

[17] Assumption 1.1 would be more problematic if combined with a time independent upper bound of the support of u_t, since the combination of the two would imply that the only ex ante optimal attention choice is 0,0, as it will be clear from case 1 of the proof of Proposition 1.5. This can be avoided either by dropping the uniformity assumption and fixing a sufficiently small $f_1(0)$ or with a time varying upper bound, and I opted for the latter.

dering literature. I build on this framework by endogenizing the choice of attention. As in Prato and Wolton (2016 and 2018),[18] I model rational inattention as the probability q and β that the voter observes the action of the politician and the state of the world before casting her vote. In particular, $q \in [0,1]$ ($\beta \in [0,1]$) is the amount of attention that the voter pays to the incumbent's action (state of the world) and hence the probability of observing the actual action (state).

Formally the voter does not observe x_t or s_t directly, but rather $\tilde{x}_t$ and $\tilde{s}_t$, and she chooses q and β at a cost $\tau C(q, \beta)$, where $Pr(\tilde{x}_t = x_t) = q$, $Pr(\tilde{x}_t = \emptyset) = 1 - q$, $Pr(\tilde{s}_t = s_t) = \beta$, $Pr(\tilde{s}_t = \emptyset) = 1 - \beta$, $\tau \geq \frac{1}{8}$. In other words, the higher the (costly) attention q (β), the greater the likelihood that the voter will observe the action chosen by the politician (state of the world), and use this information in updating her belief.[19]

The cost of attention function is strictly increasing, convex and additively separable[20] and satisfies the following conditions, useful for ensuring concavity of the maximizations. Formally, $C(q, \beta) = c(q) + c(\beta)$ where $c^{'}(0) = 0$, $c^{'}(1) = 1$, $c^{'}(j) > 0 \; \forall j \in (0, 1]$ and $c^{''}(j) \geq 0 \; \forall j \in [0, 1]$.[21]

1.2.2 Timing

I assume that the voter chooses the attention level after the incumbent's action, but obviously without knowing the action itself.[22] Hence, the timing used to derive the equilibrium attention choice is as follows:

1. Nature chooses θ, u_1, s_1 (private info of P);
2. P chooses x_1;
3. V chooses q and β at a cost $\tau C(q, \beta) \geq 0$;
4. V observes $\tilde{x}_1$, $\tilde{s}_1$ and votes;

[18] They have attention to the action only.

[19] The lower bound on τ ensures that full attention is sufficiently costly so that we can focus mainly on interior solutions, and I do not have to impose too many conditions on the parameters when I compare equilibrium and ex ante optimal attention. However, everything would go through without this lower bound, just with some extra conditions to take into account the corner solutions.

[20] Additive separability is sufficient, but not necessary, for the results in section 1.3.2, and it makes that part tractable.

[21] I impose additional restrictions whenever they are be necessary for the tractability of the problem.

[22] Formally, it is the same as letting them choose at the same time. It basically assumes that the voter cannot commit ex ante to a certain level of attention.

5. Period 1 ends and payoffs are paid;

6. Degenerate Period 2 (there is no draw of a new θ if the incumbent has been re-elected) but no elections;[23]

Payoffs are paid after the elections.

This timing is consistent with the literature (e.g., Prato and Wolton 2016) and with the idea that voters tend to acquire information close to elections.[24] Moreover, it does not require additional assumptions about the existence of commitment devices that the voters could use.[25] Ex ante optimal welfare is calculated assuming that the voter could commit to a certain level of attention, known to the politician. The solution concept I use is the perfect Bayesian Nash equilibrium (PBNE).

1.3 Results

To better understand the effect of different types of attention on the political equilibrium, their interaction and the role of attention allocation, I present the results step by step. In particular, Section 1.3.1 considers the benchmark case where V can decide how much attention to pay to the incumbent's action only, that is, she cannot learn the state. In section 1.3.2 both the total amount of attention and the allocation are endogenous.[26] In all these cases, I compare the equilibrium level of attention with the ex ante optimal one.

1.3.1 Benchmark: attention to actions only

Features of the equilibrium

In this subsection, I assume $\beta = 0$ and only look at the optimal allocation of q. The case of exogenously given $\beta \in [0, 1]$ and endogenous q is considered in Appendix B1.

[23] It is important to assume that a new state of the world is drawn in period 2. Otherwise, having a bad incumbent when the state is A would be different to having a bad incumbent when the state is B.

[24] For example, Google Trends results for "Republican Party" and "Democratic Party" in 2018 in the United States show a big increase in the few weeks before the mid-terms, with a spike on the election day and flat results during the rest of the year (both pre and post elections).

[25] In equilibrium, they would like to be able to precommit to a lower level of attention to the action, hence to acquire less information about something that has already happened and that is now a useful signal in terms of selection.

[26] In Appendix D1 I study the optimality of attention allocation by allowing V to choose attention to the state as well, but with the constraint that the sum of q and β has to be equal to a given constant.

The game is solved by backward induction. Proposition 1.1 summarizes some features of the essentially unique equilibrium that arises, noting that this does not depend on whether precommitment to a certain level of attention is possible.[27]

Proposition 1.1 *In every PBNE of this game, irrespective of the possibility of commitment,*

1. $Pr(re - elect|\tilde{x} = A) = 0$ *and* $Pr(re - elect|\tilde{x} = B) = 1$*;*

2. *C incumbent chooses* $x = B$ *when* $s = B$ *with probability 1 and* $x = B$ *when* $s = A$ *if* $u_1 \leq \delta \left(\frac{\bar{u}_2}{2} + E\right) q$*;*

3. *D incumbent chooses* $x = B$ *in both states when* $u_1 \leq \delta \left(\frac{\bar{u}_2}{2} + E\right) q$*;*

4. $Pr(re - elect|\tilde{x} = \emptyset)$ *can be anything;*

To understand this result, note that the dissonant incumbent is comparatively more likely to choose action A than the consonant one, and as a consequence the voter rewards the choice of action B. This re-election strategy gives incentives to pander, that is, to choose an action that is different from the one the politician prefers in the short term.[28] There are multiple equilibria but they are payoff-equivalent. This set of equilibria is very similar to the one derived in models without rational inattention, but Proposition 1.1 is useful because it leads to the following corollary (whose proof is obvious, hence omitted).

Definition 1 *Define "pandering" as the probability* $\gamma = F_1 \left[\delta \left(\frac{\bar{u}_2}{2} + E\right) q\right]$ *that an incumbent chooses the period 1 action that is suboptimal from his point of view.*

Corollary 1.1 *Irrespective of the availability of commitment, an increase in attention to the action increases the probability of pandering.*

This result is intuitive, since the point of pandering is to guarantee re-election, choosing the "popular" action rather than the right one. Therefore it is crucial that the voter observes this action.

[27] Proofs are in Appendix A1.

[28] In case of a consonant politician, this is always bad for the voters.

Equilibrium attention choice

Define $V_C = \delta$ and $V_D = \delta\frac{1}{2}$ as the continuation payoff from having a C and a D politician in period 2. Moreover, define $\Gamma = \pi V_C + (1-\pi)V_D$ as the expected payoff in period 2 from electing the challenger, while γ is defined as in Corollary 1.1.
V uses the re-election strategy outlined in Proposition 1.1 and solves $max_{q\in[0,1]}W$, where

$$W = q\{[Pr(x = B)](\hat{\pi}_B V_C + (1-\hat{\pi}_B)V_D) + [Pr(x = A)]\Gamma\} + (1-q)\Gamma - \tau c(q) \quad (1.1)$$

and $\hat{\pi}_B = Pr(\theta = C|\tilde{x} = B)$.[29]

The logic behind Equation (1.1) is simple: the voter observes the actual action with probability q. Upon observing action B, she re-elects the incumbent, and her expected utility is given by $\hat{\pi}_B V_C + (1-\hat{\pi}_B)V_D$. She elects the challenger when she observes $\tilde{x} = A$. With probability $1-q$ the voter observes $\emptyset$ and hence she is indifferent between re-electing or not.

From the point of view of V, $\gamma = F_1\left[\delta\left(\frac{\bar{u}_2}{2} + E\right)q^e\right]$, where q^e is the conjectured value of q chosen by the voter. In equilibrium, of course, $q^e = q^*$, where q^* is the chosen level of attention that emerges as an equilibrium outcome in this version of the game. Proposition 1.2 defines this formally.

Proposition 1.2 *The equilibrium level of attention q^* is unique, interior and implicitly defined by*

$$\pi(1-\pi)\delta\left(1 - F_1\left[\delta\left(\frac{\bar{u}_2}{2} + E\right)q^*\right]\right) = 4\tau c'(q^*) \quad (1.2)$$

Note that, if attention is costless (i.e., $\tau = 0$), then $q^* = 1$. This is because, in this case, the voter does not take into account the fact that q increases the probability of pandering. Hence, she is just considering the informative value of attention, and in an attention-for-free world she would choose the highest possible amount of attention.

[29] W includes only future expected payoffs from P's actions. The full welfare function includes also present expected payoffs, but they do not matter for the determination of the equilibrium level of attention, given that they are "sunk" when attention plays its role.

Equilibrium Comparative Statics

In terms of comparative statics:

Corollary 1.2 *The equilibrium attention level is increasing in the uncertainty over the type of incumbent ($var(\theta)$) and decreasing in the office rent (E).*

The trade off here is just between the informational value of q and its cost. Uncertainty over θ increases the value of knowing the chosen action while E, leaving everything else equal, boosts the probability of pandering, making the chosen action a poorer signal of the incumbent's type.

I also consider comparative statics on the equilibrium probability of pandering. Given that in equilibrium $q^e = q^*$, the equilibrium level of pandering is $\gamma^* = F_1\left[\delta\left(\frac{\bar{u}_2}{2} + E\right) q^*\right]$. Corollary 1.3 follows directly from Proposition 1.2.

Corollary 1.3 *The equilibrium level of pandering is increasing in the uncertainty over the incumbent's type and in the office rent.*

The point of Corollary 1.3 is that uncertainty over θ increases the equilibrium level of attention, making pandering more profitable. E has a negative effect on q^*, hence a negative indirect effect on the probability of pandering and a positive direct effect on γ^*, making being in office more attractive, and the second one always dominates.

Ex ante optimal attention

The ex ante expected welfare is

$$\begin{aligned} W^{ex} = & \frac{1}{2}\left[\pi(1-\gamma) + (1-\pi)(1-\gamma)\right] + \frac{1}{2}\left[\pi + (1-\pi)\gamma\right] \\ & + q\{[Pr(x = B)](\hat{\pi}_B V_C + (1-\hat{\pi}_B)V_D) + [Pr(x = A)]\Gamma\} + (1-q)\Gamma - \tau c(q) \end{aligned} \tag{1.3}$$

The second part of the equation above is just W, that is, the "future" part of the voter's welfare, given by the ability to select a good P for period 2. The first part is the present utility, given by period 1 policy choice, that is fixed in the equilibrium maximization but is relevant now. If the state is A both types of incumbent may choose $x_1 = B$

with probability γ, hence V gets 1 in period 1 with probability $(1-\gamma)$. If the state is B, the consonant politician always chooses $x_1 = B$ in equilibrium, and the voter gets a payoff of 1 when the dissonant politician chooses action B in state B (this happens with probability γ).

To find the optimal level of attention, the voter solves solves $max_{q\in[0,1]}W^{ex}$. Lemma 1.1 establishes that W^{ex} is strictly concave in q, therefore it is possible to look directly at the first order conditions.

Lemma 1.1 *When $\beta = 0$, the ex ante voter's welfare function is strictly concave in q.*

Solving for the ex ante optimal level of attention and comparing it with the equilibrium level, I can highlight the first important result of the paper.

Proposition 1.3 *The ex ante optimal choice of attention exists and it is unique. Moreover, it is strictly smaller than the equilibrium choice q^*.*

Intuitively when the voter chooses ex post she over-evaluates q, because she does not take into account the effect of q on welfare-depressing pandering. In particular, the effect of q on ex ante welfare can be decomposed in three parts. A positive effect on selection welfare, due to the fact that observing the action is useful in terms of discriminating between politicians; a negative effect again on selection welfare due to the fact that an increase in pandering makes the bad incumbent more similar to the good one, hence the selection process is more difficult, and finally a negative effect on policy welfare, given by the mismatch between the first period state and the chosen policy.[30] In equilibrium, however, only the first effect is relevant for the voter's choice, hence the over-attention result. In contrast the ex ante welfare calculation takes this into account as well, leading to a lower level of q.

A first consequence is that attention precommitment can help the voter to obtain a better political outcome. The second is a natural question: is it possible that a decrease in the cost of attention (captured by a decrease in τ) is overall welfare depressing when

[30]This effect can be decomposed further distinguishing between bad pandering, that is, both types of incumbent choosing B in state A, and good pandering, that is, the dissonant incumbent choosing B in state B. When β is small the former dominates the latter, hence the negative sign. A more detailed discussion is left to Section 1.3.2.

no precommitment is available? In this set up, it is possible to find conditions on the parameters where a *decrease* in τ *reduces* the welfare of the voter. This has important policy consequences, since it implies that cheaper access to politicians actions is not necessarily better for the voter. Interestingly, the channel through which this happens is pandering, that is, a proxy for populism, boosted by the increased level of attention to what politicians do. This result implies that, if social media is a tool that reduces τ,[31] making political information cheaper, then its diffusion may bring over-attention to politics and populism, reducing the welfare of the voters.

1.3.2 Endogenous attention to action and state

In this section I allow the voter to determine endogenously both types of attention, that is, q and β are chosen at a cost $\tau C(q, \beta)$.[32] Before looking at the allocation of attention, I summarize some generic features of the equilibrium. In particular, defining $r_{\tilde{x},\tilde{s}} = Pr(re - elect|\tilde{x}, \tilde{s})$, $\gamma_s^C = Pr(x = B|s, \theta = C)$ and $\gamma_s^D = Pr(x = B|s, \theta = D)$, all the PBNE of the game can be summarized as follows.

Proposition 1.4 *In every equilibrium with two types of attention:*

1. $\gamma_B^C = 1$, $\gamma_B^D = F_1\left[\delta\left(\frac{\bar{u}_2}{2} + E\right)q\right]$;
2. $\gamma_A^C = \gamma_A^D = F_1\left[\delta\left(\frac{\bar{u}_2}{2} + E\right)q(1 - \beta + \beta(r_{B,A} - r_{A,A}))\right]$.
3. $r_{B,\emptyset} = 1$, $r_{A,\emptyset} = 0$, $r_{B,B} = 1$, $r_{A,B} = 0$. $r_{B,A}$, $r_{A,A}$ *and* $r_{\emptyset,\tilde{s}}$ *are unconstrained because of indifference (or because they are off-path).*

Proposition 1.4 contains several insights. First, there are multiple equilibria, depending on the randomization chosen when the voter observes state A and the action. Second, the effect of q on pandering (i.e., on γ_A^D, γ_B^D and γ_A^C) is the same as before: more

[31]As one of the referees correctly pointed out, social media may shift attention away from politics, providing access to different content and increasing the opportunity cost of political attention. If so, in the logic of this model, it may increase the welfare of the voter. A detailed theoretical analysis of the direction of this effect is left for further research, but it is worth mentioning that two very recent empirical contributions (Mosquera et al. 2019 and Allcott et al. 2019) are consistent with social media reducing the cost of attention to politics. Both papers look at the causal effect of Facebook de-activation finding, among other things, that it reduces news knowledge and attention to politics.

[32]Appendix D1 explores what happens when the voter has to allocate a given amount of attention between action and state. The result is again equilibrium (weak) over-attention to the action and (weak) under-attention to the state of the world.

attention to the action makes pandering more desirable, for the politician. Finally, β has either no effect or a negative effect on the equilibrium level of pandering. Irrespective of the values taken by $r_{B,A}$ and $r_{A,A}$, $\frac{\partial \gamma_A^C}{\partial \beta} = \frac{\partial \gamma_A^D}{\partial \beta} \leq 0$. Note that β reduces the bad pandering in terms of period 1 welfare (i.e., politicians choosing an action that is suboptimal from the point of view of the voter, but that increases re-election chances) without affecting the good pandering (i.e., the dissonant incumbent choosing the right action in period 1).

To understand Proposition 1.4 it is useful to start from the fact that, in state A, the incentives of the two types of incumbent are perfectly aligned, hence they behave in the same way. Given this, when the voter learns that the state is A she becomes indifferent between keeping or not the incumbent, and as a consequence $r_{B,A}$ and $r_{A,A}$ are unconstrained. Looking back at the incentives for the politician, they compare their re-election chances of choosing action B vis-a-vis A depending on what they expect the voter to observe. When s is unobserved by the voter, the difference in re-election chances is 1 ($r_{B,\emptyset} = 1$ and $r_{A,\emptyset} = 0$), and it becomes $r_{B,A} - r_{A,A}$ when s is observed by V. Hence, unless $r_{B,A} - r_{A,A} = 1$, a higher conjectured β reduces the electoral reward of choosing B in state A. In state B, instead, observing or not s does not change the re-election strategy of the voter, because the dissonant incumbent has a stronger incentive to choose action A than the consonant one. Hence, the conjectured level of β does not enter in γ_B^C and γ_B^D.

For tractability[33] I make the following assumption:

Assumption 1.2 *When indifferent or off path the voter re-elects the incumbent with the same probability in every information set, hence,* $r_{B,A} = r_{A,A}$.

With two types of attention, the equilibrium objective function maximized by the voter

[33] Assumption 1.2 is useful in avoiding kinks in γ_A^D and γ_A^C (they could become 0 for sufficiently high β if I assume a different randomization), so that I can use calculus everywhere in the ex ante maximization. It does not change the overall negative effect of β on those probabilities, but it is important to point out that, in the extreme case where $r_{B,A} = 1$ and $r_{A,A} = 0$, attention to the state simply disappears from the equilibrium. In this case, then, setting $\beta = 0$ would be optimal not only in equilibrium but also in the ex ante optimal allocation, and the rest of the results would be analogous to section 1.3.1. Finally, in terms of off path beliefs, it assumes that the off path posterior is equal to the prior. Note that this makes intuitive sense because the only off path case is, sometimes, $x = B, s = A$, and both types have exactly the same incentives to choose that action (hence I cannot use D1 to refine those beliefs).

is

$$\begin{aligned}
\tilde{W} &= q\beta\{Pr(x=B,s=B)(\hat{\pi}_{B,B}V_C + (1-\hat{\pi}_{B,B})V_D) + Pr(x=A,s=B)\Gamma \\
&+ Pr(x=A,s=A)(r_{A,A}(\hat{\pi}_{A,A}V_C + (1-\hat{\pi}_{A,A})V_D) + (1-r_{A,A})\Gamma) \\
&+ Pr(x=B,s=A)(r_{B,A}(\hat{\pi}_{B,A}V_C + (1-\hat{\pi}_{B,A})V_D) + (1-r_{B,A})\Gamma) \\
&+ q(1-\beta)\{Pr(x=B)(\hat{\pi}_{B,\emptyset}V_C + (1-\hat{\pi}_{B,\emptyset})V_D) + Pr(x=A)\Gamma\} \\
&+ (1-q)\Gamma - \tau C(q,\beta)
\end{aligned}$$

Substituting the equilibrium re-election probabilities of Proposition 1.4, I obtain

$$\begin{aligned}
\tilde{W} &= q\beta\{Pr(x=B,s=B)(\hat{\pi}_{B,B}V_C + (1-\hat{\pi}_{B,B})V_D) \\
&+ (1-Pr(x=B,s=B))\Gamma\} \\
&+ q(1-\beta)\{Pr(x=B)(\hat{\pi}_{B,\emptyset}V_C + (1-\hat{\pi}_{B,\emptyset})V_D) + Pr(x=A)\Gamma\} \\
&+ (1-q)\Gamma - \tau C(q,\beta)
\end{aligned}$$

The first part of $\tilde{W}$ is the expected second period utility, conditional on observing action and state, multiplied by the probability that this happens ($q\beta$). In this case, the voter definitely re-elects when $x = B$ and $s = B$, and gets an expected payoff of Γ in all other cases (either because she chooses a new politician or because she is indifferent after observing x and s). The second part is the expected second period utility conditional on observing only the action (V re-elects only if the action is B), the third part is the expected utility conditional on observing no action, both weighted by the respective probability. Finally, the last part is the cost of attention.

The next lemma further simplifies $\tilde{W}$, showing that, from an equilibrium perspective, knowing the state is useless for the voter.

Lemma 1.2 *In equilibrium,* $Pr(x=B,s=B)(\hat{\pi}_{B,B}V_C + (1-\hat{\pi}_{B,B})V_D) + (1-Pr(x=B,s=B))\Gamma = Pr(x=B)(\hat{\pi}_{B,\emptyset}V_C + (1-\hat{\pi}_{B,\emptyset})V_D) + Pr(x=A)\Gamma$.

Lemma 1.2 is a consequence of the fact that observing the state on top of the action does not change the voter's re-election choice, hence it does not provide useful information for the decision.[34] Using lemma 1.2 I can rewrite $\tilde{W}$ as

$$\tilde{W} = q\{Pr(x = B)(\hat{\pi}_{B,\emptyset}V_C + (1 - \hat{\pi}_{B,\emptyset})V_D) + Pr(x = A)\Gamma\} + \Gamma(1 - q) - \tau C(q, \beta) \quad (1.4)$$

Equation (1.4) makes it clear that attention to the state is not going to play any role in the positive part of the objective function, and it enters only as a cost. Applying Bayes' rule and substituting the equilibrium strategies in Equation (1.4), we have

$$\tilde{W} = \frac{1}{4}\pi(1 - \pi)\delta\left(1 - \gamma_B^D\right) q + \Gamma - \tau C(q, \beta)$$

As the ex ante maximization problem is quite complex, it is necessary to impose additional conditions to guarantee the concavity of the ex ante objective function. I make some simplifying assumptions, and I discuss ways to relax them in Appendix E1. I assume the following:

Assumption 1.3 $C(q, \beta) = q^2 + \beta^2$.

This assumption makes a closed form solution possible, and highlights the conditions for concavity of the ex ante maximization (it is not an issue in the equilibrium one).

Assumption 1.4 $4\tau \geq \delta\frac{1}{\bar{u}_1}\left(\frac{\bar{u}_2}{2} + E\right)$

Assumption 1.4 is sufficient to ensure that the Hessian matrix of the ex ante maximization problem is negative semidefinite. However, as shown in Appendix E1, it is possible to ensure enough concavity even when it is violated. For simplicity, I define $K = \frac{1}{\bar{u}_1}\left(\frac{\bar{u}_2}{2} + E\right)$, [35] noticing that δK is the ex ante probability that the first period utility of P from doing his favourite action is smaller than the discounted expected utility of being in office in period 2. The following lemma describes the equilibrium attention choice.

Lemma 1.3 *The allocation of attention in equilibrium is unique and is given by* $\beta^{**} = 0$ *and* $q^{**} = \frac{\pi(1-\pi)\delta}{8\tau + \pi(1-\pi)\delta^2 K}$.

[34] I thank one referee for helping me clarifying the mechanism.

[35] By construction $\delta K < 1$.

The particular closed form of q^{**} is of course a feature of the assumptions. In general it is the value of q that solves $\frac{1}{4}\pi(1-\pi)\left[1-F_1\left(\delta\left(\frac{\bar{u}_2}{2}+E\right)q\right)\right]=\tau c'(q)$. As expected, the equilibrium level of attention to the state is zero and all the attention is put on the action of the politician.

In the ex ante case the objective function is

$$\tilde{W}^{ex}=\frac{1}{2}\left[1-\tilde{\gamma}_A+\pi+(1-\pi)\gamma_B^D\right]+\frac{1}{4}\pi(1-\pi)\delta\left(1-\gamma_B^D\right)q+\Gamma-\tau C(q,\beta) \quad (1.5)$$

where $\tilde{\gamma}_A=\gamma_A^C=\gamma_A^D$. Now $\tilde{\gamma}_A$ and γ_B^D are endogenous in the maximization. $\frac{1}{2}\left[1-\tilde{\gamma}_A+\pi+(1-\pi)\gamma_B^D\right]$ is period 1 expected utility, taking into account that in state A both types of incumbent choose the "wrong" (from the voter's period 1 perspective) action with probability $\tilde{\gamma}_A$ and if the state is B the consonant type always chooses $x=B$, while the dissonant type does so with probability γ_B^D. The second part of Equation (1.5) is the selection welfare, coming from the ability to choose a better P in period 2.

Note that q has a negative indirect effect on (1.5) through $\tilde{\gamma}_A$, that is, the probability that both types choose B when the state is A, and through the γ_B^D that enters in the selection welfare, because the dissonant type choosing B when the state is B very often makes it harder to distinguish him from a consonant type. Moreover, it has a positive indirect effect through the γ_B^D that enters in the policy welfare: the good pandering of the dissonant type behaving better in period 1. β instead has just one positive effect: it reduces the period one bad pandering. This effect is what drives the possibility of a positive amount of attention to the state in the ex ante optimal allocation.

As anticipated,

Lemma 1.4 *If Assumptions 1.3 and 1.4 are true, then the Hessian matrix of the ex ante maximization is negative semidefinite.*

Thanks to Lemma 1.4, the ex ante optimal allocation of attention is given by the system of the two partial derivatives, looking at conditions for interior solutions. The main results are as follows:

Proposition 1.5 *Comparing the ex ante optimal and the equilibrium level of attention:*

1. *The ex ante optimal level of attention to the state is always weakly higher than the equilibrium. If* $1 - \pi - 2K > 0$ *then the ex ante optimal* β *is strictly higher than the equilibrium;*

2. *If* $2\tau\delta^2\pi(1-\pi)^2 > \delta^2 K(1-\pi)(1-4\tau\pi) - 32\tau^2$ *then the ex ante optimal level of attention to the action is strictly smaller than the equilibrium;*

Proposition 1.5 makes a very straightforward point about the optimal β: as long as parameters do not produce a corner solution where both q and β are zero (note that if the ex ante optimal q is zero then the ex ante optimal β is zero as well), then the equilibrium level of attention to the state is too small. The situation would not improve with a reduction in τ, as the allocation of attention would still be unbalanced toward attention to the action. To understand the proposition, it is worth pointing out that, from an ex ante perspective, β plays an important role, despite being useless in terms of selection. In particular, it decreases $\tilde{\gamma}_A$, that is, the probability that the incumbent (irrespective of his type) chooses action B in state A. As a consequence, unless the ex ante optimum is in 0,0, it is useful to pay a positive amount of attention to the state of the world. This is ignored in equilibrium, where the politician's choices are taken as given. Importantly, β has no effect on selection welfare, in this set up.

The ex ante role of attention to the action is more complicated. In terms of future welfare (i.e., what is captured by the $\tilde{W}$ part of Equation (1.5)), it is of course directly useful for selection (as captured by the equilibrium optimization) but it also increases γ_B^D, and this is bad for the voter's expected utility from selection, as it makes the dissonant type's behaviour more similar to the behaviour of the consonant type. As a consequence, the indirect effect of q on selection welfare is always negative, hence the the commitment problem exists also in terms of pure selection effect.[36]

Looking now at the policy welfare, if $\beta = 0$, attention to the state has a decreasing effect on it as well and, as long as $\beta < \pi$, this is true in general.[37] However, from an ex ante perspective, the two types of attention are complements: a higher β makes q more desirable. If $\beta > \pi$, the derivative of present welfare with respect to q is positive.[38]

[36] I thank one of the referee for pointing out this interesting feature.

[37] To see this, note that $\frac{\partial}{\partial q}\left[1 - \tilde{\gamma}_A + \pi + (1-\pi)\gamma_B^D\right] = \delta K(\beta - \pi)$. Intuitively, in terms of policy welfare attention to the state increases both bad and good pandering. However, when β is sufficiently low, the (negative) effect of the bad pandering is stronger, as both types may choose it.

[38] Clearly, this holds with additively separable cost.

In this case the bad pandering is reduced by the high β, and as a consequence the positive effect of q on policy welfare due to good pandering dominates the negative effect. The ex ante optimal q may be higher or lower than the equilibrium, depending on whether the complementarity effect is sufficiently strong to overcome the negative effect of higher bad pandering.

Finally, note that obviously the complementarity works for β as well. The higher the ex ante optimal q, the higher the ex ante optimal β, as its welfare increasing role crucially depends on the probability that the action is observed, and hence that pandering is an issue.[39] But since this complementarity plays a role in period 1 welfare only, it is not taken into account by the voter when choosing the equilibrium level of attention.

In terms of parameters, the condition on K being sufficiently small is what determines whether the ex ante optimal q is strictly positive or not.[40] The second condition is just a comparison between the ex ante optimal q and the equilibrium q, taking into account that a higher ex ante optimal β can have cross effects on the optimal q, hence it is not always smaller than the equilibrium q. However, the condition for equilibrium over-attention to the action is easily met: $4\tau\pi > 1$ would be sufficient.

As explored in Appendix F1, the result of equilibrium under-attention to the state of the world and over-attention to the action of the politician is not a feature of this specific political agency model, at least when the cross effects are constrained by the fixed amount of attention to be allocated (or more generally when they are not too big). What is driven by this model is the $\beta^{**} = 0$ result, as it relies on the fact that different types of politician have aligned interests in one of the two states, and hence learning s on top of x does not change the re-election decision of the voter. In Appendix F1, using a model loosely based on Prat (2005) and Canes-Wrone et al. (2001), I show that there is an equilibrium with nonzero attention to the state, but this level is still smaller than what is optimal ex ante.

[39] To see this, note that $\tilde{\gamma}_A = q\delta K(1-\beta)$.

[40] Its effect on the pandering probability is linear and multiplied by K because of the uniform assumption on F_1. If the slope is very steep V prefers to commit ex ante to zero attention at all, as this drives pandering to zero as well. Note that, if $q = 0$, it is pointless to invest in β.

1.3.3 Discussion and suggestive evidence

My results are consistent with the in-depth content analysis of 24-hr news channels in Lewis et al. (2005). They show that "the news channels do not use the enormous time available to them to provide the viewer with a deeper understanding of the news world".[41] One of their case studies is particularly important in terms of attention to the action vs to the state of the world: "all three news channels covered the Government's Defence Review on 21 July, and yet none did so in the context of a broader analysis of the changing rationale behind military spending, or even by clarifying what that rationale might be."[42] Consistent with the point I make here, all the attention was focused on the Government's action (the Defence Review) and none on the state of the world, that is, the "rationale behind military spending". Lewis et al. (2005) find an effect on the supply side, but it may be driven by the demand side allocation of attention by voters, which follows the inefficiencies highlighted in the present paper.

1.4 Extension: Bundling Action and State

So far information about the action of the politician and the state has been considered unbundled. However this is not always the case: it could be that information about what the politician has done and about whether it was correct or not come together,[43] so it is interesting to analyse the implications for the optimal allocation of attention.

The set up is the same as in section 1.2, but I model attention differently. In particular, the voter does not observe x_t or s_t directly, but instead he observes a signal ζ where $Pr(\zeta = x_t, s_t) = \eta$, $Pr(\zeta = \emptyset) = 1 - \eta$. η is chosen by V at a cost $\tau C(\eta)$, where the assumptions on $C(\eta)$ are the same as in section 1.2.

Reminding that r_ζ is the probability of re-election conditional on the realization of ζ, $\gamma_s^D = Pr(x = B|\theta = D, s)$, $\gamma_s^C = Pr(x = B|\theta = C, s)$ and $K = \frac{1}{\bar{u}_1}\left(\frac{\bar{u}_2}{2} + E\right)$, I describe the equilibrium of this version of the game in Lemma 1.5.

Lemma 1.5 *In every equilibrium of this game:*

[41] Lewis et al. (2005), p. 473

[42] Lewis et al. (2005), p. 471.

[43] I thank one referee for pointing this out.

1. $\gamma_B^C = 1$, $\gamma_B^D = F_1\left[\delta\eta\left(\frac{\bar{u}_2}{2} + E\right)\right]$;

2. $\gamma_A^C = \gamma_A^D = F_1\left[\delta\eta\left(\frac{\bar{u}_2}{2} + E\right)(r_{B,A} - r_{A,A})\right]$;

3. $r_{B,B} = 1$, $r_{A,B} = 0$, $r_{B,A}$, $r_{A,A}$ *and* $r_{\emptyset}$ *are unconstrained because of indifference or off path.*

Note that Assumption 1.2 plays a more important role in this extension. In fact, it implies that $\gamma_A^C = \gamma_A^D = 0$, hence attention to politics increases the good pandering only, that is, the dissonant politician choosing B in state B, without increasing the bad pandering, that is, γ_A^C and γ_A^D. This would still be true under the assumption that $r_{A,A} > r_{B,A}$, but if $r_{B,A} > r_{A,A}$ then η would increase the bad pandering as well, with implications for the ex ante welfare calculation. However, as shown in Lemma A1.1 in Appendix A1, the result is unchanged unless I assume fairly ad-hoc randomizations.[44]

Given Lemma 1.5, V chooses η in order to maximize W_b, where

$$\begin{aligned} W_b =& \eta\left(Pr(x=B, s=B)(\hat{\pi}_{B,B}V_C + (1-\hat{\pi}_{B,B})V_D) + (1 - Pr(x=B, s=B))\Gamma\right) + \\ &+ (1-\eta)\Gamma - \tau C(\eta) \end{aligned} \tag{1.6}$$

The solution is summarized by Lemma 1.6.

Lemma 1.6 *The equilibrium attention level is interior, unique and given by the solution η^* of*

$$\frac{1}{4}\pi(1-\pi)\delta(1-\delta K\eta^*) = \tau C^{'}(\eta^*)$$

In this set up attention only increases good pandering. In terms of ex ante welfare, a high γ_B^D is good for discipline and bad for selection but overall the first effect prevails, as stated in Proposition 1.6.

Proposition 1.6 *The ex ante optimal level of attention, η^{**}, is strictly larger than the equilibrium level.*

[44] Lemma A1.1 in Appendix A1 relaxes Assumption 1.2 showing that the result of this section, that is, equilibrium under-attention, does not change as long as the randomization is such that $r_{B,A} - r_{A,A} < (1-\pi)\left(1 - \frac{1}{2}\pi\delta\right)$.

As a result, attention to bundled political information is too small with respect to the ex ante optimal level, which is consistent with the under-attention to the state of the world derived in Section 1.3.

1.5 Conclusion

Are voters allocating attention to politics optimally? I endogenize the attention choice in an otherwise standard model of pandering, highlighting two points of inefficiency. The equilibrium level of attention to politicians' actions can be too high, making pandering too attractive. Conversely, attention to the state (i.e., to what politicians should do) is too low as its effect on selection is limited, even though it could improve the incumbent's behaviour.

The results suggest that trying to increase attention to the state by reducing the cost of attention to politics is likely to make things worse, as it does not change the allocation of attention. It would just make attention to the action even more extreme. Moreover, results highlight the importance of accounting for allocation of attention and its general equilibrium effect on political behaviour. In terms of applications, if populism is a form of pandering and social media is a technological innovation that reduces the cost of attention, this paper gives a rationale for the observed correlation between populism and social media. Moreover, it highlights a general tendency to pay too much attention to what politicians are doing and too little to what they should do.

Looking at the relationship between this paper and Prato and Wolton (2016), it is worth mentioning that attention can be bad for the voter in both frameworks, but the implications of the relationship between equilibrium attention and populism are different: the correlation is positive in this paper, and negative in Prato and Wolton (2016). It would be interesting to explore these results empirically.[45] One way would be to look at the response to exogenous shocks to the cost of attention (e.g., through better mobile phone signals, access to internet where there was none or reduced/increased access to social media) by voters and politicians. Are the voters paying more attention to politics? Are more politicians choosing more populist platforms?

[45] I thank one of the referees for pointing this out.

The current very tractable set up is open to different extensions, suggesting many directions for future research. One would be a model with a multidimensional state of the world. Dimensions that are more salient for voters are likely to invite more pandering from the politician, something that the optimal attention allocation should take into account.

Another option is to microfound voters' attention in a different way, for example with a model of media entry where each outlet can decide whether to provide political news, sports or documentaries. If competition increases the provision of entertainment (where it is easier to differentiate the product), this would shift voters' attention away from hard news (or, equivalently, makes attention to politics more costly). The overall effect of this is ambiguous, as the model highlights: it reduces the politician's incentives to pander, but at the same time it worsens political selection.

Appendices

A1 Proofs

Proof of Proposition 1.1.

In terms of existence, sequential rationality requires that $\hat{\pi}_A \leq \pi$ and $\hat{\pi}_B \geq \pi$, where $\hat{\pi}_{\tilde{x}}$ is the posterior probability of the incumbent being consonant when the voter observes $\tilde{x}$.

Define $r_{\tilde{x}} = Pr(re{-}elect|\tilde{x})$, $\gamma_s^C = Pr(C\ plays\ B\ in\ state\ s)$ and $\gamma_s^D = Pr(D\ plays\ B\ in\ state\ s)$. This equilibrium implies that, from the point of view of V, $\gamma_A^C = \gamma_A^D = \gamma_B^D = F_1[\delta\left(\frac{\bar{u}_2}{2} + E\right)q] := \gamma < 1$ and $\gamma_B^C = 1$.
By Bayes rule,

$$\hat{\pi}_A = \frac{0.5(1-\gamma)\pi}{0.5(1-\gamma)\pi + (1-\gamma)(1-\pi)} < \pi$$

$$\hat{\pi}_B = \frac{(0.5\gamma + 0.5)\pi}{(0.5\gamma + 0.5)\pi + \gamma(1-\pi)} > \pi$$

hence, from the point of view of V, it is sequentially rational to re-elect after observing B and to vote for the challenger after observing A.

Given this, one needs to check optimality from the point of view of P. Starting from the D type, and given the voter's re-election rules stated above, the expected utility of the dissonant incumbent when he chooses action A, irrespective of the state of the world, is

$$\mathbb{E}U_D(x = A) = u_1 + E + \delta\left[qr_A\left(\frac{\bar{u}_2}{2} + E\right) + (1-q)r_\emptyset\left(\frac{\bar{u}_2}{2} + E\right)\right]$$

while

$$\mathbb{E}U_D(x = B) = E + \delta\left[qr_B\left(\frac{\bar{u}_2}{2} + E\right) + (1-q)r_\emptyset\left(\frac{\bar{u}_2}{2} + E\right)\right]$$

Hence, given the fact that in equilibrium $r_A = 0$ and $r_B = 1$ the D incumbent chooses action B, irrespective of the state, when $u_1 \leq \delta\left(\frac{\bar{u}_2}{2} + E\right)q$.

Moving now to the C incumbent, note first of all that, when $s = B$, all the incentives are aligned and hence he always chooses action B. When $s = A$, instead, the expected utilities are as follows:

$$\mathbb{E}U_C(x = A, s = A) = u_1 + E + \delta \left[qr_A \left(\frac{\bar{u}_2}{2} + E\right) + (1-q)r_\emptyset \left(\frac{\bar{u}_2}{2} + E\right)\right]$$

and

$$\mathbb{E}U_D(x = B, s = A) = E + \delta \left[qr_B \left(\frac{\bar{u}_2}{2} + E\right) + (1-q)r_\emptyset \left(\frac{\bar{u}_2}{2} + E\right)\right]$$

and, again, it is optimal for the C incumbent to choose action B in state A when $u_1 \leq \delta \left(\frac{\bar{u}_2}{2} + E\right) q$. This completes the existence proof. In terms of uniqueness, note that there are actually multiple equilibria, given that $\hat{\pi}_\emptyset = \pi$, hence V is indifferent in that case. However, $r_\emptyset$ does not affect P's equilibrium strategies, hence all the equilibria are identical in terms of outcomes (and strategies, with the obvious exception of $r_\emptyset$).

What needs to be shown, now, is that there are no other PBNE with different re-election probabilities after observing A or B or different strategies for P. By contradiction, suppose there is a PBNE where $r_A > 0$ and $r_B < 1$. This is sequentially rational iff $\hat{\pi}(\tilde{x} = A) \geq \pi \Rightarrow \gamma_A^D + \gamma_B^D \geq \gamma_A^C + \gamma_B^C$ and $\hat{\pi}(\tilde{x} = B) \leq \pi \Rightarrow \gamma_A^D + \gamma_B^D \geq \gamma_A^C + \gamma_B^C$.
Note that $\gamma_A^C = F_1[q\delta \left(\frac{\bar{u}_2}{2} + E\right)(r_B - r_A)]$, $\gamma_B^C = 1 - F_1[-q\delta \left(\frac{\bar{u}_2}{2} + E\right)(r_B - r_A)]$, $\gamma_A^D = \gamma_B^D = F_1[q\delta \left(\frac{\bar{u}_2}{2} + E\right)(r_B - r_A)]$.
Hence, $\gamma_A^D + \gamma_B^D \geq \gamma_A^C + \gamma_B^C$ is equivalent to $F_1[q\delta \left(\frac{\bar{u}_2}{2} + E\right)(r_B - r_A)] + 1 - F_1[-q\delta \left(\frac{\bar{u}_2}{2} + E\right)(r_B - r_A)] \leq 2F_1[q\delta \left(\frac{\bar{u}_2}{2} + E\right)(r_B - r_A)]$.
This simplifies to $F_1[q\delta \left(\frac{\bar{u}_2}{2} + E\right)(r_B - r_A)] \geq 1 - F_1[-q\delta \left(\frac{\bar{u}_2}{2} + E\right)(r_B - r_A)]$.
Note that, if $r_B \geq r_A$, $F_1[-q\delta \left(\frac{\bar{u}_2}{2} + E\right)(r_B - r_A)] = 0$ and hence the equilibrium exists iff $F_1[q\delta \left(\frac{\bar{u}_2}{2} + E\right)(r_B - r_A)] \geq 1$, which is impossible because F is strictly increasing, $F_1[\bar{u}_1] = 1$ and $\bar{u}_1 > \delta \left(\frac{\bar{u}_2}{2} + E\right) > \delta \left(\frac{\bar{u}_2}{2} + E\right) q(r_B - r_A)$.
If instead $r_A > r_B$, $F_1[q\delta \left(\frac{\bar{u}_2}{2} + E\right)(r_B - r_A)] = 0$ hence the equilibrium exists iff $F_1[q\delta \left(\frac{\bar{u}_2}{2} + E\right)(r_A - r_B)] \geq 1$. Again, this is impossible because $\bar{u}_1 > \delta \left(\frac{\bar{u}_2}{2} + E\right) > \delta \left(\frac{\bar{u}_2}{2} + E\right) q(r_A - r_B)$.
Hence, there are no equilibria where $r_A > 0$ and $r_B < 1$. Applying the same logic, it is possible to also rule out equilibria where $r_A = 0$ and $r_B < 1$ and where $r_A > 0$ and $r_B = 1$.
Hence, the sole re-election strategy in a SPNE of this game is $r_A = 0$ and $r_B = 1$ and as a consequence the sole equilibrium strategy for the incumbent is the one described

in Proposition 1.1. ■

Proof of Proposition 1.2.

As stated in the main body of the paper, the part of the welfare function relevant for the voter's maximization is $W = q\{[Pr(x = B)](\hat{\pi}_B V_C + (1 - \hat{\pi}_B)V_D) + [Pr(x = A)]\Gamma\} + (1 - q)\Gamma - \tau C(q)$. It follows from the law of iterated expectations: with probability q the action is observed. If it is B, then the incumbent (consonant with probability $\hat{\pi}_B$) is confirmed and $\hat{\pi}_B V_C + (1 - \hat{\pi}_B)V_D$ is the second period expected utility. If $x = A$ or if $\tilde{x} = \emptyset$ the incumbent is replaced by a challenger, consonant with probability π. As a consequence, the expected period 2 utility is given by Γ. Note that

$$
\begin{aligned}
W &= q\{Pr(x = B|\theta = C)\pi V_C + Pr(x = B|\theta = D)(1 - \pi)V_D - \Gamma(1 - Pr(x = A))\} + \Gamma - \tau c(q) \\
&= q\left\{\left[\frac{1}{2} + \frac{1}{2}\gamma_A^C\right]\pi V_C + \left[\frac{1}{2}\gamma_A^D + \frac{1}{2}\gamma_B^D\right](1 - \pi)V_D - \Gamma(Pr(x = B))\right\} + \Gamma - \tau c(q) \\
&= q\frac{1}{2}\left[1 + \gamma_A^C - \pi\left(1 + \gamma_A^C\right) - (1 - \pi)\left(\gamma_A^D + \gamma_B^D\right)\right]\pi V_C + \\
&+ q\frac{1}{2}\left[\gamma_B^D + \gamma_A^D - \pi\left(1 + \gamma_A^C\right) - (1 - \pi)\left(\gamma_A^D + \gamma_B^D\right)\right](1 - \pi)V_D + \Gamma - \tau c(q) \\
&= q\frac{1}{2}\left[\left(1 + \gamma_A^C - \gamma_A^D - \gamma_B^D\right)(1 - \pi)\pi V_C - \left(1 + \gamma_A^C - \gamma_A^D - \gamma_B^D\right)\pi(1 - \pi)V_D\right] + \Gamma - \tau c(q) \\
&= q\frac{1}{2}(1 - \gamma)\pi(1 - \pi)\frac{1}{2}\delta + \Gamma - \tau c(q) \\
&= \frac{1}{4}\pi(1 - \pi)\delta(1 - \gamma)q + \Gamma - \tau c(q)
\end{aligned}
$$

Where the first equality is just application of Bayes' rule, the second is based on substitution of the equilibrium strategies, the third uses the definition of Γ, the fourth is just collection of terms and the fifth one uses the fact that, in equilibrium $\gamma_A^C = \gamma_B^D = \gamma_A^D = \gamma$. A similar approach is used throughout the paper to simplify the equilibrium objective function.

Since γ has already been decided (i.e., it is a function of q^e), the solution, conjecturing an interior one, is obtained differentiating only with respect to q and setting the first order conditions equal to 0. Note that, given the assumptions about $C(q)$, the problem is concave. Finally, Equation (1.2) is obtained by setting the equilibrium condition $q^e = q^*$.

This can be written as

$$\pi(1-\pi)\delta\left(1-F_1\left[\delta\left(\frac{\bar{u}_2}{2}+E\right)q^*\right]\right)=4\tau c^{'}(q^*)$$

Uniqueness follows from the fact that the LHS is strictly decreasing in q, the RHS is strictly increasing in q. The interior solution follows from $LHS(q=0)>RHS(q=0)$ and $LHS(q=1)<RHS(q=1)$. For the last inequality, $\tau\geq\frac{1}{8}$ is a sufficient condition, noticing that the LHS is never above $\pi(1-\pi)$, whose maximum is $\frac{1}{4}$. ■

Proof of Corollary 1.2.

Recalling that the variance of the incumbent's type is $var(\theta)=\pi(1-\pi)$, the implicit function describing q^* is defined as $G(q^*(var(\theta),E),var(\theta),E)=var(\theta)\delta(1-F_1[\delta\left(\frac{\bar{u}_2}{2}+E\right)q^*])=4\tau c^{'}(q^*)$.

By implicit function theorem,

$$\frac{dq^*}{dvar(\theta)}=\frac{\delta(1-F_1[\delta\left(\frac{\bar{u}_2}{2}+E\right)q^*])}{var(\theta)\delta^2\left(\frac{\bar{u}_2}{2}+E\right)f_1(\delta\left(\frac{\bar{u}_2}{2}+E\right)q^*)+4\tau c''(q^*)}>0$$

$$\frac{dq^*}{dE}=-\frac{var(\theta)\delta^2q^*f_1(\delta\left(\frac{\bar{u}_2}{2}+E\right)q^*)}{var(\theta)\delta^2\left(\frac{\bar{u}_2}{2}+E\right)f_1(\delta\left(\frac{\bar{u}_2}{2}+E\right)q^*)+4\tau c''(q^*)}<0$$

■

Proof of Corollary 1.3.

Applying the chain rule,

$$\frac{d\gamma^*}{dvar(\theta)}=\left[\delta\left(\frac{\bar{u}_2}{2}+E\right)f_1(\delta\left(\frac{\bar{u}_2}{2}+E\right)q^*)\right]\frac{dq^*}{dvar(\theta)}>0$$

because $\frac{dq^*}{dvar(\theta)}>0$.

$$\frac{d\gamma^*}{dE}=\delta\left[q^*+\frac{dq^*}{dE}\left(\frac{\bar{u}_2}{2}+E\right)\right]f_1(\delta\left(\frac{\bar{u}_2}{2}+E\right)q^*)$$

Then, $\frac{d\gamma^*}{dE}>0\Rightarrow 1>\frac{\left(\frac{\bar{u}_2}{2}+E\right)var(\theta)\delta^2f_1(\delta\left(\frac{\bar{u}_2}{2}+E\right)q^*)}{var(\theta)\delta^2\left(\frac{\bar{u}_2}{2}+E\right)f_1(\delta\left(\frac{\bar{u}_2}{2}+E\right)q^*+4\tau C''(q^*)}$
which is always true. ■

Proof of Lemma 1.1.

Define for simplicity $K = \frac{1}{\bar{u}_1}\left(\frac{\bar{u}_2}{2} + E\right)$. The ex ante welfare function is given by

$$\begin{aligned} W^{ex} &= \frac{1}{2}\left[\pi(1-\gamma) + (1-\pi)(1-\gamma)\right] + \frac{1}{2}\left[\pi + (1-\pi)\gamma\right] \\ &+ q\{[Pr(x=B)](\hat{\pi}_B V_C + (1-\hat{\pi}_B)V_D) + [Pr(x=A)]\Gamma\} + (1-q)\Gamma - \tau c(q) \end{aligned}$$

where the first line is period 1 expected utility and period 2 is the same as W. Note that, given the equilibrium strategy for the politician, if the state is A then both types choose the sub-optimal action with probability γ (hence V gets 1 with probability $1-\gamma$) and if the state is B then the consonant type always chooses the voter's preferred action, while the dissonant type chooses it with probability γ.

It is possible to rewrite it as

$$W^{ex} = \frac{1}{2}\left[1 - \pi - \pi\delta K q\right] + \frac{1}{4}\pi(1-\pi)\delta(1-\delta K q)q - \tau c(q)$$

The second derivative of Equation (1.3) with respect to q is

$$\frac{\partial^2 W^{ex}}{\partial q^2} = -\frac{1}{2}\pi(1-\pi)\delta^2 K - \tau c^{''}(q)$$

which is negative in the whole interval. ■

Proof of Proposition 1.3.

The (ex ante) optimal level of attention, if interior, is implicitly defined by its first order conditions as

$$\pi(1-\pi)\delta(1-\delta K q) = 2\delta\pi K + (1-\pi)\pi\delta^2 K q + 4\tau c^{'}(q) \qquad \text{(A1.1)}$$

Note that the LHS of (A1.1) and (1.2) is exactly the same. However, the RHS of (A1.1) is above the RHS of (1.2) for every interior q. Moreover, a corner solution at 1 is excluded by the lower bound on τ, while a corner solution at 0 would also imply equilibrium over-attention to the action. ■

Proof of Proposition 1.4.

As a reminder, $\gamma_s^C = Pr(C\ plays\ B\ in\ state\ s)$ and $\gamma_s^D = Pr(D\ plays\ B\ in\ state\ s)$. Moreover, $r_{\tilde{x},\tilde{s}} = Pr(re-elect|\tilde{x},\tilde{s})$.

First of all, the voter re-elects the incumbent if $\hat{\pi}_{\tilde{x},\tilde{s}} > \pi$, chooses the challenger if $\hat{\pi}_{\tilde{x},\tilde{s}} < \pi$ and is indifferent otherwise.
For given γ_s^D and γ_s^C, it is easy to see that

$$\hat{\pi}_{\emptyset,\tilde{s}} = \pi$$

hence $r_{\emptyset,\tilde{s}} \in [0,1]$.

$$\hat{\pi}_{A,\emptyset} > \pi \Leftrightarrow \gamma_A^D + \gamma_B^D > \gamma_A^C + \gamma_B^C$$

hence

$$r_{A,\emptyset} = \begin{cases} 0 & \text{if } \gamma_A^D + \gamma_B^D < \gamma_A^C + \gamma_B^C \\ [0,1] & \text{if } \gamma_A^D + \gamma_B^D = \gamma_A^C + \gamma_B^C \\ 1 & \text{if } \gamma_A^D + \gamma_B^D > \gamma_A^C + \gamma_B^C \end{cases}$$

$$\hat{\pi}_{B,\emptyset} > \pi \Leftrightarrow \gamma_A^D + \gamma_B^D < \gamma_A^C + \gamma_B^C$$

hence

$$r_{B,\emptyset} = \begin{cases} 1 & \text{if } \gamma_A^D + \gamma_B^D < \gamma_A^C + \gamma_B^C \\ [0,1] & \text{if } \gamma_A^D + \gamma_B^D = \gamma_A^C + \gamma_B^C \\ 0 & \text{if } \gamma_A^D + \gamma_B^D > \gamma_A^C + \gamma_B^C \end{cases}$$

$$\hat{\pi}_{A,A} > \pi \Leftrightarrow \gamma_A^D > \gamma_A^C$$

hence

$$r_{A,A} = \begin{cases} 1 & \text{if } \gamma_A^D > \gamma_A^C \\ [0,1] & \text{if } \gamma_A^D = \gamma_A^C \\ 0 & \text{if } \gamma_A^D < \gamma_A^C \end{cases}$$

$$\hat{\pi}_{B,B} > \pi \Leftrightarrow \gamma_B^C > \gamma_B^D$$

hence

$$r_{B,B} = \begin{cases} 0 & \text{if } \gamma_B^D > \gamma_B^C \\ [0,1] & \text{if } \gamma_B^D = \gamma_B^C \\ 1 & \text{if } \gamma_B^D < \gamma_B^C \end{cases}$$

$$\hat{\pi}_{A,B} > \pi \Leftrightarrow \gamma_B^D > \gamma_B^C$$

hence

$$r_{A,B} = \begin{cases} 1 & \text{if } \gamma_B^D > \gamma_B^C \\ [0,1] & \text{if } \gamma_B^D = \gamma_B^C \\ 0 & \text{if } \gamma_B^D < \gamma_B^C \end{cases}$$

$$\hat{\pi}_{B,A} > \pi \Leftrightarrow \gamma_A^C > \gamma_A^D$$

hence

$$r_{B,A} = \begin{cases} 0 & \text{if } \gamma_A^D > \gamma_A^C \\ [0,1] & \text{if } \gamma_A^D = \gamma_A^C \\ 1 & \text{if } \gamma_A^D < \gamma_A^C \end{cases}$$

Moving to the best response correspondences of P, for conjectured levels of attention β^e and q^e, he compares $\mathbb{E}U_\theta(x = B, s)$ with $\mathbb{E}U_\theta(x = A, s)$. Therefore,

$$\gamma_A^D = Pr\left(u_1 \leq \delta\left(\frac{\bar{u}_2}{2} + E\right) q^e \left[(1-\beta^e)(r_{B,\emptyset} - r_{A,\emptyset}) + \beta^e(r_{B,A} - r_{A,A})\right]\right) \tag{A1.2}$$

$$\gamma_B^D = Pr\left(u_1 \leq \delta\left(\frac{\bar{u}_2}{2} + E\right) q^e \left[(1-\beta^e)(r_{B,\emptyset} - r_{A,\emptyset}) + \beta^e(r_{B,B} - r_{A,B})\right]\right) \tag{A1.3}$$

$$\gamma_A^C = Pr\left(u_1 \leq \delta q^e\left(\frac{\bar{u}_2}{2} + E\right) \left[(1-\beta^e)(r_{B,\emptyset} - r_{A,\emptyset}) + \beta^e(r_{B,A} - r_{A,A})\right]\right) \tag{A1.4}$$

$$\gamma_B^C = Pr\left(u_1 + \delta q^e\left(\frac{\bar{u}_2}{2} + E\right) \left[(1-\beta^e)(r_{B,\emptyset} - r_{A,\emptyset}) + \beta^e(r_{B,B} - r_{A,B})\right] \geq 0\right) \tag{A1.5}$$

Clearly, in equilibrium $\gamma_A^D = \gamma_A^C$, hence $r_{A,A}$ and $r_{B,A}$ are unconstrained because of indifference. The same is true for $r_{\emptyset,\tilde{s}}$ since no relevant information is transmitted.

I claim that, in every equilibrium, it must be that $r_{B,B} = 1$, $r_{A,B} = 0$, $r_{B,\emptyset} = 1$ and $r_{A,\emptyset} = 0$. Given that $\gamma_A^D = \gamma_A^C$, those are sequentially rational voting strategies if $\gamma_B^C > \gamma_B^D$. Replacing them in Equations (A1.5) and (A1.3) the result, as required, is

$$\gamma_B^C = 1 > \gamma_B^D = F_1\left[\delta\left(\frac{\bar{u}_2}{2} + E\right) q^e\right]$$

Replacing in (A1.4) and (A1.2) the result is that

$$\gamma_A^C = \gamma_A^D = F_1\left[\delta\left(\frac{\bar{u}_2}{2}+E\right)q^e(1-\beta^e+\beta^e(r_{B,A}-r_{A,A}))\right]$$

To show that there are no other equilibria I use contradiction. First, suppose that there exists an equilibrium where $\gamma_B^C < \gamma_B^D$, and as a consequence $r_{B,B} = 0$, $r_{A,B} = 1$, $r_{B,\emptyset} = 0$ and $r_{A,\emptyset} = 1$. Replacing those values in Equations (A1.5) and (A1.3) the result is

$$\gamma_B^C = Pr\left(u_1 \geq \delta q^e\left(\frac{\bar{u}_2}{2}+E\right)\right) > 0 = \gamma_B^D$$

hence a contradiction.

Finally, suppose that $\gamma_B^C = \gamma_B^D$. Defining $\delta\left(\frac{\bar{u}_2}{2}+E\right)q^e\left[(1-\beta^e)(r_{B,\emptyset}-r_{A,\emptyset})+\beta^e(r_{B,B}-r_{A,B})\right] := D$, this requires $Pr(u_1 \leq D) = Pr(u_1 \geq -D)$, which never holds.

For the last point of the Proposition, if $q=0$ or if $\beta \geq \frac{1}{1-r_{B,A}+r_{A,A}}$ then $\gamma_A^D = \gamma_A^C = 0$, hence observing $x = B, s = A$ is off path. Note that there are no off path beliefs (and consequential randomizations) that are consistent with being off path and that would lead to different equilibria. In case of $q = 0$, in equilibrium $\gamma_A^D = \gamma_B^D = \gamma_A^C = 0$ and $\gamma_B^C = 1$ and there is no $r_{B,A}$ that changes this (hence every off path belief works). In case of $\beta \geq \frac{1}{1-r_{B,A}+r_{A,A}}$, then either off path beliefs and randomizations are such that $r_{B,A} > r_{A,A}$, and as a consequence there is no feasible β such that $x = B, s = A$ is off path, or $r_{B,A} \leq r_{A,A}$ and hence it is optimal, for that value of β, to choose $\gamma_A^D = \gamma_A^C = 0$. ■

Proof of Lemma 1.2.

First, note that

$$Pr(x=B,s=B)(\hat{\pi}_{B,B}V_C+(1-\hat{\pi}_{B,B})V_D)+(1-Pr(x=B,s=B))\Gamma =$$

$$= \frac{1}{2}\left[\pi V_C + \gamma_B^D(1-\pi)V_D\right]+(1-Pr(x=B,s=B))\Gamma$$

Moreover,

$$
\begin{aligned}
& Pr(x=B)(\hat{\pi}_{B,\emptyset}V_C + (1-\hat{\pi}_{B,\emptyset})V_D) + Pr(x=A)\Gamma = \\
& = \frac{1}{2}\left(\gamma_A^C + \gamma_B^C\right)\pi V_C + \frac{1}{2}\left(\gamma_A^D + \gamma_B^D\right)(1-\pi)V_D + Pr(x=A)\Gamma \\
& = Pr(x=B, s=B)(\hat{\pi}_{B,B}V_C + (1-\hat{\pi}_{B,B})V_D) + \\
& + \frac{1}{2}\left(\gamma_A^C\right)\pi V_C + \frac{1}{2}\left(\gamma_A^D\right)(1-\pi)V_D + Pr(x=A)\Gamma \\
& = Pr(x=B, s=B)(\hat{\pi}_{B,B}V_C + (1-\hat{\pi}_{B,B})V_D) + \\
& + \frac{1}{2}\tilde{\gamma}_A\left(\pi V_C + (1-\pi)V_D\right) + Pr(x=A)\Gamma \\
& = Pr(x=B, s=B)(\hat{\pi}_{B,B}V_C + (1-\hat{\pi}_{B,B})V_D) + \\
& + Pr(x=B, s=A)\Gamma + Pr(x=A)\Gamma \\
& = Pr(x=B, s=B)(\hat{\pi}_{B,B}V_C + (1-\hat{\pi}_{B,B})V_D) + \\
& + (1 - Pr(x=B, s=B))\Gamma
\end{aligned}
\tag{A1.6}
$$

Where the first equality is just a straightforward application of Bayes' rule. The second follows from the fact that, as $\gamma_B^C = 1$, $\frac{1}{2}\left[\pi V_C + \gamma_B^D(1-\pi)V_D\right] = Pr(x = B, s = B)(\hat{\pi}_{B,B}V_C + (1-\hat{\pi}_{B,B})V_D)$. The third from the fact that, $\gamma_A^C = \gamma_A^D$, we can define both of them $\tilde{\gamma}_A$. The fourth from the fact that $\frac{1}{2}\tilde{\gamma}_A = Pr(x = B, s = A)$ and $\pi V_C + (1-\pi)V_D = \Gamma$. The last from the fact that $Pr(x = B, s = A) + Pr(x = A) = 1 - Pr(x = B, s = B)$. ■

Proof of Lemma 1.3.

The voter chooses to maximize $\tilde{W}$ with respect to q and β, taking as given the pandering decision of the politician. As shown in the main body of the paper,

$$
\tilde{W} = q\left(\frac{1}{4}\pi(1-\pi)\delta(1-\gamma_B^D)\right) + \left(\pi\delta + (1-\pi)\delta\frac{1}{2}\right) - \tau(q^2 + \beta^2) \tag{A1.7}
$$

From Equation (A1.7), β enters only negatively in the objective function, and hence its solution is $\beta^{**} = 0$. Replacing above, noticing the concavity of the objective function and differentiating with respect to q, the solution (conjecturing an interior solution) is given by

$$\left(\frac{1}{4}\pi(1-\pi)\delta(1-\gamma_B^D)\right) = 2\tau q$$

Combing this with $\gamma_B^D = \delta K q^e$ and the equilibrium condition $q^e = q^{**}$, it turns out that q^{**} is the fixed point of

$$\left(\frac{1}{4}\pi(1-\pi)\delta(1-\delta K q^{**})\right) = 2\tau q^{**}$$

As the LHS is decreasing in q and the RHS is increasing in q, the solution is unique. It is interior as long as $LHS(q=1) \leq RHS(q=1)$, noticing that of course $LHS(q = 0) > RHS(q=0)$.

This happens when $\tau \geq \frac{\pi(1-\pi)(1-\delta K)\delta}{8}$. Note that again $\tau \geq \frac{1}{8}$ is sufficient to guarantee an interior solution. ■

Proof of Lemma 1.4.

As derived in the main body of the paper, the objective function of the ex ante maximization is

$$\tilde{W}^{ex} = \frac{1}{2}(1-\delta K q(1-\beta) + \pi + (1-\pi)\delta K q) + q\left(\frac{1}{4}\pi(1-\pi)\delta(1-\delta K q)\right) +$$
$$+ \left(\pi\delta + (1-\pi)\delta\frac{1}{2}\right) - \tau(q^2+\beta^2)$$

where the first part is period 1 expected utility, the second and the third is period 2 expected utility and the last is the cost of attention.

To determine the Hessian matrix, note that

$$\frac{\partial^2 \tilde{W}^{ex}}{\partial q^2} = -\frac{1}{2}\pi(1-\pi)\delta^2 K - 2\tau < 0$$

$$\frac{\partial^2 \tilde{W}^{ex}}{\partial \beta^2} = -2\tau < 0$$

$$\frac{\partial^2 \tilde{W}^{ex}}{\partial \beta \partial q} = \frac{\partial^2 \tilde{W}^{ex}}{\partial q \partial \beta} = \frac{1}{2}\delta K$$

As $\frac{\partial^2 \tilde{W}^{ex}}{\partial q^2}$ and $\frac{\partial^2 \tilde{W}^{ex}}{\partial \beta^2}$ are both negative, the remaining condition for the negative semidefinitiveness of the Hessian is

$$\frac{\partial^2 \tilde{W}^{ex}}{\partial q^2} * \frac{\partial^2 \tilde{W}^{ex}}{\partial \beta^2} - \left(\frac{\partial^2 \tilde{W}^{ex}}{\partial \beta \partial q}\right)^2 \geq 0$$

that simplifies to

$$\pi(1-\pi)\delta^2 K\tau + (4\tau)^2 \geq (\delta K)^2$$

As a consequence, $4\tau \geq \delta K$ is a sufficient condition for this to be true. ■

Proof of Proposition 1.5.

Thanks to Lemma 1.4 and the fact that the constraints are all linear, Kuhn-Tucker conditions are necessary and sufficient for a maximum.

Hence, the Lagrangian function is

$$L = \tilde{W}^{ex} + \lambda_1(1-q) + \lambda_2(1-\beta)$$

The conditions are

$q \geq 0$ $q\frac{\partial L}{\partial q} = 0$ $q \leq 1$ $\frac{\partial L}{\partial q} \leq 0$

$\beta \geq 0$ $\beta\frac{\partial L}{\partial \beta} = 0$ $\beta \leq 1$ $\frac{\partial L}{\partial \beta} \leq 0$

$\lambda_1 \geq 0$ $\lambda_2 \geq 0$

$\lambda_1(1-q) = 0$ $\lambda_2(1-\beta) = 0$

and there are multiple cases to be considered.

Case 1: $\beta = 0$ and $q = 0$.

This requires $\lambda_1 = 0$ and $\lambda_2 = 0$. To be a maximum, it must be that $\frac{\partial L}{\partial q} \leq 0$ that requires, after substitution, $1 - \pi - 2K \leq 0$.

Case 2: β and q both interior, hence $\lambda_1 = 0$ and $\lambda_2 = 0$. The solution is given by the system of $\frac{\partial L}{\partial q} = 0$ and $\frac{\partial L}{\partial \beta} = 0$. The solution is given by

$$q = \frac{2\tau\pi\delta(1-\pi-2K)}{4\tau\pi(1-\pi)\delta^2 K + 16\tau^2 - (\delta K)^2}$$

$$\beta = \frac{1}{2}\frac{\delta^2 K\pi(1-\pi-2K)}{4\pi(1-\pi)\delta^2 K\tau + 16\tau^2 - (\delta K)^2}$$

which requires $1-\pi-2K \geq 0$. Note that $\frac{2\tau\pi\delta(1-\pi-2K)}{4\tau\pi(1-\pi)\delta^2 K+16\tau^2-(\delta K)^2} < 1$ implies $16\tau^2 - (\delta K)^2 > 2\tau\pi\delta\,[1-\pi-2K-2K(1-\pi)\delta]$.

Case 3: q interior and $\beta = 1$.

This requires $\lambda_1 = 0$ and $\lambda_2 > 0$. Note that, whatever $q < 1$ can be found, when substituting into $\frac{\partial L}{\partial \beta} = 0$ the latter requires $2\lambda_2 = \delta K q - 4\tau < 0$, hence a contradiction.

Case 4: $q = 1$ and β interior.

This requires $\lambda_1 > 0$ and $\lambda_2 = 0$. The solution for β is $\frac{\delta K}{4\tau}$ however, when replacing into $\frac{\partial L}{\partial q} = 0$ it simplifies into $\lambda_1 = \frac{(\delta K)^2}{8\tau} - \frac{1}{2}\pi(1-\pi)\delta^2 K - 2\tau + \frac{1}{4}\pi(1-\pi)\delta - \frac{1}{2}\delta K\pi$, which is greater than zero whenever $16\tau^2 - (\delta K)^2 < 2\tau\pi\delta\,[1-\pi-2K-2K(1-\pi)\delta]$.

Case 5: $q = \beta = 1$.

This requires $\lambda_1 > 0$ and $\lambda_2 > 0$. When substituting into $\frac{\partial L}{\partial \beta} = 0$ the latter requires $2\lambda_2 = \delta K - 4\tau < 0$, hence a contradiction.

Case 6: $q = 0$ and β interior.

This requires $\lambda_1 = 0$ and $\lambda_2 = 0$. When replacing, however, the solution for β is only 0, hence a contradiction.

Case 7: q interior and $\beta = 0$.

This requires $\lambda_2 = 0$. However, $\frac{\partial L}{\partial \beta} \leq 0$ only if $q = 0$, hence a contradiction.

Case 8: $q = 0$ and $\beta = 1$.

Then, $\frac{\partial L}{\partial \beta} = 0$ implies a negative λ_2, hence a contradiction.

Case 9: $q = 1$ and $\beta = 0$.

This requires $\lambda_2 = 0$ and $\frac{\partial L}{\partial \beta} \leq 0$. However, when substituting, we obtain $\frac{\partial L}{\partial \beta} = \frac{1}{2}\delta K > 0$, hence a contradiction.

To sum up, the ex ante maximization problem has just three solutions. $q = \beta = 0$ if $1 - \pi - 2\delta K \leq 0$.

$$q = \frac{2\tau\pi\delta(1-\pi-2K)}{4\tau\pi(1-\pi)\delta^2 K + 16\tau^2 - (\delta K)^2}$$
$$\beta = \frac{1}{2}\frac{\delta^2 K\pi(1-\pi-2K)}{4\pi(1-\pi)\delta^2 K\tau + 16\tau^2 - (\delta K)^2}$$

if $1 - \pi - 2K > 0$ and $16\tau^2 - (\delta K)^2 > 2\tau\pi\delta\left[1 - \pi - 2K - 2K(1-\pi)\delta\right]$.

$$q = 1$$
$$\beta = \frac{\delta K}{4\tau}$$

if $1 - \pi - 2K > 0$ and $16\tau^2 - (\delta K)^2 \leq 2\tau\pi\delta\left[1 - \pi - 2K - 2K(1-\pi)\delta\right]$.
As the equilibrium level of β is always 0, $1-\pi-2K > 0$ guarantees equilibrium (strict) under-attention to the state of the world.

The last part of Proposition 1.5 is a simple comparison between $q^{**} = \frac{\pi(1-\pi)\delta}{8\tau+\pi(1-\pi)\delta^2 K}$ and the interior ex ante optimal q, looking for parameters where the latter is smaller. Note that it is possible to find a sufficient condition for equilibrium over-attention to the action, that is, $4\tau\pi > 1$. ■

Proof of Lemma 1.5.

In terms of re-election strategy, note that we can use the same comparisons as in the proof of Lemma 1.3. For any given conjectured η and generic re-election strategies,

$$\begin{aligned}\gamma_A^D &= Pr\left[u_1 + (\eta r_{A,A} + (1-\eta)r_\emptyset)\,\delta\left(\frac{\bar{u}_2}{2} + E\right) \leq (\eta r_{B,A} + (1-\eta)r_\emptyset)\,\delta\left(\frac{\bar{u}_2}{2} + E\right)\right] \\ &= Pr\left(u_1 \leq \eta\delta(r_{B,A} - r_{A,A})\left(\frac{\bar{u}_2}{2} + E\right)\right)\end{aligned}$$

and for the same logic

$$\gamma_B^D = Pr\left(u_1 \leq \eta\delta(r_{B,B} - r_{A,B})\left(\frac{\bar{u}_2}{2} + E\right)\right)$$

$$\gamma_A^C = Pr\left(u_1 \leq \eta\delta(r_{B,A} - r_{A,A})\left(\frac{\bar{u}_2}{2} + E\right)\right)$$

$$\gamma_B^C = Pr\left(u_1 + \eta\delta(r_{B,B} - r_{A,B})\left(\frac{\bar{u}_2}{2} + E\right) \geq\right)$$

Note that, as $\gamma_A^C = \gamma_A^D$, $\hat{\pi}_{A,A} = \hat{\pi}_{B,A} = \pi$. Moreover, $\hat{\pi}_{\emptyset} = \pi$, so $r_{B,A}$, $r_{A,A}$ and $r_{\emptyset}$ are unconstrained. Using Assumption 1.2, I set them equal.

Finally, I claim that in equilibrium $r_{B,B} > r_{A,B}$. This requires $\gamma_B^C > \gamma_B^D$, so that $r_{B,B} = 1$ and $r_{A,B} = 0$. Plugging into the above equations, those re-election probabilities bring to $\gamma_B^C = 1$ and $\gamma_B^D = F_1\left(\eta\delta\left(\frac{\bar{u}_2}{2} + E\right)\right) < 1$. Hence, the lemma defines an equilibrium.

To show uniqueness (up to the indifference-related randomizations), note that if $r_{B,B} = r_{A,B}$ then $\gamma_B^C = 1$ and $\gamma_B^D = 0$, but $r_{B,B} = r_{A,B}$ requires $\gamma_B^C = \gamma_B^D$, hence a contradiction. Finally, suppose that there is an equilibrium where $r_{B,B} < r_{A,B}$. As a consequence, $r_{B,B} = 0$, $r_{A,B} = 1$, $\gamma_B^D = 0$ and $\gamma_B^C = 1 - F_1\left(\eta\delta\left(\frac{\bar{u}_2}{2} + E\right)\right) > 0$, which is again a contradiction.

Finally, note that if $\eta = 0$ or if $r_{B,A} \leq r_{A,A}$ $x = B, s = A$ is off path, hence Bayes' Rule cannot be used to define $r_{B,A}$. I claim that there are no off path beliefs that change the equilibrium structure outlined above. If beliefs (and the re-election strategy that follows) are such that $r_{B,A} \leq r_{A,A}$, then choosing $\gamma_A^C = \gamma_A^D = 0$ is optimal. Off path beliefs and the consequent re-election strategy such that $r_{B,A} > r_{A,A}$ are inconsistent with the fact of being off path, because in that case $\gamma_A^C = \gamma_A^D > 0$. Finally, if $\eta = 0$ then there is no value of $r_{B,A}$ that changes the equilibrium strategies in that case. ∎

Proof of Lemma 1.6.

V chooses $\eta \in [0,1]$ to maximize W_b, where

$$W_b = \eta\left(Pr(x = B, s = B)(\hat{\pi}_{B,B}V_C + (1 - \hat{\pi}_{B,B})V_D) + (1 - Pr(x = B, s = B))\Gamma\right) + (1-\eta)\Gamma - \tau C(\eta)$$

Replacing the equilibrium probabilities established by Lemma 1.5 and re-arranging, the objective function becomes

$$W_b = \frac{1}{4}\pi(1-\pi)\delta(1-\gamma_B^D)\eta + \Gamma - \tau C(\eta)$$

Noticing that the maximization is concave, the equilibrium value of η, if interior, sets $\frac{\partial W_b}{\partial \eta} = 0$ keeping γ_B^D fixed and then imposing the rationality condition $\eta^e = \eta^*$.
Looking at the implicit equation defining η^*, it is clear that the LHS is strictly decreasing in η (because γ_B^D is increasing in η) and the RHS is strictly increasing, hence there is a unique solution. Moreover, $LHS(\eta = 0) > 0 = RHS(\eta = 0)$. Finally, the assumption on τ guarantees an interior solution, as the LHS of the implicit equation is strictly smaller than $\frac{1}{8} = min\left[\tau C^{'}(\eta = 1)\right]$ ■

Proof of Proposition 1.6.

From an ex ante perspective, V chooses η to maximize

$$W_b^{ex} = \frac{1}{2}\left[1 + \pi + (1-\pi)\delta K\eta\right] + W_b$$

where the first part is just period 1 expected welfare. Looking at the first order condition, the ex ante optimal attention solves

$$\frac{1}{4}\pi(1-\pi)\delta(1-\delta K\eta) + \frac{1}{2}(1-\pi)\delta K - \frac{1}{4}\pi(1-\pi)\delta^2 K\eta = \tau C^{'}(\eta)$$

The RHS is the same as the condition for the equilibrium level of attention. The LHS is the same plus $\frac{1}{2}(1-\pi)\delta K - \frac{1}{4}\pi(1-\pi)\delta^2 K\eta$. Note that, however, we can re-arrange the additional element as $\frac{1}{2}(1-\pi)\delta K\left(1 - \frac{1}{2}\pi\delta\eta\right) > 0\ \forall\eta \in [0,1]$.
As a consequence, the LHS of the ex ante condition is strictly above the LHS of the equilibrium condition, hence $\eta^{**} > \eta^*$, that is, we have equilibrium under-attention. ■

Lemma A1.1 *If Assumption 1.2 does not hold, then a sufficient condition for equilibrium under-attention is* $r_{B,A} - r_{A,A} < (1-\pi)\left(1 - \frac{1}{2}\pi\delta\right)$.

Proof of Lemma A1.1.

If Assumption 1.2 does not apply, then $r_{B,A}$ and $r_{A,A}$ are unconstrained. Define $\epsilon = r_{B,A} - r_{A,A}$, and as a consequence $\gamma_A^D = \gamma_A^C = \delta K \eta \epsilon$. Nothing changes in the derivation of the equilibrium attention η^*. In terms of ex ante attention, instead, it is easy to show that it solves

$$\frac{1}{4}\pi(1-\pi)\delta(1-\delta K\eta) + \frac{1}{2}\delta K\left[1-\pi-\epsilon-\frac{1}{2}\pi(1-\pi)\delta\eta\right] = \tau C'(\eta)$$

because period 1 expected utility is now $\frac{1}{2}\left[1-\delta K\eta\epsilon + \pi + (1-\pi)\delta K\eta\right]$.
A sufficient condition for $\eta^{**} > \eta*$ is $\frac{1}{2}\delta K\left[1-\pi-\epsilon-\frac{1}{2}\pi(1-\pi)\delta\eta\right] > 0\ \forall \eta \in [0,1]$, and this happens as long as $\epsilon < (1-\pi)\left(1-\frac{1}{2}\pi\delta\right)$, as claimed. ■

B1 Endogenous attention to actions, exogenous state revelation

This appendix considers the case where only q is allocated endogenously by V, but I allow $\beta \in [0,1]$. I am assuming that the voter has to allocate attention if she wants to learn the action of the politician, but there is a positive, exogenous probability β that the period 1 state is learned before the election. As only q is endogenous, the cost structure is simply $c(q)$. Trivially, the results of Proposition 1.4 apply to this case as well, with the difference that β is a parameter known by all the players. For tractability, I keep Assumption 1.2 and the result stated in Lemma 1.2 applies in this case as well. In equilibrium the voter's problem is

$$max_{q\in[0,1]} W^\beta$$

where, keeping the same notation as the main body of the paper,

$$W^\beta = \frac{1}{4}\pi(1-\pi)\delta\left(1-\gamma_B^D\right)q + \Gamma - \tau c(q) \tag{B1.1}$$

Equation (B1.1) is already instructive. First of all, for the same logic as in Section 1.3.2, β does not enter in this maximization and also it does not affect γ_B^D, hence it has no effect on the equilibrium attention to the action. The following lemma summarizes the equilibrium attention.

Lemma B1.1 *The allocation of attention to the action in equilibrium $q^{\beta,*}$ is unique,* *interior and implicitly defined by*

$$\frac{1}{4}\pi(1-\pi)\delta\left(1-F_1\left[\delta\left(\frac{\bar{u}_2}{2}+E\right)q^{\beta,*}\right]\right)=\tau c^{'}(q^{\beta,*})$$

The proof of Lemma B1.1 is the same as the one of Proposition 1.2.

When I look at the ex ante optimal attention allocation, however, things are different and β plays a role. In particular, note that the objective function is given by

$$W^{\beta,ex}=\frac{1}{2}\left[1-\tilde{\gamma}_A+\pi+(1-\pi)\gamma_B^D\right]+W^{\beta} \tag{B1.2}$$

where now both $\tilde{\gamma}_A = F_1\left[\delta\left(\frac{\bar{u}_2}{2}+E\right)q\left(1-\beta\right)\right] = \delta Kq(1-\beta)$ and $\gamma_B^D = F_1\left[\delta\left(\frac{\bar{u}_2}{2}+E\right)q\right] = \delta Kq$ are endogenous in the maximization. As in the body of the paper, the first part of Equation (B1.2) is the present part of the voter's welfare, that is, the one given by the policy choice in period one, and the second part is the selection part. Note that β does not enter in the selection part, but it plays a role in the present welfare. In particular, as in Section 1.3.2, from the point of view of period 1 welfare β reduces the bad pandering $\tilde{\gamma}_A$, that is, the probability that P chooses action B in state A, without affecting the good pandering, that is, the probability that the dissonant politician chooses B in state B.

This effect of q has important consequences for the ex ante optimal allocation of q. In terms of selection welfare, a higher q is always bad because it increases γ_B^D, making the dissonant incumbent more similar to the consonant type and hence making selection harder. Hence, if we look only at the selection welfare, in equilibrium V always tends to put too much attention with respect to the ex ante optimal level. However, in terms of present welfare, things may be different, because γ_B^D enters positively. As a consequence, the effect of q on period 1 welfare is in principle ambiguous, as it enters positively with γ_B^D and negatively with $\tilde{\gamma}_A$. As long as β is not too high, the same dynamic as in Section 1.3.1 applies, but a sufficiently high β can make the effect of q on present ex ante welfare positive, and this can even be stronger than the negative effect of q on selection welfare, implying that there could be equilibrium under-attention.

Proposition B1.1 *Define $q^{\beta,**}$ the ex ante optimal attention to the action when β is* *exogenous. If $\beta < \pi$ then $q^{\beta,*} > q^{\beta,**}$, that is, there is equilibrium over-attention to the action.*

Proof of Proposition B1.1.

V chooses $q \in [0,1]$ to maximize Equation B1.2. Taking the derivative,

$$\frac{\partial W^{\beta,ex}}{\partial q} = \frac{1}{4}\pi(1-\pi)\delta\,(1-\delta Kq) - \tau c^{'}(q) - \frac{1}{4}\pi(1-\pi)\delta^2 Kq + \frac{1}{2}(1-\pi)\delta K - \frac{1}{2}\delta K(1-\beta)$$

where the first and second element are the same of the equilibrium maximization, the third element is the additional (negative) effect of q on selection welfare that enters through γ_B^D, the fourth part is the (positive) effect of q on first period welfare again through γ_B^D and the last part is the (negative) additional effect of q on present welfare coming through $\tilde{\gamma}_A$.

Note that the second order conditions are satisfied. Moreover, if we rewrite the FOC as

$$\frac{1}{4}\pi(1-\pi)\delta\,(1-\delta Kq) - \frac{1}{4}\pi(1-\pi)\delta^2 Kq + \frac{1}{2}(1-\pi)\delta K - \frac{1}{2}\delta K(1-\beta) = \tau c^{'}(q)$$

it is easy to see that the LHS is strictly decreasing in q and the RHS is strictly increasing in q, hence there will be a unique solution (it may be a corner of course).

A sufficient condition for equilibrium over-attention is that, $\forall q$,

$$-\frac{1}{4}\pi(1-\pi)\delta^2 Kq + \frac{1}{2}(1-\pi)\delta K - \frac{1}{2}\delta K(1-\beta) < 0$$

As it must hold for every q, I replace $q = 0$ and simplify the terms, obtaining $\beta < \pi$.

■

C1 Generic prior on the state

The main body of the paper assumes a flat prior on the state. In this appendix we show how the conditions are modified when we allow for a generic prior $Pr(s_t = A) = p \in (0,1)$. The rest of the set up is the same as in the body of the paper.

First of all, starting from the framework of Section 1.3.2, I show that most of the basic results are unchanged. I consider separately the two cases from Section 1.3.1 and 1.3.2.

General results

Proposition C1.1 *In every equilibrium with two types of attention:*

1. $\gamma_B^C = 1$, $\gamma_B^D = F_1\left[\delta\left(\frac{\bar{u}_2}{2} + E\right)q\right]$;

2. $\gamma_A^C = \gamma_A^D = F_1\left[\delta\left(\frac{\bar{u}_2}{2} + E\right)q(1-\beta+\beta(r_{B,A} - r_{A,A}))\right]$.

3. $r_{B,\emptyset} = 1$, $r_{A,\emptyset} = 0$, $r_{B,B} = 1$, $r_{A,B} = 0$. $r_{B,A}$, $r_{A,A}$ *and* $r_{\emptyset,\tilde{s}}$ *are unconstrained because of indifference (or because off-path).*

Proof of Proposition C1.1.

As a reminder, $\gamma_s^C = Pr(C\ plays\ B\ in\ state\ s)$ and $\gamma_s^D = Pr(D\ plays\ B\ in\ state\ s)$. Moreover, $r_{\tilde{x},\tilde{s}} = Pr(re-elect|\tilde{x},\tilde{s})$.

First of all, the voter re-elects the incumbent if $\hat{\pi}_{\tilde{x},\tilde{s}} > \pi$, chooses the challenger if $\hat{\pi}_{\tilde{x},\tilde{s}} < \pi$ and is indifferent otherwise.
For given γ_s^D and γ_s^C, it is easy to see that

$$\hat{\pi}_{\emptyset,\tilde{s}} = \pi$$

hence $r_{\emptyset,\tilde{s}} \in [0,1]$.

$$\hat{\pi}_{A,\emptyset} > \pi \Leftrightarrow p\gamma_A^D + (1-p)\gamma_B^D > p\gamma_A^C + (1-p)\gamma_B^C$$

hence

$$r_{A,\emptyset} = \begin{cases} 0 & \text{if } p\gamma_A^D + (1-p)\gamma_B^D < p\gamma_A^C + (1-p)\gamma_B^C \\ [0,1] & \text{if } p\gamma_A^D + (1-p)\gamma_B^D = p\gamma_A^C + (1-p)\gamma_B^C \\ 1 & \text{if } p\gamma_A^D + (1-p)\gamma_B^D > p\gamma_A^C + (1-p)\gamma_B^C \end{cases}$$

$$\hat{\pi}_{B,\emptyset} > \pi \Leftrightarrow p\gamma_A^D + (1-p)\gamma_B^D < p\gamma_A^C + (1-p)\gamma_B^C$$

hence

$$r_{B,\emptyset} = \begin{cases} 1 & \text{if } p\gamma_A^D + (1-p)\gamma_B^D < p\gamma_A^C + (1-p)\gamma_B^C \\ [0,1] & \text{if } p\gamma_A^D + (1-p)\gamma_B^D = p\gamma_A^C + (1-p)\gamma_B^C \\ 0 & \text{if } p\gamma_A^D + (1-p)\gamma_B^D > p\gamma_A^C + (1-p)\gamma_B^C \end{cases}$$

$$\hat{\pi}_{A,A} > \pi \Leftrightarrow \gamma_A^D > \gamma_A^C$$

hence

$$r_{A,A} = \begin{cases} 1 & \text{if } \gamma_A^D > \gamma_A^C \\ [0,1] & \text{if } \gamma_A^D = \gamma_A^C \\ 0 & \text{if } \gamma_A^D < \gamma_A^C \end{cases}$$

$$\hat{\pi}_{B,B} > \pi \Leftrightarrow \gamma_B^C > \gamma_B^D$$

hence

$$r_{B,B} = \begin{cases} 0 & \text{if } \gamma_B^D > \gamma_B^C \\ [0,1] & \text{if } \gamma_B^D = \gamma_B^C \\ 1 & \text{if } \gamma_B^D < \gamma_B^C \end{cases}$$

$$\hat{\pi}_{A,B} > \pi \Leftrightarrow \gamma_B^D > \gamma_B^C$$

hence

$$r_{A,B} = \begin{cases} 1 & \text{if } \gamma_B^D > \gamma_B^C \\ [0,1] & \text{if } \gamma_B^D = \gamma_B^C \\ 0 & \text{if } \gamma_B^D < \gamma_B^C \end{cases}$$

$$\hat{\pi}_{B,A} > \pi \Leftrightarrow \gamma_A^C > \gamma_A^D$$

hence

$$r_{B,A} = \begin{cases} 0 & \text{if } \gamma_A^D > \gamma_A^C \\ [0,1] & \text{if } \gamma_A^D = \gamma_A^C \\ 1 & \text{if } \gamma_A^D < \gamma_A^C \end{cases}$$

Moving to the best response correspondences of P, for conjectured levels of attention β^e and q^e, he compares $\mathbb{E}U_\theta(x = B, s)$ with $\mathbb{E}U_\theta(x = A, s)$. This is not affected by p

as it is known by the politician. Therefore,

$$\gamma_A^D = Pr\left(u_1 \leq \delta\left(\frac{\bar{u}_2}{2} + E\right) q^e \left[(1-\beta^e)(r_{B,\emptyset} - r_{A,\emptyset}) + \beta^e(r_{B,A} - r_{A,A})\right]\right) \quad \text{(C1.1)}$$

$$\gamma_B^D = Pr\left(u_1 \leq \delta\left(\frac{\bar{u}_2}{2} + E\right) q^e \left[(1-\beta^e)(r_{B,\emptyset} - r_{A,\emptyset}) + \beta^e(r_{B,B} - r_{A,B})\right]\right) \quad \text{(C1.2)}$$

$$\gamma_A^C = Pr\left(u_1 \leq \delta q^e \left(\frac{\bar{u}_2}{2} + E\right) \left[(1-\beta^e)(r_{B,\emptyset} - r_{A,\emptyset}) + \beta^e(r_{B,A} - r_{A,A})\right]\right) \quad \text{(C1.3)}$$

$$\gamma_B^C = Pr\left(u_1 + \delta q^e \left(\frac{\bar{u}_2}{2} + E\right) \left[(1-\beta^e)(r_{B,\emptyset} - r_{A,\emptyset}) + \beta^e(r_{B,B} - r_{A,B})\right] \geq 0\right) \quad \text{(C1.4)}$$

Clearly, in equilibrium $\gamma_A^D = \gamma_A^C$, hence $r_{A,A}$ and $r_{B,A}$ are unconstrained because of indifference. The same is true for $r_{\emptyset,\tilde{s}}$ since no relevant information is transmitted.

I claim that in every equilibrium it must be that $r_{B,B} = 1$, $r_{A,B} = 0$, $r_{B,\emptyset} = 1$ and $r_{A,\emptyset} = 0$. Given that $\gamma_A^D = \gamma_A^C$, those are sequentially rational voting strategies if $\gamma_B^C > \gamma_B^D$. Replacing them in Equations (C1.4) and (C1.2) the result is

$$\gamma_B^C = 1 > \gamma_B^D = F_1\left[\delta\left(\frac{\bar{u}_2}{2} + E\right) q^e\right]$$

Replacing in (C1.3) and (C1.1) the result is that

$$\gamma_A^C = \gamma_A^D = F_1\left[\delta\left(\frac{\bar{u}_2}{2} + E\right) q^e(1 - \beta^e + \beta^e(r_{B,A} - r_{A,A}))\right]$$

To show that there are no other equilibria I use contradiction. First, suppose that there exists an equilibrium where $\gamma_B^C < \gamma_B^D$, and as a consequence $r_{B,B} = 0$, $r_{A,B} = 1$, $r_{B,\emptyset} = 0$ and $r_{A,\emptyset} = 1$. Replacing those values in Equations (C1.4) and (C1.2) the result is

$$\gamma_B^C = Pr\left(u_1 \geq \delta q^e \left(\frac{\bar{u}_2}{2} + E\right)\right) > 0 = \gamma_B^D$$

hence a contradiction.

Finally, suppose that $\gamma_B^C = \gamma_B^D$. Defining

$$\delta\left(\frac{\bar{u}_2}{2} + E\right) q^e \left[(1-\beta^e)(r_{B,\emptyset} - r_{A,\emptyset}) + \beta^e(r_{B,B} - r_{A,B})\right] := D$$

this requires $Pr(u_1 \leq D) = Pr(u_1 \geq -D)$, which never holds.

For the last point of the Proposition, note that, if $q = 0$ or if $\beta \geq \frac{1}{1-r_{B,A}+r_{A,A}}$ then $\gamma_A^D = \gamma_A^C = 0$, hence observing $x = B, s = A$ is off path. Note that there are no off path beliefs (and consequential randomizations) that are consistent with being off path and that would lead to different equilibria. In case of $q = 0$, in equilibrium $\gamma_A^D = \gamma_B^D = \gamma_A^C = 0$ and $\gamma_B^C = 1$ and there is no $r_{B,A}$ that changes this (hence every off path belief works). In case of $\beta \geq \frac{1}{1-r_{B,A}+r_{A,A}}$, then either off path beliefs and randomizations are such that $r_{B,A} > r_{A,A}$, and as a consequence there is no feasible β such that $x = B, s = A$ is off path, or $r_{B,A} \leq r_{A,A}$ and hence it is optimal, for that value of β, to choose $\gamma_A^D = \gamma_A^C = 0$. ■

Lemma 1.2 has its equivalent in this section as well, and I keep Assumption 1.2.

Lemma C1.1 *In equilibrium, for any prior p on the state,* $Pr(x = B, s = B)(\hat{\pi}_{B,B}V_C + (1-\hat{\pi}_{B,B})V_D) + (1-Pr(x = B, s = B))\Gamma = Pr(x = B)(\hat{\pi}_{B,\emptyset}V_C + (1-\hat{\pi}_{B,\emptyset})V_D) + Pr(x = A)\Gamma$.

Proof of Lemma C1.1.

First, note that

$$Pr(x = B, s = B)(\hat{\pi}_{B,B}V_C + (1 - \hat{\pi}_{B,B})V_D) + (1 - Pr(x = B, s = B))\Gamma =$$

$$= (1 - p)\left[\pi V_C + \gamma_B^D(1 - \pi)V_D\right] + (1 - Pr(x = B, s = B))\Gamma$$

Moreover,

$$
\begin{aligned}
& Pr(x=B)(\hat{\pi}_{B,\emptyset}V_C+(1-\hat{\pi}_{B,\emptyset})V_D)+Pr(x=A)\Gamma = \\
&= \left(p\gamma_A^C+(1-p)\gamma_B^C\right)\pi V_C + \\
&+ \left(p\gamma_A^D+(1-p)\gamma_B^D\right)(1-\pi)V_D+Pr(x=A)\Gamma \\
&= Pr(x=B,s=B)(\hat{\pi}_{B,B}V_C+(1-\hat{\pi}_{B,B})V_D)+ \\
&+ p\left(\gamma_A^C\right)\pi V_C+p\left(\gamma_A^D\right)(1-\pi)V_D+Pr(x=A)\Gamma \\
&= Pr(x=B,s=B)(\hat{\pi}_{B,B}V_C+(1-\hat{\pi}_{B,B})V_D)+ \\
&+ p\tilde{\gamma}_A\left(\pi V_C+(1-\pi)V_D\right)+Pr(x=A)\Gamma \\
&= Pr(x=B,s=B)(\hat{\pi}_{B,B}V_C+(1-\hat{\pi}_{B,B})V_D)+ \\
&+ Pr(x=B,s=A)\Gamma+Pr(x=A)\Gamma \\
&= Pr(x=B,s=B)(\hat{\pi}_{B,B}V_C+(1-\hat{\pi}_{B,B})V_D)+ \\
&+ (1-Pr(x=B,s=B))\Gamma
\end{aligned} \tag{C1.5}
$$

Where the first equality is just a straightforward application of Bayes' rule. The second follows from the fact that, as $\gamma_B^C=1$, $(1-p)\left[\pi V_C+\gamma_B^D(1-\pi)V_D\right] = Pr(x = B,s=B)(\hat{\pi}_{B,B}V_C+(1-\hat{\pi}_{B,B})V_D)$. The third from the fact that, $\gamma_A^C=\gamma_A^D$, we can define both of them $\tilde{\gamma}_A$. The fourth from the fact that $p\tilde{\gamma}_A=Pr(x=B,s=A)$ and $\pi V_C+(1-\pi)V_D=\Gamma$. The last from the fact that $Pr(x=B,s=A)+Pr(x=A) = 1-Pr(x=B,s=B)$. ■

As a consequence, the objective function maximized in equilibrium is

$$
\begin{aligned}
W^p &= q\{Pr(x=B)(\hat{\pi}_{B,\emptyset}V_C+(1-\hat{\pi}_{B,\emptyset})V_D)+Pr(x=A)\Gamma\}+\Gamma(1-q)-\tau C(q,\beta) \\
&= q(1-p)^2\pi(1-\pi)(1-\gamma_B^D)\delta+\Gamma-\tau C(q,\beta)
\end{aligned} \tag{C1.6}
$$

where the second line of C1.6 is just straightforward substitution and rearrangement. The objective function of the ex ante maximization is also only slightly different, and it is now given by

$$W^{p,ex} = \left[p\left(1-\tilde{\gamma}_A\right)+(1-p)\left(\pi+(1-\pi)\gamma_B^D\right)\right]+W^p \tag{C1.7}$$

Endogenous q with $\beta = 0$

It is useful to assume that $\tau \geq \frac{\delta}{4}$, so that the equilibrium level of attention is interior. The cost function is of course just $c(q)$ and, for any conjectured level of attention q^e, $\gamma_B^D = \gamma_A^C = \gamma_A^D = \delta K q^e$. The next proposition shows that, for p not too small, there is equilibrium over-attention to the action.

Proposition C1.2 *For every $p > \bar{p}$, with $\bar{p} < \frac{1}{2}$, there is equilibrium over-attention to the action*

Proof of Proposition C1.2.

First, note that the equilibrium level of attention $q^{p,*}$ is unique, interior and implicitly given by

$$(1-p)^2\pi(1-\pi)(1-\delta K q^{p,*})\delta = c'\left(q^{p,*}\right)$$

This follows from the differentiation of (C1.6) with respect to q, keeping γ_B^D exogenous and noticing that the second order conditions are trivially satisfied, and then imposing the rationality condition.

The equilibrium is interior because $LHS(q=0) > 0 = RHS(q=0)$ and $LHS(q=1) < RHS(q=0)$ and it is unique because the LHS is strictly decreasing in q while the RHS is strictly increasing.

Looking now at Equation (C1.7), note that the effect of q on the policy part of the welfare function is now given by $-\delta K p + (1-p)(1-\pi)\delta K$. Overall, the result of the derivative of (C1.7) treating everything as endogenous is given by

$$(1-p)^2\pi(1-\pi)(1-\delta K q)\delta - \delta K p + (1-p)(1-\pi)\delta K - q(1-p)^2\pi(1-\pi)\delta^2 K = c'(q)$$

A sufficient condition for equilibrium over-attention is that, for every q,

$$-\delta K p + (1-p)(1-\pi)\delta K - q(1-p)^2\pi(1-\pi)\delta^2 K < 0$$

As it must hold for every q, I set $q = 0$ and this simplifies as

$$p > \frac{1-\pi}{2-\pi} = \bar{p} < \frac{1}{2}$$

■

Intuitively, a very small p means that state A is unlikely, that is, that bad pandering is something V should not worry about too much. As a consequence, the positive effect of q in increasing the good pandering becomes predominant, and it could even dominate the (unchanged) negative effect on selection of the good pandering.

Endogenous q and β

For this set of results I use again Assumption 1.3 and, instead of Assumption 1.4, I now assume the following (for the same reasons of global concavity of the ex ante maximization problem):

Assumption C1.1 $2\tau \geq p\delta K$

In terms of results,

Lemma C1.2 *The allocation of attention in equilibrium is unique and is given by* $\beta^{p,**} = 0$ *and* $q^{p,**} = \frac{(1-p)^2\pi(1-\pi)\delta}{2\tau+(1-p)^2\pi(1-\pi)\delta^2 K}$.

Proof of Lemma C1.2.

From Equation (C1.6), β enters only negatively in the objective function, and hence its solution is $\beta^{p,**} = 0$. Replacing above, noticing the concavity of the objective function and differentiating with respect to q, the solution (conjecturing an interior one) is given by

$$(1-p)^2\pi(1-\pi)\delta(1-\gamma_B^D) = 2\tau q$$

Combing this with $\gamma_B^D = \delta K q^e$ and the equilibrium condition $q^e = q^{p,**}$, it turns out that $q^{p,**}$ is the fixed point of

$$(1-p)^2\pi(1-\pi)\delta(1-\delta K q^{p,**}) = 2\tau q^{p,**}$$

As the LHS is decreasing in q and the RHS is increasing in q, the solution is unique. It is interior as long as $LHS(q=1) \leq RHS(q=1)$, noting that $LHS(q=0) > RHS(q=0)$. This happens when $\tau \geq \frac{(1-p)^2\pi(1-\pi)(1-\delta K)\delta}{2}$. Note that again $\tau \geq \frac{\delta}{4}$ is sufficient to guarantee an interior solution. ■

The result of Lemma C1.2 is just the adaptation of what stated in Lemma 1.3. In terms of comparative statics, the following corollary summarizes the additional result:

Corollary C1.1 *The equilibrium level of attention to the state is decreasing in p.*

Proof of Corollary C1.1.

Straightforward differentiation of $q^{p,**}$ with respect to p. ■

Intuitively the result makes sense: if state A is very likely, then there is nothing to learn from the actions of P, because both C and D types behave in the same way.

Proposition C1.3 *Comparing the ex ante optimal and the equilibrium level of attention:*

1. *The ex ante optimal level of attention to the state is always weakly higher than the equilibrium. If $K\left[(1-p)(1-\pi)-p\right]+(1-p)^2\pi(1-\pi)>0$ then the ex ante optimal β is strictly higher than the equilibrium;*

2. *If $4\tau^2((1-p)(1-\pi)-p)+(1-p)^2\pi(1-\pi)\delta^3K(2\tau((1-p)(1-\pi)-p)+p^2) < 2\tau\left((1-p)^2\pi(1-\pi)\delta\right)^2$ then the ex ante optimal level of attention to the action is strictly smaller than the equilibrium;*

Proof of Proposition C1.3.

First, I show that Assumption C1.1 is sufficient for the global concavity of the objective function (C1.7).

To determine the Hessian matrix, note that

$$\frac{\partial^2 W^{p,ex}}{\partial q^2} = -2(1-p)^2\pi(1-\pi)\delta^2K - 2\tau < 0$$

$$\frac{\partial^2 W^{p,ex}}{\partial \beta^2} = -2\tau < 0$$

$$\frac{\partial^2 W^{p,ex}}{\partial \beta \partial q} = \frac{\partial^2 W^{p,ex}}{\partial q \partial \beta} = p\delta K$$

As $\frac{\partial^2 W^{p,ex}}{\partial q^2}$ and $\frac{\partial^2 W^{p,ex}}{\partial \beta^2}$ are both negative, the remaining condition for the negative semidefinitiveness of the Hessian is

$$\frac{\partial^2 W^{p,ex}}{\partial q^2} * \frac{\partial^2 W^{p,ex}}{\partial \beta^2} - \left(\frac{\partial^2 W^{p,ex}}{\partial \beta \partial q}\right)^2 \geq 0$$

that simplifies to

$$(1-p)^2 \pi (1-\pi) \delta^2 K 4\tau + (2\tau)^2 \geq (p\delta K)^2$$

As a consequence, $2\tau \geq p\delta K$ is a sufficient condition for this to be true.

Given the above, Kuhn-Tucker conditions are necessary and sufficient for a maximum. Hence, the Lagrangian function is

$$L = W^{p,ex} + \lambda_1(1-q) + \lambda_2(1-\beta)$$

The conditions are

$q \geq 0$ $q\frac{\partial L}{\partial q} = 0$ $q \leq 1$ $\frac{\partial L}{\partial q} \leq 0$

$\beta \geq 0$ $\beta\frac{\partial L}{\partial \beta} = 0$ $\beta \leq 1$ $\frac{\partial L}{\partial \beta} \leq 0$

$\lambda_1 \geq 0$ $\lambda_2 \geq 0$

$\lambda_1(1-q) = 0$ $\lambda_2(1-\beta) = 0$

and there are multiple cases to be considered.

Case 1: $\beta = 0$ and $q = 0$.

This requires $\lambda_1 = 0$ and $\lambda_2 = 0$. To be a maximum, it must be that $\frac{\partial L}{\partial q} \leq 0$ that requires, after substitution, $K\left[(1-p)(1-\pi) - p\right] + (1-p)^2\pi(1-\pi) \leq 0$.

Case 2: β and q both interior, hence $\lambda_1 = 0$ and $\lambda_2 = 0$. The solution is given by the system of $\frac{\partial L}{\partial q} = 0$ and $\frac{\partial L}{\partial \beta} = 0$. The solution is given by

$$q = \frac{2\tau\delta\left[K\left[(1-p)(1-\pi)-p\right]+(1-p)^2\pi(1-\pi)\right]}{4\tau(1-p)^2\pi(1-\pi)\delta^2K+4\tau^2-(p\delta K)^2}$$

$$\beta = \frac{p\delta^2K\left[K\left[(1-p)(1-\pi)-p\right]+(1-p)^2\pi(1-\pi)\right]}{4\tau(1-p)^2\pi(1-\pi)\delta^2K+4\tau^2-(p\delta K)^2}$$

which requires $K\left[(1-p)(1-\pi)-p\right]+(1-p)^2\pi(1-\pi) > 0$.

Case 3: q interior and $\beta = 1$.

This requires $\lambda_1 = 0$ and $\lambda_2 > 0$. Note that, whatever $q < 1$ can be found, when substituting into $\frac{\partial L}{\partial \beta} = 0$ the latter requires $\lambda_2 = p\delta Kq - 2\tau < 0$, hence a contradiction.

Case 4: $q = 1$ and β interior.

This requires $\lambda_1 > 0$ and $\lambda_2 = 0$. The solution for β is $\frac{p\delta K}{2\tau}$ however, when replacing into $\frac{\partial L}{\partial q} = 0$ it simplifies into $\lambda_1 = -p\delta K + \frac{(p\delta K)^2}{2\tau} + (1-p)(1-\pi)\delta K + (1-p)^2\pi(1-\pi)\delta(1-2\delta K) - 2\tau$, hence this is possible when the RHS is greater than 0.

Case 5: $q = \beta = 1$.

This requires $\lambda_1 > 0$ and $\lambda_2 > 0$. When substituting into $\frac{\partial L}{\partial \beta} = 0$ the latter requires $2\lambda_2 = p\delta K - 2\tau < 0$, hence a contradiction.

Case 6: $q = 0$ and β interior.

This requires $\lambda_1 = 0$ and $\lambda_2 = 0$. When replacing, however, the solution for β is only 0, hence a contradiction.

Case 7: q interior and $\beta = 0$.

This requires $\lambda_2 = 0$. However, $\frac{\partial L}{\partial \beta} \leq 0$ only if $q = 0$, hence a contradiction.

Case 8: $q = 0$ and $\beta = 1$.

Then, $\frac{\partial L}{\partial \beta} = 0$ implies a negative λ_2, hence a contradiction.

Case 9: $q = 1$ and $\beta = 0$.

This requires $\lambda_2 = 0$ and $\frac{\partial L}{\partial \beta} \leq 0$. However, when substituting, we obtain $\frac{\partial L}{\partial \beta} = p\delta K > 0$, hence a contradiction.

To sum up, the ex ante maximization problem has just three solutions. $q = \beta = 0$ if $K\left[(1-p)(1-\pi) - p\right] + (1-p)^2\pi(1-\pi) \leq 0$.

$$q = \frac{2\tau\delta\left[K\left[(1-p)(1-\pi)-p\right] + (1-p)^2\pi(1-\pi)\right]}{4\tau(1-p)^2\pi(1-\pi)\delta^2 K + 4\tau^2 - (p\delta K)^2}$$

$$\beta = \frac{p\delta^2 K\left[K\left[(1-p)(1-\pi)-p\right] + (1-p)^2\pi(1-\pi)\right]}{4\tau(1-p)^2\pi(1-\pi)\delta^2 K + 4\tau^2 - (p\delta K)^2}$$

if $K\left[(1-p)(1-\pi) - p\right] + (1-p)^2\pi(1-\pi) > 0$ and $-p\delta K + \frac{(p\delta K)^2}{2\tau} + (1-p)(1-\pi)\delta K + (1-p)^2\pi(1-\pi)\delta(1-2\delta K) - 2\tau \leq 0$.

$$q = 1$$

$$\beta = \frac{\delta K}{4\tau}$$

if $K\left[(1-p)(1-\pi) - p\right] + (1-p)^2\pi(1-\pi) > 0$ and $-p\delta K + \frac{(p\delta K)^2}{2\tau} + (1-p)(1-\pi)\delta K + (1-p)^2\pi(1-\pi)\delta(1-2\delta K) - 2\tau > 0$.

As the equilibrium level of β is always 0, $K\left[(1-p)(1-\pi) - p\right] + (1-p)^2\pi(1-\pi) > 0$ guarantees equilibrium (strict) under-attention to the state of the world.

The last part of the proposition is a simple comparison between $q^{p,**}$ and the interior ex ante optimal q, looking for parameters where the latter is smaller. ■

Proposition C1.3 is just the equivalent of Proposition 1.5 for generic p.

D1 Fixed total level of attention

In this appendix the voter is allowed to allocate attention between the action and the state. I assume that the total amount of attention is fixed, exogenous and equal to α.

Hence, the voter is now allowed to choose q and β with the constraint that $q+\beta=\alpha$. Clearly, the main features of the equilibrium are those summarized by Proposition 1.4, and I keep Assumption 1.2.

The voter's problem is now to $max_{q,\beta}\tilde{\tilde{W}}$ subject to $q+\beta=\alpha\leq 1$, where

$$\begin{aligned}\tilde{\tilde{W}} &= q\beta\{Pr(x=B,s=B)(\hat{\pi}_{B,B}V_C+(1-\hat{\pi}_{B,B})V_D)+Pr(x=A,s=B)\Gamma \\ &+ Pr(x=A,s=A)(r_{A,A}(\hat{\pi}_{A,A}V_C+(1-\hat{\pi}_{A,A})V_D)+(1-r_{A,A})\Gamma) \\ &+ Pr(x=B,s=A)(r_{B,A}(\hat{\pi}_{B,A}V_C+(1-\hat{\pi}_{B,A})V_D)+(1-r_{B,A})\Gamma) \\ &+ q(1-\beta)\{Pr(x=B)(\hat{\pi}_{B,\emptyset}V_C+(1-\hat{\pi}_{B,\emptyset})V_D)+Pr(x=A)\Gamma\} \\ &+ (1-q)\Gamma\end{aligned}$$

In equilibrium $\hat{\pi}_{A,A}=\hat{\pi}_{B,A}=\pi$ so, substituting the constraint, this can be rewritten as

$$\begin{aligned}\tilde{\tilde{W}} &= q(\alpha-q)\{Pr(x=B,s=B)(\hat{\pi}_{B,B}V_C+(1-\hat{\pi}_{B,B})V_D) \\ &+ (1-Pr(x=B,s=B))\Gamma\} \\ &+ q(1-(\alpha-q))\{Pr(x=B)(\hat{\pi}_{B,\emptyset}V_C+(1-\hat{\pi}_{B,\emptyset})V_D)+Pr(x=A)\Gamma\} \\ &+ (1-q)\Gamma\end{aligned} \tag{D1.1}$$

Proposition D1.1 clarifies how attention is allocated, in equilibrium.

Proposition D1.1 *In equilibrium, the voter chooses* $q^{***}=\alpha$ *and* $\beta^{***}=0$.

Proof of Proposition D1.1.

While choosing the equilibrium level of attention, the voter differentiates (D1.1) keeping as a constant the pandering probabilities of the incumbent. Noticing that, once replacing the actual values and taking into account that $\gamma_B^C=1$ and $\gamma_A^C=\gamma_A^D$, $Pr(x=B,s=B)(\hat{\pi}_{B,B}V_C+(1-\hat{\pi}_{B,B})V_D)+(1-Pr(x=B,s=B))\Gamma=Pr(x=B)(\hat{\pi}_{B,\emptyset}V_C+(1-\hat{\pi}_{B,\emptyset})V_D)+Pr(x=A)\Gamma$, and defining for simplicity $K=\frac{1}{\bar{u}_1}\left(\frac{\bar{u}_2}{2}+E\right)$, the objective function simplifies to

$$\tilde{\tilde{W}}=\frac{1}{4}\pi(1-\pi)\delta(1-\gamma_B^D)q$$

and as a consequence, differentiating with respect to q,

$$\frac{1}{4}\pi(1-\pi)\delta(1-\gamma_B^D) = 0$$

which leads to a corner solution where q^{***} is as big as possible, that is, equal to α. As a consequence, $\beta^{***} = 0$. ■

Basically, the point is the same as in Section 1.3.2. As attention to the state does nothing for the voter in terms of selection, then for every conjectured γ_B^D she best responds allocating all the attention to the action (i.e., choosing $q = \alpha$), and the politician is conjecturing this correctly in equilibrium.

In this set up, the ex ante welfare function is given by

$$\tilde{\tilde{W}}^{ex} = \frac{1}{2}\left[1 - F_1\left(\delta\left(\frac{\bar{u}_2}{2} + E\right)q\left(1-(\alpha-q)\right)\right) + \pi + (1-\pi)F_1\left(\delta\left(\frac{\bar{u}_2}{2} + E\right)q\right)\right] + \tilde{\tilde{W}} \tag{D1.2}$$

where the first part of the equation is simply the present part of the welfare function, given by the policy choices of the current incumbent. Lemma D1.1 establishes that the objective function is concave in q.

Lemma D1.1 *For every $q \in [0,\alpha]$ $\frac{\partial^2 \tilde{\tilde{W}}^{ex}}{\partial q^2} \leq 0$.*

Proof of Lemma D1.1.

First note that (D1.2) can be simplified to

$$\frac{1}{2}\left[1 - \delta K q(1-(\alpha-q)) + \pi + (1-\pi)\delta K q\right] + \frac{1}{4}\pi(1-\pi)(1-\delta K q) + \Gamma \tag{D1.3}$$

Its second derivative with respect to q is given by

$$-\delta K - \frac{1}{2}\pi(1-\pi)\delta^2 K < 0$$

as required. ■

From this I solve the optimization problem, giving the following result:

Proposition D1.2 *If* $K > \frac{\pi(1-\pi)}{2(\alpha+\pi+\alpha\pi(1-\pi)\delta)}$ *then the ex ante optimal allocation of attention is such that* $\beta^{ex} > 0$ *and* $q^{ex} < \alpha$.

Proof of Proposition D1.2.

Using lemma D1.1, it is sufficient to look at the first order conditions of (D1.3) for maxima. Hence, the ex ante optimal q solves

$$-\frac{1}{2}\delta K(1-\alpha+2q)+\frac{1}{2}(1-\pi)\delta K+\frac{1}{4}\pi(1-\pi)\delta(1-2\delta Kq)$$

This can be written as:

$$\frac{1}{2}\delta K(\alpha-\pi)+\frac{1}{4}\pi(1-\pi)\delta=\delta Kq\left[1+\frac{1}{2}\pi(1-\pi)\delta\right]$$

As the RHS is strictly increasing in q and the RHS is strictly decreasing in q, the maximum is unique.

To have equilibrium misallocation of attention, it is enough that the maximum is below α, and this happens whenever $LHS(q=\alpha) < RHS(q=\alpha)$. Replacing values from above and writing solving for K, the condition simplifies to

$$K > \frac{\pi(1-\pi)}{2\left(\alpha+\pi+\alpha\pi(1-\pi)\delta\right)}$$

■

Intuitively, it may be that - even ex ante - it is optimal to allocate all the attention to the action. However, Proposition D1.2 establishes that while the extreme allocation always occurs in equilibrium, this may not be the case ex ante, if the positive effect of β in reducing pandering is taken into account and sufficiently big.

E1 Relaxing the assumptions of Section 1.3.2

If the explicit functional form assumed in 1.3 is replaced with a generic $\tau C(q,\beta)$, the condition for the negative semidefiniteness of the Hessian becomes

$$\frac{1}{4}\pi(1-\pi)\delta(1+\delta)K\tau C''_{\beta\beta}+\tau^2 C''_{\beta\beta}C''_{qq}-\left(\frac{1}{2}\delta K-\tau C''_{\beta q}\right)^2\geq 0$$

which holds as long as the second derivative of $C(q,\beta)$ is sufficiently big everywhere. Obviously, it is not possible to characterize closed form solutions for a generic cost function.

Assumption 1.4 is a sufficiency one: it can be relaxed without losing the concavity. Note that concavity is violated for sufficiently small values of τ. In this case, even though the full analysis is complicated, it seems likely that the solution would move toward $\beta = q = 1$.

F1 Attention, pandering and competence

Set up

This appendix tests the robustness of my result to a different class of political agency models, where politicians differ in their level of competence rather than in their alignment with the interests of the voters. The set up I use is similar to Prat (2005) and loosely based on Canes-Wrone et al (2001). In this appendix, $s_t \in \{A, B\}$ with $Pr(s_t = A) = p > \frac{1}{2}$, hence state A is the ex ante most likely state.

Politicians can be of two types: $\theta \in \{H, L\}$ where $Pr(\theta = H) = \pi$. Both types have the same utility as the consonant politician of the main body of the paper. The difference here is that type H politicians are highly competent and they observe perfectly s_t before taking the action. As a further simplification, I assume that H types always choose $x_t = y_t$.[46]. If $\theta = L$, instead, the knowledge of the state is not perfect. Formally, I assume that, rather than observing s_t, the politician observes a signal $y_t \in \{A, B\}$

[46] This is equivalent to assuming that their distribution of u_1 has a sufficiently high lower bound. It simplifies a lot the equilibrium structure in the present set up, where the utility from matching the state is random from the point of view of the voters. I cannot use a set up closer to Prat (2005) or Canes-Wrone et al. (2001), where u is common knowledge, because the equilibrium with pandering involves a mixed strategy where attention does not enter in the politician's randomization.

such that $Pr(y_t = s_t|\theta = L) = g \in (p, 1)$ and $Pr(y_t = s_t|\theta = H) = 1$. Hence, y_t is, for the low type, informative about the true state and stronger than the prior (hence $Pr(s_t = B|y_t = B) > Pr(s_t = A|y_t = B)$). Without re-election concerns, the low type politician should just follow the signal. However, as noticed by Ashowrth and Shotts (2010), the asymmetric burden of proofs implied by choosing the ex ante popular action vis-a-vis the other one creates pandering incentives. Everything else is the same as in the main body of the paper.

An equilibrium with pandering

Defining for simplicity the second period utility of a low quality politician $V_2^L = g\mathbb{E}(u_2) + E$ and as $\gamma_y^H = Pr(x_1 = A|\theta = H, y_1)$ and $\gamma_y^L = Pr(x_1 = A|\theta = L, y_1)$, I first derive sufficient conditions for an equilibrium with pandering.

Proposition F1.1 *If* $\gamma_A^H = 1$, $\gamma_B^H = 0$ *and* $\frac{1}{\bar{u}_1} V_2^L < \frac{(2p-1)(1-g)(g-p)}{(g+p-2pg)^2}$, *there is a "pandering equilibrium" where* $r_{A,\emptyset} = 1$, $r_{B,\emptyset} = 0$, $r_{A,A} = 1$, $r_{B,A} = 0$, $r_{B,B} = 1$ *and* $r_{A,B} = 0$, *with* $\gamma_A^L = 1$ *and* $\gamma_B^L \in [0, 1)$.

Proof of Proposition F1.1.

First of all, note that as usual the voter keeps the incumbent if $\hat{\pi}_{\tilde{x},\tilde{s}} > \pi$. Moreover, by Bayes' rule,

$$\hat{\pi}_{A,\emptyset} \geq \pi \Rightarrow p\gamma_A^H + (1-p)\gamma_B^H \geq p(g\gamma_A^L + (1-g)\gamma_B^L) + (1-p)(g\gamma_B^L + (1-g)\gamma_A^L)$$

$$\hat{\pi}_{B,\emptyset} \geq \pi \Rightarrow p\gamma_A^H + (1-p)\gamma_B^H \leq p(g\gamma_A^L + (1-g)\gamma_B^L) + (1-p)(g\gamma_B^L + (1-g)\gamma_A^L)$$

$$\hat{\pi}_{A,A} \geq \pi \Rightarrow \gamma_A^H \geq g\gamma_A^L + (1-g)\gamma_B^L$$

$$\hat{\pi}_{B,A} \geq \pi \Rightarrow \gamma_A^H \leq g\gamma_A^L + (1-g)\gamma_B^L$$

$$\hat{\pi}_{B,B} \geq \pi \Rightarrow \gamma_B^H \leq g\gamma_B^L + (1-g)\gamma_A^L$$

$$\hat{\pi}_{A,B} \geq \pi \Rightarrow \gamma_B^H \geq g\gamma_B^L + (1-g)\gamma_A^L$$

Assuming that $\gamma_A^H = 1$ and $\gamma_B^H = 0$, and hence focusing on the incentives of the low type, define $\hat{p}_y = Pr(s = A|y)$, that is, the updated beliefs on the state being A after signal realization y. Note that, as before, for generic values of q and β conjectured by P,

$$\begin{aligned}\gamma_A^L &= Pr(\mathbb{E}U_P(x=A, y=A, \theta=L) \geq \mathbb{E}U_P(x=B, y=A, \theta=L))\\ &= Pr((2\hat{p}_A-1)u_1 \geq V_2^L\left[q\beta\hat{p}_A(r_{B,A}-r_{A,A})+q\beta(1-\hat{p}_A)(r_{B,B}-r_{A,B})\right]+\\ &+V_2^L\left[q(1-\beta)(r_{B,\emptyset}-r_{A,\emptyset})\right])\end{aligned}$$

$$\begin{aligned}\gamma_B^L &= Pr(\mathbb{E}U_P(x=A, y=B, \theta=L) \geq \mathbb{E}U_P(x=B, y=B, \theta=L))\\ &= Pr((1-2\hat{p}_B)u_1 \leq V_2^L\left[q\beta(1-\hat{p}_B)(r_{A,B}-r_{B,B})+q\beta\hat{p}_B(r_{A,A}-r_{B,A})\right]+\\ &+V_2^L\left[q(1-\beta)(r_{A,\emptyset}-r_{B,\emptyset})\right])\end{aligned}$$

Given the (candidate) equilibrium re-election strategies outlined above, we have that $\gamma_A^L = 1$ and

$$\gamma_B^L = F_1\left[\frac{q}{1-2\hat{p}_B}V_2^L(1-2\beta(1-\hat{p}_B))\right]$$

Given P's actions, it is clear that $\hat{\pi}_{B,B} > \pi$ and $\hat{\pi}_{A,A} > \pi$. However, $\hat{\pi}_{A,\emptyset} > \pi$ requires $\gamma_B^L < \frac{(2p-1)(1-g)}{g+p-2pg}$ and the assumption $\frac{1}{\bar{u}_1}V_2^L < \frac{(2p-1)(1-g)(g-p)}{(g+p-2pg)^2}$ guarantees that this is the case for every q and β.[47] ■

Importantly, in equilibrium γ_B^L is actually a piecewise function:

$$\gamma_B^L = \begin{cases}\frac{1}{\bar{u}_1}\frac{V_2^L}{1-2\hat{p}_B}q(1-2\beta^e(1-\hat{p}_B)) & \text{if } \beta^e \leq \frac{1}{2(1-\hat{p}_B)}\\ 0 & \text{if } \beta^e > \frac{1}{2(1-\hat{p}_B)}\end{cases}$$

where $\hat{p}_B = Pr(s=A|y=B) = \frac{(1-g)p}{p+g-2gp}$.

It is now possible to write down the utility function that V maximizes when choosing attention. Defining V_H and V_L the expected period 2 payoff from having a high or low

[47]Note that, with those re-election strategies, relaxing the assumption about the high type always following the signal would imply $\gamma_A^H = 1$ and $\gamma_B^H = F_1\left[qV_2^H(1-2\beta)\right]$, that is not necessarily zero. Moreover, it is not easy to compare γ_B^H and γ_B^L, because on the one hand $1-2\beta < 1-2\beta(1-\hat{p}_B)$, but on the other hand $V_2^H = \mathbb{E}(u_2)+E > V_2^L$, as the high type does not make mistakes in period 2. Hence, transforming the high quality politician into a signal-abiding robot allows me to avoid all those case-by-case complications focusing on the interaction of attention and pandering incentives.

quality politician in power in that period, $\Gamma = \pi V_H + (1-\pi)V_L$ and $\Delta V = V_H - V_L = \delta(1-g)$, V chooses q and β to maximize:

$$\pi + (1-\pi)\left[g - (g-p)\gamma_B^L\right] + q\beta[Pr(x=B,s=B)(\hat{\pi}_{B,B}V_H + (1-\hat{\pi}_{B,B})V_L) +$$
$$+Pr(x=A,s=A)(\hat{\pi}_{A,A}V_H + (1-\hat{\pi}_{A,A})V_L) + (1 - Pr(x=A,s=A) - Pr(x=B,s=B))\Gamma] +$$
$$+q(1-\beta)\left[Pr(x=A)(\hat{\pi}_{A,\emptyset}V_H + (1-\hat{\pi}_{A,\emptyset})V_L) + (1-Pr(x=A))\Gamma\right] + (1-q)\Gamma - \tau C(q,\beta) \tag{F1.1}$$

The first part of (F1.1) is the present welfare and the rest is the expected period 2 welfare (taking into account re-election strategies) and finally the attention cost (on which I keep the same assumptions as in the rest of the paper). Collecting terms in $q\beta$ and in q, applying Bayes' rule and substituting in $\gamma_A^H = 1$, $\gamma_B^H = 0$ and $\gamma_A^L = 1$, this can be simplified as

$$\pi + (1-\pi)\left[g - (g-p)\gamma_B^L\right] + q\beta\left[(1-\pi)\pi 2(1-p)\Delta V(1 - g + g\gamma_B^L)\right] +$$
$$+ q\left[\pi(1-\pi)\Delta V((2p-1)(1-g) - \gamma_B^L(p(1-g) + (1-p)g))\right] + \Gamma - \tau C(q,\beta) \tag{F1.2}$$

Obviously, γ_B^L is to be treated exogenously when deriving the equilibrium level of attention and endogenously in the ex ante case. The maximization with respect to two different variables and the non linearity of the resulting system of two first order conditions, together with the kink in γ_B^L, imply that there are some serious tractability issues.

However, there are already some interesting results. First of all, the effect of q and β on pandering (captured by γ_B^L in this case) is analogous to the main body of the paper (as long as γ_B^L is positive). Attention to the action increases pandering while attention to the state reduces it. Moreover, pandering has a negative effect on present welfare and it is easy to show that as long as γ_B^L is positive, the effect of pandering on future welfare is negative as well. Note that this second effect can be decomposed in two parts: when both the action and the state are observed pandering is useful for selection, as it consists in the low quality incumbent choosing the wrong action (and hence he can be spotted and replaced). Instead, when only the action is observed, pandering affects selection negatively, as it makes the low quality incumbent more similar to the high quality one. For sufficiently low β, and importantly for every β such that pandering happens with positive probability, the negative effect dominates.

Given this, the basic effects of the different types of attention captured by the paper are robust to this different model of political agency: the equilibrium choice focuses on selection, ignoring the indirect effect of the two types of attention on pandering, and as a consequence there is a tendency to choose too high a level of attention to the actions and too small a level of attention to the state of the world.

Whether this happens or not in the equilibrium allocation depends on cross-effects as well, and in this set up they can be quite complex, in the ex ante optimal choice. An important difference from the model used in the main body of the paper is that any complementarity between the two types of attention comes from the selection part of the welfare function, rather than from the first period part. The reason is that there is not good pandering in this set up, as the sole pandering is that of the low quality player ignoring relevant information for re-election concerns. Hence, knowing the state on top of the action helps in selecting a better period 2 incumbent. In contrast, discouraging pandering makes this ability to select on the basis of both action and state less effective, as it makes the low quality politician more similar to the high quality type.

Generally speaking, the importance of those cross-effects makes the general problem with endogenous attention difficult to solve fully. Cross effects are simple in the case of the fixed (and free) amount of attention to be allocated, hence it is clear from there that the equilibrium q is higher than the ex ante optimal level and hence the equilibrium β is too low (but not zero, as this is a feature of a model of political agency with "ideological" incumbent, that sometimes has interests aligned with the consonant politician). Given the tractability issues outlined above, I leave the full solution of the model with costly endogenous allocation of the two types of attention (the equivalent of section 1.3.2) to further research, focusing here on the tractable fixed amount of attention case. However, it is possible to show through numerical simulation that there are parameter values where the cross effects are sufficiently strong to lead to equilibrium over-attention with respect to both the action and the state.

Fixed total level of attention

As in Appendix D1, I assume that $q+\beta=1$, that is, V allocates a fixed (and free) unit of attention between action and state. As a consequence, substituting the constraint in the objective function, in equilibrium V chooses q to maximize $\tilde{\tilde{\tilde{W}}}$, where

$$\tilde{\tilde{\tilde{W}}} = q(1-q)\left[(1-\pi)\pi 2(1-p)\Delta V(1-g+g\gamma_B^L)\right] + \\ + q\left[\pi(1-\pi)\Delta V((2p-1)(1-g)-\gamma_B^L(p(1-g)+(1-p)g))\right] + \Gamma$$

taking γ_B^L as given. Note that, defining q^e the conjectured value of q,

$$\gamma_B^L = \begin{cases} \frac{1}{\bar{u}_1}\frac{V_2^L}{1-2\hat{p}_B}q^e(1-2(1-q^e)(1-\hat{p}_B)) & \text{if } q^e \geq \bar{q} \\ 0 & \text{if } q^e < \bar{q} \end{cases}$$

where $\bar{q} = \frac{g-p}{2(1-p)g}$.

First of all, I implicitly characterize the equilibrium attention choice.

Lemma F1.1 *The equilibrium level of attention is unique, strictly positive and given by:*

1. *the implicit solution of $H(\gamma_B^L(q^*)))+Q(\gamma_B^L(q^*)))=2q^*H(\gamma_B^L(q^*)))$, where where $H(\gamma_B^L(q^*)))=\left[(1-\pi)\pi 2(1-p)\Delta V(1-g+g\gamma_B^L(q^*))\right]$ and $Q(\gamma_B^L(q^*)))=\left[\pi(1-\pi)\Delta V((2p-1)(1-g)-\gamma_B^L(q^*)(p(1-g)+(1-p)g))\right]$, if $\frac{1}{\bar{u}_1}V_2^L > \frac{(1-g)(4p-3)(g-p)}{(p+g-2pg)(p+g-2pg+2g(1-p))}$;*
2. *1 otherwise.*

Proof of Lemma F1.1.

From the point of view of V, $\frac{\partial^2 \tilde{\tilde{\tilde{W}}}}{\partial q^2} < 0$. As a consequence, any interior equilibrium is, by first order condition and the rationality condition, the fixed point q^* of

$$H(\gamma_B^L(q^*))) + Q(\gamma_B^L(q^*))) = 2q^*H(\gamma_B^L(q^*))) \tag{F1.3}$$

where $H(\gamma_B^L(q^*))) = \left[(1-\pi)\pi 2(1-p)\Delta V(1-g+g\gamma_B^L(q^*))\right]$ and
$Q(\gamma_B^L(q^*))) = \left[\pi(1-\pi)\Delta V((2p-1)(1-g) - \gamma_B^L(q^*)(p(1-g)+(1-p)g))\right]$.

The LHS of (F1.3) is flat in q below $\bar{q}$ and then it is increasing, while the RHS is always strictly increasing and both are continuous everywhere with a kink at $q = \bar{q}$. Obviously $LHS(q=0) > 0 = RHS(q=0)$. Moreover, by substitution it is easy to see that $LHS(q=1) < RHS(q=1)$ implies $\frac{1}{\bar{u}_1}V_2^L > \frac{(1-g)(4p-3)(g-p)}{(p+g-2pg)(p+g-2pg+2g(1-p))}$.
Trivially, $\frac{\partial RHS}{\partial q} > \frac{\partial LHS}{\partial q}$ for $q < \bar{q}$. Moreover, by substitution, a sufficient condition for $\frac{\partial RHS}{\partial q} > \frac{\partial LHS}{\partial q}$ when $q \geq \bar{q}$ is $2(1-p)g(1-2q) - (p+q-2pq) \leq 0$. Noticing that q has to be at least $\frac{g-p}{2(1-p)g}$, the condition simplifies to $p-g$, which is always negative.
Hence, the RHS is always steeper in q than the LHS, it starts from below and ends above if $\frac{1}{\bar{u}_1}V_2^L > \frac{(1-g)(4p-3)(g-p)}{(p+g-2pg)(p+g-2pg+2g(1-p))}$. As a consequence, there is one interior solution of Equation (F1.3). If instead $\frac{1}{\bar{u}_1}V_2^L \leq \frac{(1-g)(4p-3)(g-p)}{(p+g-2pg)(p+g-2pg+2g(1-p))}$ the unique solution is $q = 1$. ■

Define q^* as the equilibrium level of attention to the action. Looking now at the ex ante case, it is quite clear that the objective function has a kink at $q = \bar{q}$. However, the problem is concave both above and below the kink, hence the maximum is either at one of the corners or at the kink, or it can be obtained via first order conditions. The objective function, using (F1.2), is

$$\tilde{\tilde{\tilde{W}}}^{ex} = P(\gamma_B^L(q)) + q(1-q)H(\gamma_B^L(q)) + qQ(\gamma_B^L(q)) + \Gamma \tag{F1.4}$$

where $P(\gamma_B^L(q)) = \pi + (1-\pi)\left[g - (g-p)\gamma_B^L\right]$ and $H(\gamma_B^L(q))$ and $Q(\gamma_B^L(q))$ are as above. Of course now V can differentiate with respect to γ_B^L as well. The next proposition summarizes the findings, showing that the equilibrium level of attention to the action is never too low. It can only be the right level or too high, depending on parameters of the model. As a consequence, the equilibrium level of attention to the state can only be either correct or too low.

Proposition F1.2 *There is never equilibrium under-attention to the action. In particular:*

1. *If parameters are such that $q^* \leq \bar{q}$, the ex ante optimal level of attention is the*

same as the equilibrium level;

2. *If parameters are such that $q^* \in (\bar{q}, 1)$, there is equilibrium over-attention to the state and under-attention to the action;*

3. *If parameters are such that $q^* = 1$, there is equilibrium over-attention to the state and under-attention to the action as long as the ex ante optimal level of q is interior. Otherwise, the equilibrium level of attention and the ex ante optimal one are the same.*

Proof of Proposition F1.2.

First of all, from Equation (F1.4), the second order conditions are trivially satisfied if $q < \bar{q}$, as $\gamma_B^L = 0$ there. When $q \geq \bar{q}$,

$$\frac{\partial \tilde{\tilde{\tilde{W}}}^{ex}}{\partial q} = H + Q - 2qH + \frac{\partial P}{\partial \gamma_B^L}\frac{\partial \gamma_B^L}{\partial q} + q\frac{\partial \gamma_B^L}{\partial q}\left[\frac{\partial H}{\partial \gamma_B^L} + \frac{\partial Q}{\partial \gamma_B^L}\right] - q^2\frac{\partial \gamma_B^L}{\partial q}\frac{\partial H}{\partial \gamma_B^L}$$

By substitution, $\frac{\partial P}{\partial \gamma_B^L}\frac{\partial \gamma_B^L}{\partial q} + q\frac{\partial \gamma_B^L}{\partial q}\left[\frac{\partial H}{\partial \gamma_B^L} + \frac{\partial Q}{\partial \gamma_B^L}\right] - q^2\frac{\partial \gamma_B^L}{\partial q}\frac{\partial H}{\partial \gamma_B^L} < 0$ implies $-(g-p)(1 - q\pi\delta(1-g)(g-p)) - q^2\pi 2(1-p)\delta(1-g)g < 0$ that is always true.

Moreover, from the derivation on the different slopes of the RHS and LHS of the equilibrium fixed point condition, I know that $\frac{\partial}{\partial q}(H + Q - 2qH) < 0$. A sufficient condition for concavity is that, defining $Z = \frac{\partial P}{\partial \gamma_B^L}\frac{\partial \gamma_B^L}{\partial q} + q\frac{\partial \gamma_B^L}{\partial q}\left[\frac{\partial H}{\partial \gamma_B^L} + \frac{\partial Q}{\partial \gamma_B^L}\right] - q^2\frac{\partial \gamma_B^L}{\partial q}\frac{\partial H}{\partial \gamma_B^L}$, $\frac{\partial Z}{\partial q} < 0$. Taking into account that, for $q \geq \bar{q}$, $\frac{\partial \gamma_B^L}{\partial q} > 0$, this requires $(1-\pi)\pi\Delta V(g - p) - 2q(1-\pi)\pi 2(1-p)\Delta V g < 0$. The condition simplifies to

$$4(1-p)gq > (g-p) \tag{F1.5}$$

Noticing that I am considering the case of $q \geq \bar{q}$ (for $q < \bar{q}$ concavity is trivially satisfied), I substitute $q = \frac{g-p}{2(1-p)g}$ in (F1.5). This simplifies to $2 > 1$.

In terms of comparison between the equilibrium attention allocation and the ex ante optimal choice, we consider different cases. For $q \leq \bar{q}$, the equation that defines the equilibrium and the first order condition of the ex ante maximization are exactly the same, because $\gamma_B^L = 0$. For $q > \bar{q}$, the equilibrium solves $H + Q = 2qH$ and the ex ante optimum solves $H + Q + Z = 2qH$, with $Z < 0$. So, the LHS of the equation that defines the ex ante optimal attention is the same as the equilibrium equation for $q \leq \bar{q}$ and strictly below the other one otherwise, while the RHS is always the same.

As a consequence, if parameters are such that $q^* \leq \bar{q}$, then this would be the solution of the ex ante maximization as well. If instead parameters are such that $q^* \in (\bar{q}, 1)$, then the ex ante optimal attention to the action is smaller than the equilibrium one (the RHS is the same but the LHS is smaller because of Z) and as a consequence the ex ante optimal attention to the state is higher than the equilibrium choice.

Finally, if parameters are such that $q^* = 1$, it can be either that $H(q = 1) + Q(q = 1) + Z(q = 1) < 2H(q = 1)$, so the ex ante optimal attention to the action is smaller than the equilibrium level, or that $H(q = 1) + Q(q = 1) + Z(q = 1) \geq 2H(q = 1)$ and the two allocations are identical. ■

Chapter 2

The Price of Silence. Media Competition, Capture and Electoral Accountability

2.1 Introduction

A free media has been seen as a powerful guarantor of political accountability, both theoretically (e.g. Besley 2006) and empirically (e.g. Ferraz and Finan 2008, Snyder and Stromberg 2010). However, the media may be powerful enough to determine an electoral outcome and to promote a bad candidate, even when voters are fully rational and "Bayesian" (Prat 2014, Anderson and McLaren 2012, Enikolopov et al. 2011). As a consequence, an incumbent politician may be interested in controlling what media outlets report, to present a positive image to voters and to stay in power. This paper looks at the effect of competition in influencing the incentives towards media capture, with novel results.

The current literature stresses the positive effect of competition on media freedom: increasing the number of outlets means that the bad politician has more publishers to deal with, so capture is more costly. However, there is a more subtle effect, because the competitive pressure decreases profits, and firms with smaller financial margins may be more willing to sacrifice editorial independence for political money.[1] Hence, they

[1] Although in a different context (i.e. advertisement driven media bias), this is consistent with the

are cheaper to capture.

Overall, the direction of this trade off is not trivial: the "positive" view of the role of competition in deterring capture can be questioned if more competition means smaller (and financially weaker) media outlets, less able to inform the public opinion and to resist to political pressure. Media outlets may be more numerous, but are they freer? Under some conditions, more competition can actually harm the media's independence from political influence, as the empirical section of this paper suggests. This highlights the need for a deeper understanding of the forces behind this trade off.

This paper makes two contributions: on the empirical side, it is the first[2] to point out the existence of a robust negative relationship between competition in the media market and media freedom from political influence. On the theoretical side, the model provides an explanation for the counter-intuitive empirical results, showing that potential risks to media independence from the political power are high not only when competition is too little, but also when it is too much. Moreover, the model highlights that the standard "positive" result of competition and capture relies on restrictive assumptions about voters' behaviour and media outlets' profits. From a policy perspective, both the empirics and the theory stress the risk, in terms of editorial independence, of excessive competition in the media sector.

More specifically, the theoretical model is a natural extension of the seminal contribution by Besley and Prat (2006). It shows that, relaxing some of the assumptions in a fairly natural way,[3] high competition is actually bad for media freedom, as the cost of media capture is driven to zero as the number of outlets goes to infinity. Moreover, the relationship can be non-monotonic, overall, meaning that media capture occurs when competition is either very weak or very strong. Intuitively, increasing the number of media outlets can make capture overall more expensive for the politician, as long as the number of outlets that need to be silenced increases in the same way, and the politician pays monopoly profits to every captured outlet. But competition has a decreasing

results of Beattie et al. (2017), where they show that online competition for advertisement increases media bias.

[2]To the best of the author's knowledge.

[3]In this paper, voters are allowed to be heterogeneous in their interest for politically related news and profits are convex in readership. Note that I refer to readers as consumers of media contents for simplicity. However, the paper is agnostic with respect to the type of media firm (newspapers, televisions, websites etc).

effect on the influence that each individual outlet has on voters: basically, with more competition every outlet is able to inform a smaller fraction of the electorate. In fact, even established media outlets have limits in their ability to inform voters, as the recent spreading of fake news highlights. Hence, when there are enough outlets, the politician may be willing to allow for some free media outlets, since they will not be powerful enough to change electoral outcomes. This decreases the profits that captured outlet would make by rejecting the bribe from the politician, hence they are cheaper. The overall effect of this trade off induced by competition (more outlets to be silenced, but each of the is cheaper) depends on the relationship between readership and profits. When it is convex (I provide a micro-foundation where media market is modelled as a two sided one) then I show the existence of an equilibrium where high competition makes capture cheaper, overall.

This result may sound surprising and somewhat counter-intuitive. But does it also make sense in the real world? Of course, it is very hard to measure media capture empirically, since "influence" is difficult to measure (and it can be generated in many different ways: from bribing to buying advertising space). The empirical section of this paper suggests that high competition in the mass media market is strongly correlated with a *reduction* in media freedom from political influence. The identification strategy exploits the staggered digitization of terrestrial television in Europe: a technological change that allowed for a more efficient use of the spectrum and, hence, for the entry of new players in the market. In an event study analysis, this paper shows that the sign of the relationship between digitization (and hence competition) and the freedom of the media from political influence is negative and strongly significant. And this happens precisely in countries with high level of pre-treatment competition, suggesting that media capture is easier in those places.

Also, importantly, it is possible to find anecdotal confirmation for the two steps leading to this result: firstly, competition has a negative effect on outlets' profitability. Secondly, financially weak outlets may be more willing to accept political influence.[4]

[4]Additional anecdotal evidence comes from Drago et al. (2014). They find that the entry of newspapers increases the re-election probability of the incumbent: this is at odds with Besley and Prat (2006), but it is consistent with the results of this paper. Moreover, they find that "competition matters", meaning that the effect of entry does not disappear as more newspapers participate in the market, and in this model I will try to go into more depth than the standard monopoly *versus* duopoly approach.

For the first step, Cagé (2014) has precise estimations about the effects of an exogenous increase in media competition on profitability, noticing that "entry reduces the circulation of incumbent newspapers by nearly 25%", and this implies a 20 to 36 per cent reduction in revenues, 14 to 29 per cent reduction in size and a 9 to 13 per cent reduction in the share of hard news. Moving to the second step, many different sources show that the media is more easily influenced when revenues are low. The 2015 edition of the Media Sustainability Index (IREX 2015), for example, stresses that "[o]verall the issue seems to be that media have been weakened by a poor economy and been preyed upon by political money, or political pressure has weakened the economic environment in which media operate, thus *making it easier for political money to distort the market and put independent media at a strong disadvantage*" (IREX 2015, p. ix, italics added). One example is Bulgaria, where "most traditional media operate at a financial loss, which leads to compromises with editorial independence. With few exceptions, the big advertisers enjoy complete media support. As public institutions remain the biggest advertisers, any government regardless of its political affiliation receives media support" (IREX 2015, p. xi, italics added). Others are Albania,[5] FYR of Macedonia[6] and Turkey[7].

In the theoretical model, voters act as principal while the incumbent (of good or bad type) is the agent. Media outlets may receive a common, verifiable signal about the type of the incumbent, and may be bribed by the bad politician not to reveal it. Hence, outlets face a trade off between publishing the signal and enjoying the higher audience-related revenues and accepting the politician's bribe instead. In contrast to the basic version of Besley and Prat (2006), in this paper not all the voters are interested in politically related news, hence not all of them will become aware of the incumbent's

[5]The website worldaudit.org points out the risk of political influence on financially weak media in Albania: "many independent media outlets are hampered by a lack of revenue. Publishers and media owners tend to dictate editorial policy based on political and economic affiliations, which, together with the employment insecurity journalists face, nurtures a culture of self-censorship".

[6]According to fairpress.eu (Iloska 2014), the country is characterized by a "*large number of broadcast and print outlets* in comparison with its population size". Despite this, "Macedonian media are subject to *severe interference and political pressure*. [...] In the last few years the Government was regularly criticized for its liberal use of promotional advertising, leading to increased financial dependence of media and increasing the number of outlets that favour its position".

[7]The Turkish newspaper "Today's Zaman" (Zibak 2015) recently claimed that "the diversity and abundance of media outlets in Turkey do not necessarily guarantee the presence of a free, independent and pluralistic media". This highlights two things: on the one hand, the existence of free media even in a country with potentially serious problems of media capture; on the other hand, the fact that a very competitive market is not enough to avoid these problems.

type if only a small fraction of the outlets is publishing the signal. Moreover, audience-related revenues are not constrained to be linear in readership. This allows me to check the robustness of the results to different specifications of the profit function, noticing that a profit function convex in readership can emerge from a media market modelled as a two-sided one, where outlets are selling content to the readers and advertising space to a monopolist advertiser, who seeks to place advertisements where the readers are.

In this setting, increasing the number of outlets increases the number of players that the incumbent politician may want to silence (as in Besley and Prat 2006), but it may also *reduce* the cost of capture by reducing the "business-as-usual" profits for captured outlets,[8] which would have to compete with more subjects and hence attract lower audience-related rents. As a consequence, they are more willing to sacrifice their independence in exchange for money. When the profit function is convex in the readership, as the number of outlets goes to infinity, there exists an equilibrium where media competition makes capture easier.

This result follows from the combined effect of the two modifications highlighted before. First of all, when the politician does not need to capture the whole industry, the bribe he has to pay to each captured outlet is no longer the monopoly profits, but it is reduced by the competition effect. As the number of outlets increases, the readership gains from publishing the signal are divided with more and more free outlets, hence the individual bribe to be paid becomes smaller. However, at the same time, there are more outlets to be silenced. The total cost of capture is simply the product between the number of outlets to be silenced and the individual bribe and, when the profit function is convex in the readership, the "reduction in bribe" effect prevails on the "increasing in number" one, reducing the total cost of capture as the number of outlets goes to infinity. Note that both modifications are necessary.[9]

On top of this, the model shows conditions on the parameters where the relationship is non-monotonic, i.e. capture happens when competition is either too little or too

[8] I.e. the profits that each captured outlet would make by rejecting the bribe and publishing the signal.

[9] The online appendix of Besley and Prat (2006) considers an extension where the politician does not need to capture the whole industry, pointing out that this does not change the result. This happens precisely because the linearity assumption on the profit function is maintained.

much. Essentially, this happens when an increase in competition, keeping constant the number of free outlets, increases the total cost of capture (because the politician has to silence more of them). But, at some point, more competition implies that the politician can allow some outlets to be free. Hence, the reduction in the "business-as-usual" profits kicks in and, as competition goes to infinity, the total cost of capture goes to zero. Finally, since the voters' ex ante welfare is strictly decreasing in the ex ante probability of successful media capture, the effect of competition on welfare can be negative as well.

The remainder of this chapter is as follows. Section 2.2 briefly reviews the related literature. Section 2.3 presents an empirical analysis of the digitization of terrestrial television in Europe which, by showing a negative effect of competition on media freedom, motivates the theoretical part of the paper. Section 2.4 describes the setting of the model and comments on the main assumption, while I derive the equilibrium and its consequences on welfare in section 2.5. Section 2.6 provides a two-sided market based microfoundation for the profit function used in the basic model and 2.7 concludes.

2.2 Related literature

This paper contributes to different branches of the economic literature, from the political economy of mass media, to the economic regulation of the media market. There is particular attention on media capture and on the effects that media competition has on political outcomes via media capture.

The relationship between media and politics has been widely studied in the political economy literature, and detailed and comprehensive reviews are provided by Prat and Strömberg (2013), Strömberg (2015), Gentzkow et al. (2016) and Puglisi and Snyder (2016).

Media capture can be seen as a particular way of endogenizing supply-driven media bias,[10] with the important difference that, in media capture models, the political pref-

[10] The vast political economy literature on media acknowledges the existence of a demand driven media bias as well. See, for example, Strömberg (2004), Mullainathan and Shleifer (2005), Gentzkow and Shapiro (2006), Bernhardt et al. (2008), Chan and Stone (2013), Andina-Diaz and Garcia-Martinez (2016). Models of supply driven media bias (but without an explicit model of media capture) are for example Baron (2006), Anderson and McLaren (2012), Duggan and Martinelli (2011) and Prat (2014), Hafer et al. (2017), Levy et al. (2017).

erences of media outlets are endogenous, and hence determined and constrained by the parameters of the game. Generally speaking, competition is seen as sometimes problematic in models of demand-driven bias, while the current literature considers it a good deterrent in the case of supply-driven bias.

Prat (2016) offers a good summary of the literature on media capture. The seminal paper is Besley and Prat (2006): the basic framework of their model is the starting point for the model in this paper. Given the necessity of capturing the whole industry, in equilibrium the politician will either silence every outlet or none of them, and has to pay monopoly profits to each of them. As a consequence, the cost of media capture (and hence politicians' turnover and voters' welfare) monotonically increases in the number of outlets.

Petrova (2008) and Corneo (2006) look at media capture too, in cases where the capturing entities are interest groups or particular factions of society. Drufuca (2014) extends Besley and Prat (2006), endogenizing voters' informational choices, but she does not focus on the effects of competition and the market structure. Vaidya and Gupta (2016) study the effect of media competition on corruption via media capture, finding a mixed result: when the probability that bad news about the incumbent can be discredited is sufficiently high, then it may be cheaper to capture a duopoly than a monopoly, while the opposite is true when the probability is low. However, this result is obtained in a setting where signals are not perfect and voters are not fully rational, meaning that they ignore the possibility of interactions between the politician and media outlets. Gehlbach and Sonin (2014) look at the link between media bias and a continuous outcome that the government is interested in – political "mobilization"[11] - while Kibris and Kocak (2017) look at how the presence of social media can make capture more or less effective.

It is difficult to find empirical evidence for media capture from the politicians, since it is neither visible nor precisely defined. A very recent review of empirical research on media capture is in Enikolopov and Petrova (2016). Their overall finding is that competition helps in reducing media capture, even if the method for testing the presence

[11] The modelling strategy of this paper shares with Gehlbach and Sonin (2014) the importance given to a precise structure for the advertising market, related to readership. It takes it further by modelling it in a precise two-sided setting. However, the political setting is quite different (and closer to Besley and Prat 2006) since it does not assume a particular "bad" nature of the government, and citizens are interested in being governed by a "good" type.

of media capture is generally indirect.[12]

Probably the most direct insights on how media capture works come from McMillan and Zoido's (2004) study on Fujimori's media control mechanism in Perù, showing the huge costs he had to pay to silence media channels (with respect to members of congress or judges), and from Szeidl and Szucs (2017), pointing out that media capture, in Hungary, occurs through advertisement misallocation. Di Tella and Franceschelli (2011) find a negative correlation between advertisements bought by the government and the space dedicated to scandals in Argentinian newspapers, while Hamilton (2004) relates the emergence of independent (rather than politically affiliated) newspapers to the increasingly important advertising market. Petrova (2011) has a similar result. Finally, Petrova (2008) finds empirical support for the theory that media freedom is decreasing in income inequality in democracies.

Looking at a different type of capture (from advertisers rather than from politicians) Beattie et al. (2017) find that newspapers tend to provide less coverage of car recalls when they involve their advertisers, and that this effect is mitigated by competition in the number of outlets but exacerbated by competition for advertisement.

This paper is distinguished from standard models of media capture because it looks more closely at the media market. In particular, this model is related to those stressing the importance of the "two-sidedness" of the market of media (Argentesi and Ivaldi 2005)[13] and of the role of the advertising market, highlighted for example in Germano and Meier (2013) and Blasco et al. (2015).

Finally, this paper adds to the literature on media market regulation. Since there is no general agreement about what a "healthy" broadcasting sector is (Seabright and von Hagen 2007), this paper highlights that the role of competition, in the context of media capture, may be more complex than the standard answer given by the political economy literature.[14]

[12] For example, Galvis et al. 2016 looks at the effect of competition on media bias in US late 19th century newspapers. This can be due to political capture, but it is more generic than that.

[13] Ellman and Germano (2009), for example, use a two-sided market model in order to see the effect of the advertising industry on media bias, while Godes et al. (2009) and Dukes (2006) use a two-sided model to study how media competition affects pricing strategies and content quality choices respectively. This paper "incorporates" their approach (media outlets selling copies to the readers, i.e. competing on prices, and advertising space to the advertisers, i.e. competing on quantity) in a standard media capture political agency framework.

[14] In particular, Polo (2005) studies whether market incentives are powerful enough to guarantee internal and external pluralism in the media market, finding that the differentiation triggered by

Outside the formal economic modelling, Hollifield (2006) highlights the possibility of a reverse-U shaped relationship between competition and journalistic performance, listing many different channels that may contribute to it. In contrast to that paper, this work focuses solely on media capture, deriving the non-monotonic relationship as a result of a formal, game theoretical model and looking directly at the effects of political and commercial forces.

2.3 Suggestive Empirical Evidence

The empirical identification of a causal effect of competition on media capture is problematic. Firstly, media capture is rarely explicitly observed and recorded. Secondly, the effect of competition in the media market on political outcomes is normally identified using local variations in the number of media outlets (e.g. Drago et al. 2014, Cagé 2014, Galvis et al. 2016), but media capture (or its proxy: media freedom from political influence) is measured at the national level. Hence, instead of using local variations, this paper exploits the staggered implementation, at a country level, of the digitization of terrestrial television, i.e. a policy that was expected to increase the level of competition in the TV market.

As the timing of digitization is decided by each different government, the identification is not perfect, but it is the best that can be done given the restrictions on the outcome variable. I find strong evidence of a robust negative correlation between digitization and media freedom, and anecdotal evidence highlighting the positive relationship between digitization and competition. Moreover, splitting the sample between countries with above- and below-average levels of pre-treatment competition in the TV market[15] shows that this effect comes precisely from those countries with a *high* level of

the market does not necessarily extend to political views, and that internal pluralism is limited by the personal interests of the media owner, while Torrens (2014) looks into the optimal media market structure from a normative perspective, taking into account the possibility of media capture and finding the same results as in Besley and Prat (2006). The basic welfare trade-off is given by the fact that, in his model, more media outlets imply a lower probability of capture but also higher fixed costs to be supported without an increase in the amount of information available. It is interesting to note, however, that Torrens' (2014) main effect of the number of outlets on the possibility of capture is analogous to Besley and Prat (2006) *precisely because* he shares the same assumption about the fact that a single media outlet publishing the news is enough to inform the whole electorate.

[15] Competition is measured as the Herfindahl index calculated using the annual average daily audience market share of the top eight TV channels. Unfortunately, the current dataset does not allow me to keep track of the ownership structure of those channels, or of their content.

competition, while it is not statistically different from zero in low pre-treatment competition countries. In other words, too much competition seems to be correlated with a reduction in freedom of the media from political influence, making capture easier.

2.3.1 Background

Terrestrial television is one of three main modes for transmitting television signals (the other two are satellite and cable). It is the oldest technology and it is based on radio waves. Until the end of the 20th century, terrestrial television was based on analogue transmission that allowed for a limited number of frequencies and hence, a limited number of TV channels. Technological innovation made possible the digitization of terrestrial television, allowing for a more efficient use of the spectrum[16] and, as a consequence, an increase in the number of TV channels that can be transmitted.

Over the first decade of the 21st century, some countries started to replace analogue terrestrial signals with digital ones. In order to harmonize the process within the Continent, the European Commission[17] (EC) set 2012 as the deadline for the full replacement of analogue terrestrial signals with digital terrestrial signals. This process was supposed to reduce transmission costs, free up frequencies and give consumers a wider choice of TV channels.

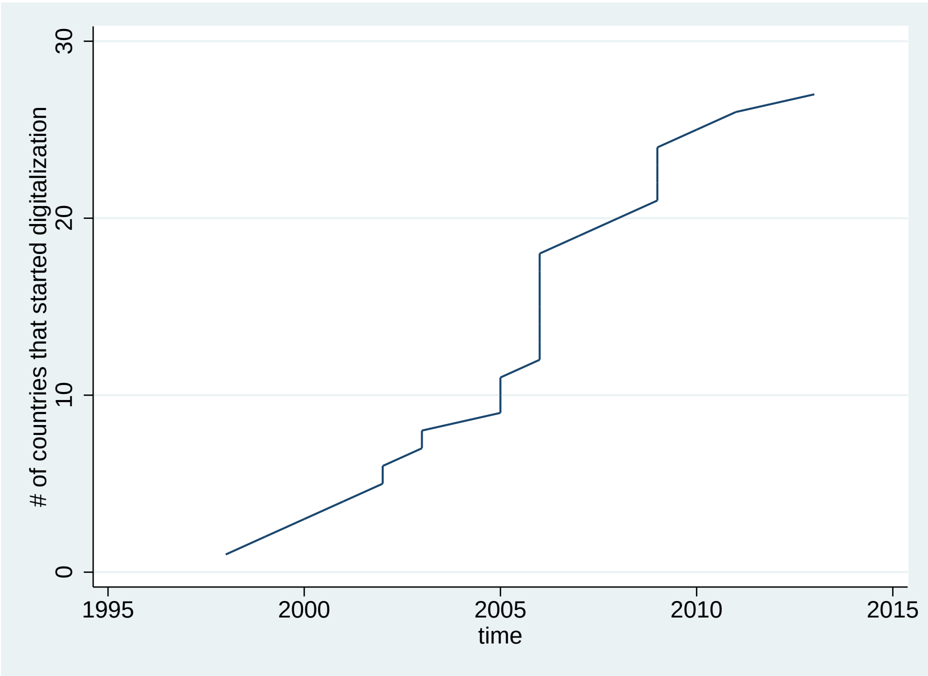
The positive relationship between terrestrial digitization and competition in the mass media market has been mentioned by multiple sources. For example, Barone et al. (2015) notice that Italy had 78 new free to air channels after the transition, "51 of which have no ties with Berlusconi or the public network" (Barone et al. 2015, p. 33), implying that new players are entering in the market. The aforementioned EC communication mentions the increased number of alternatives for consumers, while Kenny et al. (2014) stress the "key role" of digital terrestrial television in securing competition between platforms and content providers.

The empirical strategy exploits the implementation of this European decision to find something as close as possible to the causal effect of digitization of terrestrial television on media freedom from political influence. The fact that digitization happened in a

[16]Technically, multiplexes are able to "squeeze" multiple signals in a smaller space, freeing frequencies for other signals and other multiplexes.
[17]COM(2005) 204.

staggered way between countries is important, with some countries starting the process earlier than others. Figure 2.1 gives an idea of the staggered process, plotting the cumulative number of countries that started digitization as a function of time. Overall, the analysis is a reduced form one, looking at the direct effect of digitization on media freedom in an event study set up.

Figure 2.1: Digitization starting year: cumulative plot



Cumulative number of countries that started the digitization process over time

2.3.2 Empirical strategy

The switch over process took place in a staggered way between and within EU countries,[18] with some regions switching off the analogue terrestrial signal earlier than others. Hence, thanks to the EU Commission deadline, every EU country is treated, while the staggered implementation allows me to exploit a quasi-experimental set up with an event study approach, where the treatment intensity (i.e. the portion of terrestrial television that has already been digitalized) varies over time.

The treatment variable is the penetration of digital terrestrial television, i.e. the number of households receiving TV signals primarily through digital terrestrial over the

[18] On average, there are more than 5 years between the launch of the first digital terrestrial signal in a country and the cessation of analogue terrestrial signals, i.e. the end of the replacement process.

total number of households in a country $\left(\text{formally,} \frac{DTT_{i,t}}{Hous_{i,t}}\right)$.

Note, however, that this treatment is basically the composition of two different effects. First of all, the staggered implementation within countries implies that the replacement ratio, i.e. the fraction of terrestrial television households already receiving digital terrestrial television (DTT), is different at different points in time, and moving between 0 (before the launch of digital terrestrial television) and 1 (when all analogue terrestrial signals are switched off). Define the implementation part of the treatment as $I_{i,t} = \frac{DTT_{i,t}}{TerrTV_{i,t}}$, where $DTT_{i,t}$ and $TerrTV_{i,t}$ are the number of households receiving digital terrestrial television and the number of households receiving terrestrial television[19] in country i at time t respectively.

Secondly, the fraction of the population affected by the reform is different in different EU countries. The reform only affects the digital terrestrial television, so its impact depends on the number of households using primarily terrestrial television, rather than cable or satellite TV. This number is heterogeneous amongst EU countries: some are traditionally "cable" countries (e.g. Belgium, Denmark, the Netherlands), while others are traditionally "terrestrial" countries (e.g. Greece, Italy, Spain). Figure 2.2 shows the penetration of terrestrial television in Europe in 1997.

Moreover, the fraction of households receiving primarily terrestrial television (rather than cable or satellite) changes over time. Formally, the population affected by the reform is defined as $PA_{i,t} = \frac{TerrTV_{i,t}}{Hous_{i,t}}$.

Therefore, the treatment variable can be decomposed as the product of those two effects as follows:

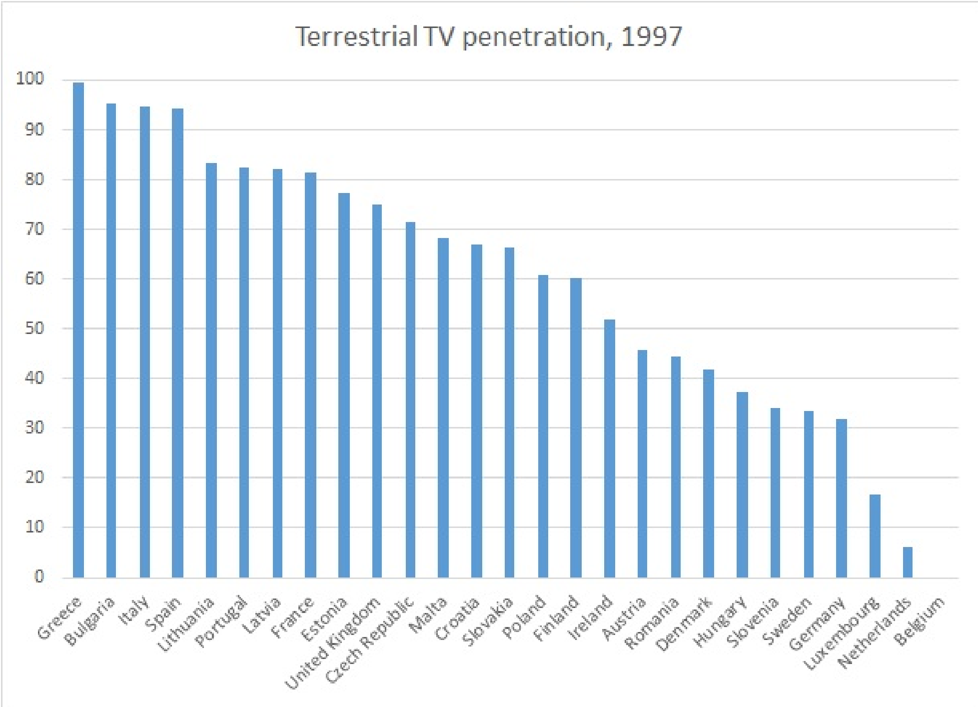
$$\begin{aligned}\frac{DTT_{i,t}}{Hous_{i,t}} &= \frac{\frac{DTT_{i,t}}{Hous_{i,t}}}{\frac{TerrTV_{i,t}}{Hous_{i,t}}} * \frac{TerrTV_{i,t}}{Hous_{i,t}} \\ &= I_{i,t} * PA_{i,t}\end{aligned}$$

Given the precision of the data it is not possible to disentangle the two parts of the treatment.[20] However, the results are robust to different ways of measuring the treat-

[19] I.e. the total number of households minus those receiving TV signals *via* cable, satellite and those not having a television.

[20] In particular, the EAO yearbooks contain a measure of digital terrestrial penetration, but not of terrestrial penetration. This can be obtained as a residual once the number of households receiving primarily cable and satellite is subtracted from the total number of households with a television, but

Figure 2.2: Terrestrial television penetration in 1997



Terrestrial television penetration, in per cent of households with a television, in 1997. The remainder is cable and satellite.

ment (dummy times pre-treatment population affected, as in Card 1992), as discussed in Appendix A2. The main specification used is

$$mf_{i,t} = \lambda_i + \mu_t + \beta_1 DTTpenetration_{i,t} + \gamma^{'} \mathbf{x}_{i,t} + \epsilon_{i,t} \tag{2.1}$$

where $mf_{i,t}$ is the Freedom House media freedom from political influence score, $DTTpenetration_{i,t}$ is defined above[21] and λ_i and μ_t are country and time fixed effects, $\mathbf{x}_{i,t}$ is a vector of time-varying control variables (log of GDP per capita, corruption perception index and low/unique chamber electoral years) and $\epsilon_{i,t}$ is the error term.

In practice, I am comparing the level of media freedom from political influence at different stages of the implementation process, and weighting this effect by the fraction

this variable is extremely noisy, with some peculiar results (e.g. a supposedly negative number of terrestrial television households). This is probably due to different measurement methods for cable and satellite penetration. An exception is the terrestrial penetration in 1997, provided directly by the 1998 Statistical Yearbook.

[21] Given the fact that there are some discrepancies in the data, I use two different measures of DTT penetration. One is calculated manually by dividing the number of households using primarily terrestrial television by the total number of households, both coded using observations from the most recent yearbook available. The other is directly provided in each yearbook and I use, for every year, the observation coming from the next yearbook (i.e. the closest one). In this paper, I report the results using the former, but almost nothing changes if I use the latter.

of population affected by the reform. In this set up β_1, under the assumption of parallel pre-trends and no other policies changing at the same time, measures the causal effect of the digitization of terrestrial television on media freedom.

The main threat to this identification strategy is from the endogeneity of the digitization decision at a country level. In other words, every EU country had to complete the switch over process, but the timing was decided by national governments. Ways to address this concern are discussed in section 2.3.5.

2.3.3 Data

The newly assembled dataset of 28 EU countries combines data from two main sources: Freedom House and the European Audiovisual Observatory. Media freedom from political influence is measured using the score of the political environment from the "Freedom of the Press" index produced every year by the American NGO Freedom House. The score goes from 1 to 40 [22] and evaluates the level of media freedom from political influence in every country, leaving aside economic and legal constraints on what the media is able to report on.[23]

All the variables about television signal penetration and the European audiovisual market (number of households, number of households receiving primarily digital terrestrial television, percentage of DTT penetration, terrestrial, cable and satellite penetration in 1997, country by country DTT launch year and year of completed switch off of analogue signal, daily audience market share of the biggest 8 TV channels, total expenditure on TV advertising) come from the European Audiovisual Observatory yearbooks (1998 edition to 2015 edition). The level of competition in the television market is measured using the Herfindahl Index, using the daily audience market share of the biggest eight channels per country per year.

In addition I collected data for GDP per capita (from the UNESCO Institute for Statistics), corruption (using the corruption perception index from Transparency International) and a dummy equal to 1 for electoral years in the lower/unique legislative

[22] The original score measures the level of influence, hence 0 is a country whose media sector is completely free from political influence and 40 is a country with the highest possible level of influence. However, to give a more intuitive interpretation to the outcome variable as a measure of media freedom, I use 41-original score.

[23] These are captured by the other two environments.

chambers (from the International Foundation for Electoral System). Table 2.1 summarizes these variables.

Table 2.1: Summary statistics

Variable	Obs.	Mean	St. Dev.	Min	Max
Media freedom	392	32.844	3.788	20	40
# of households	392	7339.625	9859.954	118	40129
DTT penetration (%)	379	12.414	17.608	0	77.83
Households with DTT	379	1206.19	3073.866	0	17690
GDP per capita	391	29522.6	13775.04	6507.8	97661.94
Corruption	387	6.293	1.863	2.6	9.9
Election year	392	0.263	0.441	0	1
Herfindahl audience	354	1345.314	560.616	397	3674.93
TV adverts	373	1067.308	1544.048	7	6201

Media freedom is 41 minus the Freedom of the Press score for the Environment B; # of households is the number of households, DTT penetration (%) is the penetration of terrestrial digital calculated using the most recent observation available, Households with DTT is the raw number of households receiving DTT calculated using the most recent observation. GDP per capita is in PPP at current international dollars. Corruption is the Corruption Perception Index, Election year is 1 if elections have been hold for the lower or for the unique chamber of the Parliament (whichever applies). Herfindahl audience uses the daily audience share of the 8 biggest TV channels; TV adverts is the expenditure for TV advertising (mill. of Euro).

2.3.4 Results

The results of regression (2.1) are summarized in Table 2.2.[24] I first estimate the pure fixed effect regression, without any control. Columns (2)-(4) show that the treatment effect is not affected by the inclusion of different controls.[25] Wild bootstrapped standard errors are added to take into account the low number of clusters, and results with group specific time trends are presented in Appendix A2.

Interestingly, a policy change that increases media competition is associated with a *decrease* in media freedom from political influence. Moreover, this result is only marginally affected by the introduction of different control variables. In terms of magnitude, a one standard deviation increase in digital terrestrial penetration is associated with a decrease in media freedom from political influence of 0.12 standard deviations.

This result is already at odds with the existing theoretical literature on media capture and competition. To better understand what is driving it, I split the sample between

[24]Reported results are estimated with a standard OLS regression, hence treating the discrete outcome variable as a continuous one, given that it takes 40 values. This is consistent with the literature (e.g. Petrova 2008).

[25]I do not control for our measure of competition and for our measure of advertisement expenditures because they are potential outcomes of the treatment.

Table 2.2: Regression results, main specification

	Dep. variable: media freedom (1)	(2)	(3)	(4)
DTT penetration	−0.027	−0.026	−0.025	−0.025
	(0.013)*	(0.011)**	(0.011)**	(0.011)**
	[0.014]*	[0.012]**	[0.012]**	[0.012]**
ln(GDP p.c.)		0.244	0.880	0.932
		(2.156)	(2.606)	(2.532)
Corruption			−0.242	−0.253
			(0.300)	(0.305)
Elections				−0.147
				(0.121)
Time FE	Y	Y	Y	Y
Country FE	Y	Y	Y	Y
N	379	377	373	373
R^2	0.483	0.457	0.466	0.467

Dependent variable: media freedom score. DTT penetration uses the most recent observation. Column 1 is the pure DID set up without controls. Columns 2-4 add one control each. Country level clustered standard errors are in parentheses. Wild bootstrap s.e. in squared parentheses, in order to take into account for the small number of clusters.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

countries whose Herfindahl Index, calculated using 1999 data, is above and below the 1999 average. This is to see whether the treatment effect is different in countries whose TV sector was already highly competitive before the digitization and countries where competition was lower.

The results, reported in table 2.3, show an interesting pattern: on the one hand, the effect of digitization on media freedom is not statistically different from zero in countries with a low level of initial competition. On the other hand, digitization has a strong and negative effect on media freedom where there was already a high level of competition.[26] In other words, this indicates the possibility that too much competition makes it easier for politicians to influence media outlets.

The model reconciles those results with the theoretical literature on media capture. This explanation holds as long as digitization is associated with higher competition, and there is evidence that this is indeed the case. In particular, Figure 2.3 shows the correlation between digital terrestrial penetration and concentration, as measured by the audience market share Herfindahl index described above. The correlation is negative and statistically significant, showing that digitization is associated with higher competition.[27]

[26]Even though the difference between the two is not significant in a regression with the full sample interacted with the high/low competition dummy.

[27]The coefficient remains negative and significant when the Herfindahl index is regressed on penetration, the controls, country and year fixed effect. A one standard deviation increase in DTT penetration

Table 2.3: Sample split by competition in 1999

Dep. variable: media freedom	Competitive 1999	
	Low comp	High comp
DTT penetration	−0.009	−0.035
	(0.012)	(0.016)**
	[0.013]	[0.018]*
Controls	*Y*	*Y*
Time FE	*Y*	*Y*
Country FE	*Y*	*Y*
N	160	213
R^2	0.525	0.476

Dependent variable: media freedom score. DTT penetration uses the most recent observation. Column 1 uses countries with a 1999 Herfindahl index above the average, Column 1 uses countries with a 1999 Herfindahl index level below the average. Controls are corruption, log of GDP per capita, election dummy. Country level clustered standard errors are in parentheses. Wild bootstrap s.e. in squared parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.3.5 Parallel trends and robustness

One concern about the validity of the identification strategy is the endogeneity (at country level) of the decision to begin the switch to digital. I address this concern in three ways. First, even if the timing of the initiation of the digitization process is at the discretion of each country, they all had to make the switch before the European Commission deadline, hence the margins for a strategic use of digitization were relatively small. Second, this endogeneity seems not to affect the parallel pre trends. To test for this, I perform an event-study analysis using logic similar to Dobkin et al. (2014). To do so, I run the following regression:

$$mf_{i,t} = \lambda_i + \mu_t + \sum_{\tau=-9}^{+9} \gamma_\tau * d_{i,\tau} + \delta' \mathbf{x}_{i,t} + \epsilon_{i,t} \tag{2.2}$$

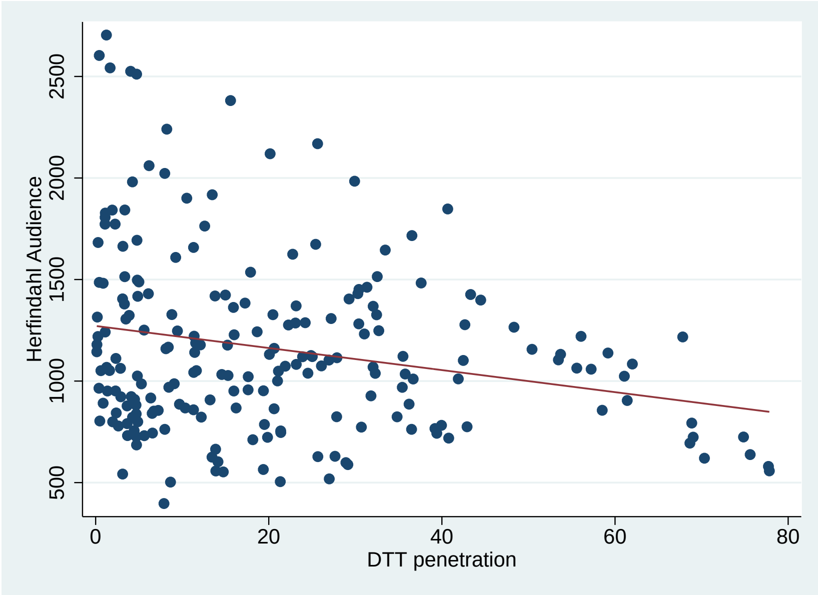
where d_τ is a dummy that takes value 1 for country i two years before the beginning of the adoption process and so on, up to nine years after the starting point of the adoption process. Since every country in the sample gets the treatment at some point in time, I normalize with respect to the coefficient of the year before the DTT launch.[28]

Figure 2.4 shows the coefficient plot of (2.2). Coefficients are not significant before the introduction of DTT and becomes significant and negative afterwards. However, the point estimation of pre-treatment coefficients is not extremely close to zero, and

is associated with a 0.15 standard deviations decrease in the Herfindahl index.

[28]In other words, $d_{i,\tau} = 0$ when $\tau = -1$.

Figure 2.3: Correlation between competition and digitization



Correlation between the Herfindahl Index (top 8 channels) using daily audience market share and DTT penetration. Correlation is -0.312, p-value < 0.01.

as a consequence I do not want to stress too much the causal interpretation of those results. They seem to point at a very robust negative correlation between digitization of terrestrial television and media freedom from political influence, with some evidences suggesting that the former may have a causal effect on the latter.

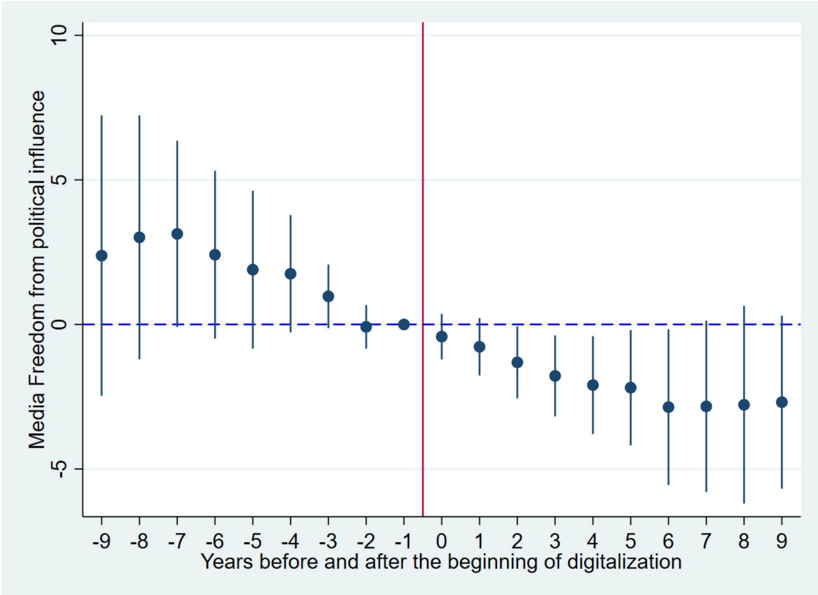
Finally, one last way to address the concern related with the endogenity of the digitization decision is to split the sample between early and late adopters. As table 2.4 shows, the two samples are quite different in terms of observables, hence it is worth enquiring whether the effect of DTT penetration acts differently.

Table 2.4: Summary statistics, early vs late adopters

	Post 2006		Pre 2006	
	mean	sd	mean	sd
Freedom from pol. influence	31.525	3.407	33.469	3.541
Households	3712.820	3857.044	10460.957	11828.639
penhh	8.161	15.310	15.846	18.759
DTT households	173.377	348.315	2080.598	3922.211
GDP p.c.	22305.891	9434.140	32392.013	9880.754
Corruption	5.125	1.246	7.001	1.805
Election year	0.270	0.446	0.254	0.436
Herf. index	1402.272	587.799	1292.241	510.356
TV adv. expenditures	411.220	413.961	1521.048	1831.167

Summary statistics for the two subsamples of countries starting the digitization process before or in 2006 and post 2006.

Figure 2.4: Pre-trends analysis



Start year effect on media freedom score, where 0 is the DTT launch year. Media freedom from political influence score is regressed on country fixed effects, year fixed effects, time varying controls and indicator variables corresponding to the number of years before and after the launch of digital terrestrial television in country i. Coefficients plot of the indicator variables with 95 per cent confidence intervals using the year before the introduction of DTT as reference point.

Table 2.5 looks at the treatment effect splitting the sample between late and early adopters (i.e. countries that started the digitization before or during 2006 vs after 2006). Obviously this implies a lower statistical power, but note that the treatment effect is negative, similar to that in the whole sample, and only marginally non significant (both p-values are below 0.15). Hence, even if early and late adopters are different in terms of observables, it seems that the treatment works on both subsamples in a very similar way. On top of this, table A2.2 in Appendix A2 show that the results are robust to the inclusion of both linear and flexible early-late adopter time trend.

Finally, note that the tests in this section are necessary only in order to give causal interpretation to the negative relationship between digitization and media freedom from political influence. Even if taken as correlations only, the results of section 2.3.4 go in the opposite direction with respect to the current view of the literature on media capture, thus providing a motivation for extending the existing theoretical analysis.

Table 2.5: Regression results, early vs late adopters

Dep. variable: media freedom	Early-late adopters	
	Pre 2006	Post 2006
DTT penetration	−0.024	−0.038
	(0.015)	(0.022)
	[0.018]	[0.029]
Controls	Y	Y
Time FE	Y	Y
Country FE	Y	Y
N	242	131
R^2	0.484	0.601

Dependent variable: media freedom score. DTT penetration uses the most recent observation. Column 1 uses countries that started the digitization before or in 2006. Column 2 uses countries that started after 2006. Controls are corruption, log of GDP per capita, election dummy. Country level clustered standard errors are in parentheses. Wild bootstrap s.e. in squared parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.4 The Model

The model builds on Besley and Prat (2006), keeping essentially the same structure (principal-agent-supervisor, as in Tirole 1986, pure adverse selection) and the same timing. There are two important modifications. Firstly, voters have heterogeneous interests in political news, hence not all of them become informed when one media outlet publishes news about the politician.[29] Secondly, the relationship between readership and outlets' profits is not necessarily linear, but it is assumed to be described by a generic iso-elastic function. Section 2.6 shows that it can be derived as the equilibrium profit of a two sided media market where outlets are selling content to the readers and advertising space to a monopolist advertiser.

2.4.1 The game

Incumbent's type and voters' payoffs

The players of this game are a politician (he), media outlets (it) and voters (she). In period 1 an incumbent of type $\theta \in \{b, g\}$ is in power, with $Pr(\theta = g) = \gamma$. θ is private information of the incumbent. The "good" incumbent always picks the policy that

[29] Note that this is, broadly speaking, consistent with the result of Durante and Knight (2012), where they point out that the change in viewing habits due to Berlusconi's control of some media outlets is only partial.

maximizes voters' welfare, while the "bad" incumbent is only a rent seeker. While in power, he earns rent equal to R. As in the baseline of Besley and Prat (2006), the problem is a purely adverse selection one, so when a good incumbent is in power voters' payoff is 1, and when a bad incumbent is in power this payoff is 0.

Signal structure

Media outlets (but not the voters) may receive a common, verifiable signal $s = \{\emptyset, b\}$ of the incumbent type. In particular, $Pr(s = b|\theta = b) = q$, $Pr(s = \emptyset|\theta = b) = 1 - q$ and $Pr(s = \emptyset|\theta = g) = 1$. Upon observing s, the incumbent offers a vector of bribes $\{t_i\}_{i=1,...,n}$ in exchange for silence.

Every media outlet then has the option to choose a reporting strategy, defined as $\tilde{s_i} \in \{b, \emptyset\}$. If $s = \emptyset$, then the only option available is $\tilde{s_i} = \emptyset$, since there are no signals and news cannot be fabricated. However, if $s = b$, each outlet can decide between reporting $\tilde{s_i} = b$, enjoying the profits from a higher readership, or accepting t_i while publishing $\tilde{s_i} = \emptyset$. A media outlet that chooses $\tilde{s_i} = b$ is referred to as an outlet that *publishes the news about the incumbent.*

Voters' types

Voters are heterogeneous in their interest in politics, meaning that only some of them will be interested in political news. Formally, there are two types of consumers/voters.[30] A fraction α of them is composed of *interested voters*, who are exactly like every voter in Besley and Prat (2006). In particular, they are interested in political news and willing to consume political news from any outlet willing to publish them. Hence, they observe whether some of the outlets have published the news about the politician and they consume that content. They are not going to consume news from any outlet otherwise. In terms of Bayesian updating, note that it is enough that one outlet publishes the news in order for them all to become informed about the signal.

The other $1 - \alpha$ voters are *rationally ignorant*, meaning that they have no interest in spending money or effort to buy political content. As a consequence, unlike the interested voters, they do not actively seek outlets reporting news about the politician,

[30] In total, consumers/voters are a large number normalized to 1.

hence they do not pay attention to whether any of the outlets has published the news about the incumbent. However, they read the media for other reasons (e.g. they like sport, gossip, gardening etc.), and hence they are equally divided amongst all the outlets, irrespective of the signal that they publish. Hence, a fraction of rationally ignorant voters may become informed about the politician's type by reading, for other reasons, one of the outlets publishing the signal.[31] Voters know their type and $\alpha \in (0, 1)$ is common knowledge. As is quite standard in these models, every reader/voter reads at most one piece of news.

As the only relevant piece of news that outlets can publish is the fact that the incumbent is bad, I define *informed* voters as those who are aware of the fact that $s = b$ has been received by media outlets (and hence $\theta = b$). Note that both interested and rationally ignorant voters can become informed about the (bad) type of the incumbent. This is a consequence of the outlet they read. In particular, every interested voter will become informed if at least one outlet publishes the signal, while the fraction of rationally ignorant voters that becomes informed depends on the fraction of outlets publishing the signal.

Media outlets

The number of media outlets, n, is exogenously given and it will be the main aspect of the comparative statics analysis. In general, media outlets derive revenues from two sources: audience-related revenues and money from the politician. Formally, the total profits are defined by Π_i. The component coming from readership is defined by $\pi(r_i) = (r_i)^\delta$, with $\delta \geq 0$. Note that $\delta = 1$ incorporates the case of Besley and Prat (2006), while $\delta > 1$ can be the equilibrium result of a two sided media market, as explained in section 2.6. The bribe offered to outlet i is defined by t_i. For simplicity, it is assumed that outlets offer their contents for free.[32]

After the vector of bribes is decided, outlets observe it and each of them simultaneously

[31] This precise assumption about the behaviour of rationally ignorant voters is not crucial for our main result. This holds as long as a fraction of rationally ignorant voters becomes informed when some outlets are publishing the signal, and this fraction is increasing in the fraction of outlets publishing the signal.

[32] This assumption makes the model more tractable, but it can be shown that it appears quite easily as an equilibrium result in a model where outlets choose the price of their contents in a competitive way and are sufficiently interested in maximizing the readership.

decides whether to accept the individual offer or not. Defining I the set of media that accepts the offer, the incumbent gets a payoff of $R - \sum_{i \in I} t_i$ if he is elected and of $- \sum_{i \in I} t_i$ if he is voted out of office. R is drawn from a distribution F_R with support $[0, +\infty)$. The distribution is common knowledge, while the realization is private information of the incumbent and the outlets.

At the end of period one, voters decide whether to confirm the incumbent or to choose a challenger that is good with probability γ. It is assumed that every voter votes sincerely.[33] In period 2 there is voters' and politicians payoff payment only.

2.4.2 Summary of the timing

1. θ is realized. If $\theta = g$ then $s = \emptyset$ with probability 1. If $\theta = b$, $s = b$ with probability q and $s = \emptyset$ with probability $(1 - q)$. The incumbent observes the media signal and decides $\{t_i\}_{i=1,\dots,n}$.

2. Each media i observes s and $\{t_i\}_{i=1,\dots,n}$ and decides whether to accept or reject t_i. If she rejects, she reports the true signal (if $s = b$) competing with the other outlets that reported the news, if she accepts she reports $\tilde{s}_i = \emptyset$.

3. Voters make readership decisions. Rationally ignorant voters do not observe the vector of reports $\{\tilde{s}_i\}_{i=1,\dots n}$ and they just split themselves among all the outlets, observing just the report of the outlet they pick, i.e. $\tilde{s}_i$. Interested voters instead observe the vector $\{\tilde{s}_i\}_{i=1,\dots n}$ and either split themselves among the outlets choosing $\tilde{s}_i = b$ or do not consume any content.

4. Consumers/voters use the information they have to update beliefs and vote. If the incumbent is voted out, the new incumbent is randomly drawn with $Pr\,(\theta = g) = \gamma$. Period 1 ends.

5. In period 2, payoffs for both periods are paid and the game ends.

[33] Note that, with two alternatives, this is a weakly undominated voting strategy.

2.5 Equilibrium and welfare

2.5.1 Solving the game

The model is solved by backward induction, focusing on showing the existence of a symmetric[34] sincere pure strategy perfect Bayesian equilibrium. I construct the equilibrium using a series of lemmas: all the proofs are in Appendix B2. Of course, it makes sense to focus only on the case where $\theta = b$ and $s = b$, since nothing interesting happens in the rest of the game.

First of all, let us look at voters' choice. Note that, irrespective of their type, the voters' information set is binary (i.e. they observe $\emptyset$ or that there is at least one b). This is because rationally ignorant voters will just observe the report of the outlet they consume, while informed voters can use the whole vector of reports. Lemma 2.1 describes their equilibrium choices, assuming sincere voting. It is reminded that an informed voter is a voter that knows that observes at least one $\tilde{s}_i = b$, hence she knows that $\theta = b$. A uninformed voter is a voter that observes only $\tilde{s}_i = \emptyset$.

Lemma 2.1 *In a sincere voting equilibrium, all uninformed voters vote for the incumbent and all informed voters vote for the challenger.*

Intuitively, in this type of model no news is good news, and this is true even when voters take into account the possibility of media capture. As a result, the politician only needs to keep half the voters uninformed to win elections and remain in power.

Given the assumption about voters' heterogeneous interest in political news, it is straightforward to derive the readership of each outlet depending on its reporting strategy. In particular, define I^C the set of outlets that rejected the bribe from the incumbent (with cardinality $m \geq 0$) and I the set of outlets that accepted the bribe. It is easy to see that $r_j = \frac{1-\alpha}{n}\ \forall j \in I$ and $r_k = \frac{\alpha}{m} + \frac{1-\alpha}{n}\ \forall k \in I^C$.[35]

In other words, n outlets enjoy some readership coming from the rationally ignorant voters, while only those outlets publishing news about the politician will enjoy the

[34] I.e. every outlet follows the same strategy.

[35] This abstracts from the outlets' pricing decisions. However, section 2.6 shows that an equilibrium with zero price can be easily obtained.

additional readership of interested voters. As a consequence,

$$\pi_k = \left(\frac{\alpha}{m} + \frac{1-\alpha}{n}\right)^{\delta} \quad \forall k \in I^C \quad \pi_j = \left(\frac{1-\alpha}{n}\right)^{\delta} \quad \forall j \in I \tag{2.3}$$

The politician determines the optimal number of outlets to silence knowing that all of the α interested voters will become informed once a single outlet publishes the news, while the fraction of the other $1-\alpha$ rationally ignorant voters that become informed is increasing in the fraction of media outlets reporting the news. Hence, if every outlet publishes the news, then all the voters become aware of the type of the politician, while if $m = 0$ then none of them are. To stay in power, given Lemma 2.1, the bad politician needs at least 50 per cent of the voters to be uninformed.[36] Because getting a higher percentage of uninformed voters is irrelevant in terms of re-election, but costly in terms of bribes, the politician knows that he can have a certain number m of outles "allowed" to publish the news about his type, without affecting his re-election. In particular, m must be such that

$$\alpha + (1-\alpha)\frac{m}{n} \leq \frac{1}{2} \tag{2.4}$$

Note that, if $\alpha \in \left[\frac{1}{2}, 1\right]$, i.e. if there is a majority of interested voters, then the result is the same as in Besley and Prat (2006), where the bad politician has to silence the whole industry. Hence, the interesting case is when $\alpha \in \left(0, \frac{1}{2}\right)$. In this respect, Hamilton (2004) points out that interested readers/voters are usually a minority, compared with the rationally ignorant. Hence, the case considered here is likely to reflect reality.

Equation (2.4) can be rearranged as

$$\frac{m}{n} \leq \frac{1-2\alpha}{2(1-\alpha)}$$

As a consequence, defining λ the fraction of outlets that the politician has to silence in order to be re-elected, we can see that, in case of capture, it is optimal for him to have

$$1 - \lambda = \frac{1-2\alpha}{2(1-\alpha)}$$

[36] As an indifference breaking rule, I give a small "incumbency advantage" to the politician in power assuming that, in case of a 50-50 result, he would be re-elected.

and hence

$$\lambda = \frac{1}{2(1-\alpha)} \tag{2.5}$$

Given that λ is not a function of n, from now on $\lceil \lambda n \rceil$, i.e. the smaller integer greater or equal to λn, defines the number of outlets that the politician silences in equilibrium.

It is now possible to characterize (in Lemma 2.2) the bribe structure and the total equilibrium cost of capture in the lowest-cost equilibrium, from the point of view of the incumbent.

Lemma 2.2 *There exists an equilibrium where, to be re-elected at the minimum cost, the politician offers a positive bribe* $t_i = \left(\frac{\alpha}{\lfloor (1-\lambda)n \rfloor + 1} + \frac{1-\alpha}{n}\right)^{\delta} - \left(\frac{1-\alpha}{n}\right)^{\delta}$ *to exactly* $\lceil \lambda n \rceil$ *outlets, where* λ *is defined in equation (2.5), and* $t_j = 0$ *to the remaining outlets.*

The formal proof of this equilibrium is in Appendix B2. Note, however, that there are no unilateral profitable deviations. Outlets not receiving any bribe cannot do anything other than publish the signal. On the other hand, the politician has to pay to every bribed outlet its outside option, i.e. the profits each of them would make if, when the other $\lceil \lambda n \rceil - 1$ outlets are captured, it would decide to deviate from accepting the offer, publishing the news and hence competing with the other $\lfloor (1-\lambda) n \rfloor$ free outlets, minus the profits the outlet is making by staying silent.

In this case, then, $m = \lfloor (1-\lambda) n \rfloor + 1$ and hence the profits from a deviation, following (2.3), would be $\left(\frac{\alpha}{\lfloor (1-\lambda)n \rfloor + 1} + \frac{1-\alpha}{n}\right)^{\delta}$. The politician has to pay the difference between this and the amount of profit made under capture to all the $\lceil \lambda n \rceil$ outlets he needs to capture. The total amount to be paid is

$$K = \lceil \lambda n \rceil \left[\left(\frac{\alpha}{\lfloor (1-\lambda) n \rfloor + 1} + \frac{1-\alpha}{n} \right)^{\delta} - \left(\frac{1-\alpha}{n} \right)^{\delta} \right] \tag{2.6}$$

Moreover, the politician cannot hope to stay in power by bribing a lower number of outlets (as seen above) or with a lower offer, since it would be rejected. Clearly, it is optimal for the bad politician to bribe the outlets if this amount is lower than the office rent he could realize by staying in power in period 2 (defined as R), and hence capture occurs, in equilibrium, when

$$\lceil \lambda n \rceil \left[\left(\frac{\alpha}{\lfloor (1-\lambda) n \rfloor + 1} + \frac{1-\alpha}{n} \right)^{\delta} - \left(\frac{1-\alpha}{n} \right)^{\delta} \right] \leq R \tag{2.7}$$

It is immediately clear that the effect of n on this condition is non trivial. On one hand, raising n implies, as in Besley and Prat (2006), that the politician has to silence more outlets. On the other hand, raising n also increases competition in the slice of the market that remains free (if this exists, of course). This reduces the outside option for every captured firm and makes capture more attractive (and cheaper, for the politician). Given this:

Proposition 2.1 *In the equilibrium described by Lemma 2.2, if $\delta > 1$ then, as n becomes large, competition drives the total cost of capture to zero.*

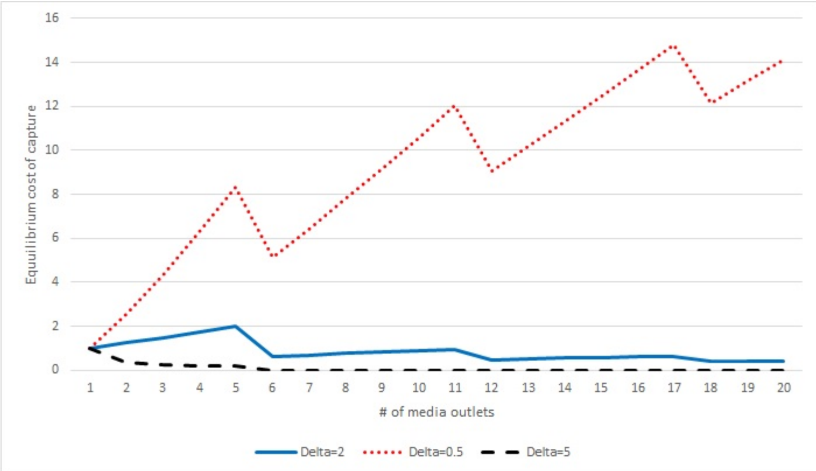
Intuitively, as the extra available readership is reduced because of competition among free media outlets, its marginal effect on profits is also getting increasingly small. As a consequence, the compensation the politician has to pay for capture gets small quite quickly. And this reduces the total cost of capture faster than the increase in the total cost caused by the higher number of firms that need to be silenced. Hence, consistent with the empirical results of section 2.3, excessive competition can actually be bad for media freedom.

Figures 2.5 and 2.6 illustrate of the effect of δ on the total equilibrium cost of capture (normalized so that $K(n = 1) = 1$ irrespective of δ). When it is below 1 (red dotted line), the cost tends to "explode", hence competition makes capture more costly (in the limit) precisely because the "increasing in number" effect dominates the "outside option reduction effect". When δ is big enough (black dashed line), then the total cost of capture is always decreasing in competition, irrespective of the number of free outlets.

The most interesting case occurs for intermediate values of δ (blue solid line), where the effect of competition on capture is non-monotonic. K still goes to 0 in the limit, but this effect kicks in only when competition is sufficiently high, i.e. there is a sufficiently large number of media outlets so that one of them publishing the news is not enough to inform the majority of the voters. Section 2.5.2 discusses this in greater detail.

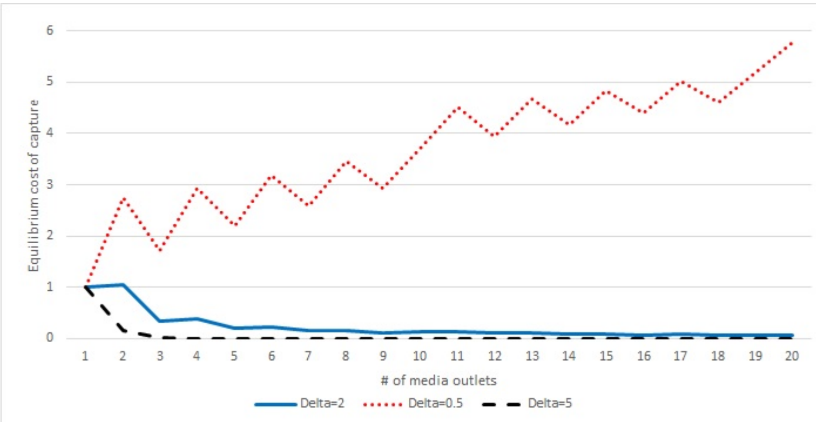
Finally, note that the basic message of Figures 2.5 and 2.6 is not affected by α.

Figure 2.5: Model simulation, high α



Total equilibrium cost of capture for different values of δ when $\alpha = 0.4$. The dotted red line is $K(\delta = 0.75)$, blue solid line is $K(\delta = 2)$, black dashed line is $K(\delta = 5)$. Everything is normalized so that $K(n = 1) = 1$ irrespective of δ.

Figure 2.6: Model simulation, low α



Total equilibrium cost of capture for different values of δ when $\alpha = 0.2$. The dotted red line is $K(\delta = 0.75)$, blue solid line is $K(\delta = 2)$, black dashed line is $K(\delta = 5)$. Everything is normalized so that $K(n = 1) = 1$ irrespective of δ.

2.5.2 Non-monotonic effect of competition

Can this negative result of competition be reconciled with the more positive message of Besley and Prat (2006)? It seems to be the case, at least for specific values of δ. In fact, it can be shown that, when $\delta \in (1, 2]$, the effect of competition on media capture can be non-monotonic.

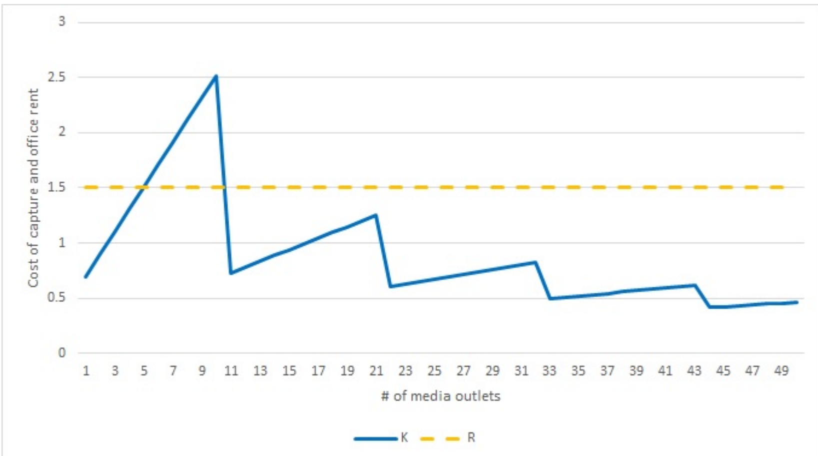
To see this, note that since $\lambda > \frac{1}{2}$, total capture will be necessary for at least $n = 1$

and $n = 2$. The aforementioned restriction on δ guarantees that, before the competition effect kicks in reducing the outside options of captured outlets (i.e. as long as $\lfloor (1-\lambda) n \rfloor = 0$), K is increasing in n. This is formalized in the following corollary:

Corollary 2.1 *In the equilibrium described by Lemma 2.2, if $\delta \in (1, 2]$, there is at least one configuration of the market with free media characterized by a number of outlets $n^* > 1$ and $R \in (K(1); K(n^*))$, then we observe media capture for $n < \underline{n}$ and for $n > \bar{n}$, so the effect of n on the possibility of media capture is non-monotonic.*

This means that media will be captured when their number is either too small or too big, since competition would, at some point, make them so cheap that it will be convenient for the incumbent to silence them. To understand the behaviour of the cost function, Figure 2.7 simulates the model with parameters consistent with Corollary 2.1.

Figure 2.7: Model simulation, non-monotonicity



Model simulation for different values of n when $R = 1.5$, $\alpha = 0.45$, $\delta = 2$.

The number of media outlets is on the x-axis, while the cost of silencing enough of them to have a bad incumbent re-elected when the outlets get the signal is on the y-axis. For the other parameters, $R = 1.5$, $\alpha = 0.45$, $\delta = 2$.

The non-monotonic effect of competition is readily apparent: the incumbent can capture enough outlets and stay in power after a bad signal is received when the blue solid line (the total equilibrium cost of capture) is below the orange dotted line (office rent). That is, when there are at most 4 or at least 11 media outlets. In between, there will be no capture as it would be too costly for the incumbent.

It is clear from the graph that rounding plays a role.[37] In particular, the cost of capture is increasing in n as long as the number of outlets is such that the politician has to silence the whole market. It "jumps down" when n goes from 10 to 11 because of the competition effect given by the possibility of having one free outlet (this lowers the outside option of every captured outlet and hence the total cost). The behaviour is similar for the remaining parts of the graph, with jumps when n goes from 21 to 22, from 32 to 33 and so on. Note, however, that the slope of the cost function is progressively decreasing as n grows, because n enters at the denominator with a power of two.

2.5.3 Discussion

In terms of assumptions, the most important difference between this model and the literature on media capture is the heterogenity in voters' interest in politics. This implies that this model does not assume that having one outlet publish the news is enough to inform the whole electorate.

Gentzkow and Shapiro (2008) use the argument of re-broadcasting as a justification for the assumption in Besley and Prat (2006). However, "silenced" media may not be willing to re-broadcast anything, since they have been compensated in for not doing so. Moreover, in reality there are many ways of re-broadcasting a news item: what matters in this model is the message transmitted to the readers about the type of the incumbent.

Moreover, it must be noted that even in places traditionally associated with media capture (e.g. Italy, Russia), the capture never reached the whole media industry (in Italy it did not even reach the whole television industry when Mr Berlusconi was simultaneously Prime Minister and owner of the three main private television stations). McMillan and Zoido (2004), for example, point out that Montesinos was not bribing the whole media industry, but was choosing the outlets with the highest viewership/readership, and this is consistent with this model. Hence, the possibility that $\lambda < 1$ should at least be considered.

[37] The graph where n is treated as a real, rather than as a natural number, is shown in Appendix C2.

The second important difference is the generic way of describing the relationship between readership and profits. Note, first of all, that the linear case is just a special case of this specification. Secondly, section 2.6 will provide a microfoundation, based on the two sided market literature, of a profit function where $\delta > 1$.

2.5.4 Welfare of the voters

It is now possible to look at the (*ex ante*) welfare of the voters, and how it is affected by competition and the presence of media capture. As already mentioned, office rents R are assumed to be drawn from a generic distribution F_R, with support $\Re^+$. As a consequence, when $\alpha < \frac{1}{2}$, the *ex ante* probability of successful media capture, conditional on having a bad incumbent in power and on $s = b$, is defined as

$$\begin{aligned}\sigma &:= Pr\left(R \geq \lceil \lambda n \rceil \left[\left(\frac{\alpha}{\lfloor (1-\lambda)\, n \rfloor + 1} + \frac{1-\alpha}{n}\right)^{\delta} - \left(\frac{1-\alpha}{n}\right)^{\delta}\right]\right) \\ &= 1 - F_R\left(\lceil \lambda n \rceil \left[\left(\frac{\alpha}{\lfloor (1-\lambda)\, n \rfloor + 1} + \frac{1-\alpha}{n}\right)^{\delta} - \left(\frac{1-\alpha}{n}\right)^{\delta}\right]\right)\end{aligned}$$

With this definition, the probability that a bad incumbent is voted out is $(1-\sigma)q$ and the expected turnover is $(1-\gamma)(1-\sigma)q$. Hence, it is decreasing in σ. In the standard Besley and Prat (2006) model, since n was monotonically decreasing σ, then n was supposed to increase the expected turnover. Interestingly, Drago et al. (2014) find that this is not the case, at least for the entry of local newspapers in Italian municipalities. In this model, when $\delta > 1$, n increases σ in the limit, falling as a consequence the expected turnover. This is consistent with Drago et al. (2014).

Finally, the *ex ante* voters' welfare is defined as

$$W := \gamma(2) + (1-\gamma)\left[q\,(1-\sigma)\,\gamma\right] \tag{2.8}$$

because the good incumbent would be re-elected for sure, leading to a payoff of 1 in each period. On the other hand, the bad incumbent is voted out with probability $q(1-\sigma)$, and in that case the new politician is good with probability γ. If confirmed, the bad incumbent would bring a payoff of 0 in both periods.

As expected, (2.8) shows that voters' welfare is decreasing in the probability of media capture. As a consequence, the effect of media competition on voters' welfare can be described as follows:

Proposition 2.2 *In the equilibrium described by lemma 2.2, if* $\alpha < \frac{1}{2}$ *and* $\delta > 1$ *competition has, in the limit, a decreasing effect on voters' welfare.*

This result suggests that some care is required when thinking about competition as a tool for avoiding media capture. Under some conditions, excessive competition can be counter-productive.

2.6 Microfounding the profit function

As mentioned in sections 2.4 and 2.5, the relationship between outlets' profits and readership pays an important role in my results, and it is important to look beyond the linearity assumption of Besley and Prat (2006). One way of micro-founding a more generic relationship is through a two-sided market approach, where outlets are assumed to be selling contents to the readers and advertising space to a monopolist advertiser, who is interested in placing advertisements where they will reach a large audience.

The timing of the media market part of the game, which determines the shape of the profit function and hence the bribe that the politician has to pay recalls Ellman and Germano (2009),[38] is as follows, starting after the choice of publication strategies.

1. Every media outlet $i \in N$ sets the price of its content, $p_i \geq 0$, achieving as a consequence a readership r_i. Interested voters/consumers are buying the content (one copy each) as long as it publishes the political news and as long as $p_i \leq \bar{p}$, where $\bar{p} \sim U[0,1]$ is a positive individual reservation price. Since these consumers/voters are interested only in the political news, and this news will be the same in every outlet, they treat the m outlets as homogeneous, and hence they will all buy their copy from the outlet setting the lowest price. Rationally ignorant voters/consumers are just equally split between the outlets with the lowest

[38] The idea behind this timing is that readership choices and editorial choices tend to be stable, while you can sign advertising contracts based on them. However, this timing choice is not crucial for the results of this paper.

price.[39]

2. Given the readership, outlets choose the unit price of advertising space, q_i.

3. Finally, the monopolist advertiser, knowing the readership and the price, chooses the quantity of advertising it wants to buy.

In this set up, every media outlet $i \in N$ derives profit from readership and from advertising space. Formally,

$$\pi_i = \pi_{r,i} + \pi_{a,i} \tag{2.9}$$

where $\pi_{r,i}$ and $\pi_{a,i}$ are the profits that outlet i makes from readership and from advertising space respectively. When the decision about the quantity and price of advertising is made, the readership of every outlet has already been determined, hence the (monopolist) advertiser's problem is

$$\max_{y_1,\dots,y_n} \sum_{i=1}^{n} r_i\sqrt{y_i} - \sum_{i=1}^{n} q_i y_i$$

where y_i is the quantity of advertising purchased from outlet i, r_i is the readership of that outlet (determined in stage 1) and q_i is the market price of a unit of advertising on outlet i, already chosen by each outlet i.

The justification for this objective function is found in the literature on advertising and the media market. As pointed out by Hamilton (2004), "once people are watching a program or reading a news entry, advertisers care about the chance to divert their attention to a commercial product". So, following this idea and similarly to the microfoundation of advertiser's demand in Ellman and Germano (2009), I assume that each reader/consumer's demand for the advertised good is increasing and concave in the quantity of advertising in the publication he buys.[40] In other words, the quantity of advertising space boosts the demand for the advertised good at a decreasing rate. As in Ellman and Germano (2009), the advertiser is interested in the total demand for the good, so she multiplies the individual demand determined by the quantity of advertising on outlet i by the readership of that outlet, summing across all outlets pub-

[39] As long as this price is below their reservation price.

[40] The squared root is chosen for convenience. It is easy to see that, defining $z \in (0,1)$ the exponent on y_i, then $\delta = \frac{1}{1-z} > 1$, hence the negative limit effect of competition is always there. If $z \in (0, \frac{1}{2}]$ then the non-monotonicity in the effect of competition on media capture is always there as well.

lishing the signal. This leads directly to the objective function. Note that, as in Dukes (2006), the objective function is additively separable in the media outlets; moreover it is linear in the readership, as in Ellman and Germano (2009) and it exhibits decreasing marginal returns in the quantity of advertising space as in Godes et al. (2009).

The problem is concave and, from first order conditions, the inverse demand function for every outlet is given by

$$y_i = \left(\frac{r_i}{2q_i}\right)^2 \quad \forall i \in N$$

Moving now to the outlets, each one will choose q_i to maximize its profits, knowing how the market would react to its decision. As a consequence, the problem for every outlet i is:

$$\begin{aligned} &\max_{q_i} \, (q_i - c)\, y_i \\ \text{s.t.} \quad & y_i = \left(\frac{r_i}{2q_i}\right)^2 \end{aligned} \tag{2.10}$$

where c is the cost of hosting advertising space (in terms of foregone "useful" space in a newspaper or on a website etc.) that is strictly positive but arbitrarily small. For reasons that will be justified shortly, it is assumed $c \leq \frac{\alpha}{8}$.

Solving (2.10),

$$q_i^* = 2c$$

and

$$y_i^* = \frac{r_i^2}{16c^2}$$

So, the profits from advertising for the media outlet are given by

$$\pi_{a,i}^* = \frac{r_i^2}{16c}$$

that is strictly increasing in r_i. Note that the advertiser is also making positive profits on every outlet, since

$$\pi_{ADV,i}^* = r_i \frac{r_i}{4c} - 2c\frac{r_i^2}{16c^2} = \frac{r_i^2}{8c}$$

where $\pi_{ADV,i}^*$ is the equilibrium profit level for the advertiser on outlet i.

Going backwards to stage 1, note that outlets may, in principle, adopt a different pricing strategy depending on whether they are publishing the signal or not, hence whether they are trying to "attract" also interested readers or not. However, it is easy to note that, as long as $m \geq 2$ and $n - m \geq 2$, the standard features of a Bertrand competition apply in this setting. Every outlet will try to undercut the others in order to reach more readers. Given the zero marginal cost assumption, the only equilibrium is the one where $p_i = p_j = 0$ for $\forall \quad i, j \in N$. The readership decision is the same as above, hence the profits are a convex function of outlets' readership. In particular,

$$\pi_{i,b} = \frac{1}{16c}\left(\frac{\alpha}{m} + \frac{1-\alpha}{n}\right)^2 \quad \forall i \in I^C \quad \pi_{i,\emptyset} = \frac{1}{16c}\left(\frac{1-\alpha}{n}\right)^2 \quad \forall i \in I$$

For cases where Bertrand competition does not apply, the following lemma is sufficient for a zero price equilibrium:

Lemma 2.3 *If $c \leq \frac{\alpha}{8}$ then every outlet, irrespective of the market configuration and on whether they are captured or not, finds optimal to set a price equal to 0 in order to maximize the readership.*

As proved in Lemma 2.3, as long as c is small enough[41] even the monopolist outlet behaves in the same way as in the "Bertrand competition" environment, and as a consequence $\pi^*_{r,i} = 0$ and $\pi^*_i = \pi^*_{a,i} = \frac{r_i^2}{16c}$ and all the rationally ignorant voters are equally divided between all the outlets.

Finally, note that many types of media outlets offer their content for free (websites, most of the TV channels, radio channels, free newspapers), deriving 100 per cent of their profits from advertising and, as pointed out by Ellman and Germano (2009), revenues from advertising account for 50-80 % of the total revenues of "standard" newspapers. Hence, the assumption is not far from reality.

2.7 Conclusion

This paper shows that competition in the mass media market does not have a universally positive role in deterring media capture by bad politicians. On the empirical

[41]The assumption of $c \leq \frac{\alpha}{8}$ is made for simplicity, since even captured outlets have a readership and hence a pricing decision to make. Hence, for example if $n = 2$ and the politician needs to capture every outlet and the optimal outside option choice for an outlet that rejects the bribe involves a positive price, then the other one may find it optimal to impose a positive and slightly lower price. But this would affect the readership of those outlets and hence the number of readers/voters informed and so on. Hence, this assumption plays a role only when Bertrand competition does not apply.

side, it is possible to identify the negative effect on media freedom from political influence of a competition-increasing reform (digitization of terrestrial television in the European Union). Interestingly, this effect seems to be driven by countries with high level of pre-treatment competition, highlighting that the risks comes from excessive competition.

The theoretical model provides a rationale for those results, reconciling the theoretical literature on media capture with the empirical evidence. It relaxes two assumptions from the media capture literature (voters' homogeneity in interest for politics and profits linear in readership) and derives conditions for a negative limit effect and for a non-monotonic effect of competition on the total equilibrium cost of capture. The profit function necessary for those results can arise from a media market modelled as a two sided one.

These findings are relevant in a complex world where the internet seems to allow for proliferation of media outlets and the independence of media from political influence is under threat in many contexts.

This paper is a first step toward a better understanding of the effect of media competition on media freedom from political influence. Two main points are open to further research. First, the role of ideology, and whether ideologically biased media make media capture more or less effective. Second, it would be important to measure media freedom from political influence at a local, sub-national level, in order to use this type of variation for identification purposes.

Appendices

A2 Robustness Checks

I present here two alternative empirical strategies, showing that the main point of the results of Section 2.3 (a negative, significant sign) is robust to those different specifications.

First of all, table A2.1 uses an approach similar to Card (1992), where the treatment is measured by a dummy variable (equal 0 before the beginning of the switch over process, and 1 when it starts) multiplied by the terrestrial television penetration measured in 1997, to capture the difference in affected population discussed in section 2.3. Despite the measurement error implied by this specification (the digitization process takes time, while the treatment variable is 1 immediately after the beginning of the process) the coefficient is negative and significant.

Table A2.1: Robustness: different treatment variable

	Dep. variable: media freedom			
	(1)	(2)	(3)	(4)
Treatment*population	−0.013	−0.014	−0.013	−0.013
	(0.006)**	(0.005)**	(0.005)**	(0.005)**
	[0.007]*	[0.007]**	[0.007]*	[0.006]*
ln(GDP p.c.)		1.898	2.263	2.322
		(2.137)	(2.589)	(2.628)
Corruption			−0.163	−0.177
			(0.322)	(0.327)
Elections				−0.129
				(0.117)
Time FE	Y	Y	Y	Y
Country FE	Y	Y	Y	Y
N	378	377	374	374
R^2	0.453	0.462	0.449	0.450

Dependent variable: media freedom score. Treatment*population is the interaction of a dummy = 1 when the digitization process begins and the penetration of terrestrial television in 1997. Only observations before the beginning of digitization or after the ASO are used. Column 1 is the pure DID set up without controls. Columns 2-4 add one control each. Country level clustered standard errors are in parentheses. Wild bootstrap s.e. in squared parentheses, in order to take into account for the small number of clusters.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Finally, I look at the effect of different specific time trends on the treatment effect. In particular, Table A2.2 repeats the estimation of (2.1) controlling for subgroup specific linear and flexible time trends (where the chosen subgroups are countries with terrestrial television penetration in 1999 above and below 50 per cent and early and late adopters as specified above) and country specific linear time trends. The result is robust to all the specifications but the last one, and the coefficients of columns 1-4 are similar to the specification in the main body of the paper. The result of column 5 is not particularly worrying: Borusyak and Jaravel (2016) show that this type of time trends should not be added to event study analysis, such as the present one.

Table A2.2: Robustness: time trends

	Dep. variable: media freedom				
	(1)	(2)	(3)	(4)	(5)
DTT penetration	-0.024	-0.024	-0.025	-0.025	-0.009
	$(0.013)^{*}$	$(0.014)^{*}$	$(0.012)^{**}$	$(0.013)^{*}$	(0.012)
	$[0.014]^{*}$	[0.015]	$[0.013]^{*}$	$[0.014]^{*}$	[0.012]
Terr. linear trend	Y	N	N	N	N
Terr. flexible trend	N	Y	N	N	N
Early-late linear trend	N	N	Y	N	N
Early-late flexible trend	N	N	N	Y	N
Country specific linear trend	N	N	N	N	Y
Controls	Y	Y	Y	Y	Y
Time FE	Y	Y	Y	Y	Y
Country FE	Y	Y	Y	Y	Y
N	352	352	373	373	373
R^2	0.473	0.500	0.488	0.499	0.712

Dependent variable: media freedom score. DTT penetration uses the most recent observation. Column 1 controls for terrestrial specific linear time trends, Column 2 for terrestrial specific flexible time trends, Column 3 for early-late adopter specific linear time trend, Column 4 for early-late adopter flexible time trend, Column 5 for country specific linear time trends. Controls are corruption, log of GDP per capita, election dummy. Country level clustered standard errors are in parentheses. Wild bootstrap s.e. in squared parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

B2 Proofs

Proof of of Lemma 2.1.

Let us start with interested voters, who behave exactly as the voters in Besley and Prat (2006), i.e. they will all become immediately informed if at least one outlet publishes the signal and they will all observe $\tilde{s_i} = \emptyset$ otherwise. Note that

$$Pr\left(\theta = g | \text{at least one } \tilde{s_i} = b\right) = 0 < \gamma$$

Moreover, suppose the voters believe that the bad politician is able to silence all the media with probability[42] $\sigma^{'} \in [0,1]$. Then, again by Bayes' rule,

$$\begin{aligned}
&Pr\left(\theta = g|\{\tilde{s_i}\}_{i=1,\dots n} = \emptyset\right) \\
&= \frac{Pr\left(\{\tilde{s_i}\}_{i=1,\dots n} = \emptyset|\theta = g\right)Pr\left(\theta = g\right)}{Pr\left(\{\tilde{s_i}\}_{i=1,\dots n} = \emptyset|\theta = g\right)Pr\left(\theta = g\right) + Pr\left(\{\tilde{s_i}\}_{i=1,\dots n} = \emptyset|\theta = b\right)Pr\left(\theta = b\right)} \\
&= \frac{\gamma}{\gamma + \left[(1-q) + q\sigma'\right](1-\gamma)}
\end{aligned}$$

From this, I derive that

$$Pr\left(\theta = g|\{\tilde{s_i}\}_{i=1,\dots n} = \emptyset\right) \geq \gamma$$

is true when

$$\left(1 - \sigma^{'}\right)q \geq 0$$

and hence it is never sequentially rational for an interested voter observing $\{\tilde{s_i}\}_{i=1,\dots n} = \emptyset$ to vote out the incumbent.[43]

Consider now the fraction of uninformed rationally ignorant voters. The single voter, despite knowing his type, does not know if she is observing $\tilde{s_i} = \emptyset$ because she picked up one of the silenced outlets or because there is no signal to be transmitted. Of course, she is still rational and hence she assigns a probability σ to the event of (partial or total) media capture[44] and a probability η to her belonging to the fraction of rationally ignorant voters that would become informed in case of the publication of the signal and capture. In particular, η is equal to the ratio between the number of free outlets in case of capture over the total number of outlets. In equilibrium I will show that $\eta = \frac{m}{n}$, where m is the total number of free outlets and n is the total number of outlets. Note that the rationally ignorant voter knows n and is able to work out how many outlets

[42] Voters do not know the realization of R, but they know that it is a random variable drawn from a generic cumulative distribution function F_R. Note that $\sigma^{'}$ is the probability of total media capture conditional on having a bad incumbent and on the media having received the signal about his type.

[43] As a tie-breaking rule, I am assuming that the indifferent voter will confirm the incumbent.

[44] Note that σ and $\sigma^{'}$ are not the same thing. The latter is the probability that the whole industry is silenced. This is relevant for the interested voters, since they all become informed if a single outlet publishes the news, hence when only partial capture is necessary, then $\sigma^{'} = 0$. On the other hand, σ is the probability that some outlets are captured (both conditional on having a bad incumbent and $s = b$). Note that the voters are able to work out the fraction of media outlets a bad incumbent needs to capture in order to stay in power and the fact that a bad incumbent will either capture just enough outlets or none of them.

can be left free by a bad incumbent in case of successful capture. In other words,

$$Pr\left(\tilde{s}_i = \emptyset | \text{capture}, \theta = b, s = b\right) = 1 - \frac{m}{n}$$

So, every rationally ignorant voter observing $\tilde{s}_i = \emptyset$ would update her belief in the following way:

$$Pr\left(\theta = g | \tilde{s}_i = \emptyset\right) = \frac{\gamma}{\gamma + \left[(1-q) + q\sigma\left(1 - \frac{m}{n}\right)\right](1-\gamma)}$$

This happens because, when the incumbent is a bad type, the voter observes $\tilde{s}_i = \emptyset$ when there is no signal to be transmitted (this happens with probability $(1-q)$) and when there is a signal to be transmitted, but there is capture and she belongs to the fraction of voters that stays uninformed (and this event has probability $q\sigma\left(1 - \frac{m}{n}\right)$). Given this,

$$Pr\left(\theta = g | \tilde{s}_i = \emptyset\right) \geq \gamma$$

is true when

$$1 \geq \sigma\left(1 - \frac{m}{n}\right)$$

So it is always optimal for the rationally ignorant voter observing $\tilde{s}_i = \emptyset$ to confirm the incumbent. Trivially, they will vote out the incumbent after $\tilde{s}_i = b$ since $Pr\left(\theta = g | \tilde{s}_i = b\right) = 0$. ■

Proof of Lemma 2.2.

To characterize formally the equilibrium in the game between the politician and the outlets, I follow closely the proof used in the 2001 working paper version of Besley and Prat (2006).

First of all, it is without loss of generality to order the observed vectors of transfers $\{t_i\}_{i=1,\dots,n}$ such that $t_1 \leq t_2 \leq \dots \leq t_n$. Defining $I_i = 1$ if outlet i accepts the offer and $I_i = 0$ if she rejects it, I characterize the best response recursively.

Given I_1, I_2, ..., I_{i-1}, $I_i = 1$ if and only if

$$t_i \geq \left[\left(\frac{\alpha}{i - \sum_{j<i} I_j} + \frac{1-\alpha}{n}\right)^{\delta} - \left(\frac{1-\alpha}{n}\right)^{\delta}\right] \tag{B2.1}$$

Note that i should accept only if it thinks that $t_i \geq \left[\left(\frac{\alpha}{n-\sum_{j\neq i} I_j} + \frac{1-\alpha}{n}\right)^{\delta} - \left(\frac{1-\alpha}{n}\right)^{\delta}\right]$. Since the LHS is non-decreasing in i and the RHS is non-increasing in i, there must be a $k \in [0, n]$ such that $I_i = 1$ for every $i \geq k+1$ and $I_i = 0$ otherwise.
Note that this implies $t_k < \left[\left(\frac{\alpha}{k-\sum_{j<k} I_j} + \frac{1-\alpha}{n}\right)^{\delta} - \left(\frac{1-\alpha}{n}\right)^{\delta}\right] = \left[\left(\frac{\alpha}{k} + \frac{1-\alpha}{n}\right)^{\delta} - \left(\frac{1-\alpha}{n}\right)^{\delta}\right]$ and $t_{k+1} \geq \left[\left(\frac{\alpha}{k+1} + \frac{1-\alpha}{n}\right)^{\delta} - \left(\frac{1-\alpha}{n}\right)^{\delta}\right]$. As a consequence, an outlet $i \leq k$ does not want to deviate and accept the offer because

$$t_i \leq t_k < \left[\left(\frac{\alpha}{k} + \frac{1-\alpha}{n}\right)^{\delta} - \left(\frac{1-\alpha}{n}\right)^{\delta}\right] \leq \left[\left(\frac{\alpha}{i - \sum_{j<i} I_j} + \frac{1-\alpha}{n}\right)^{\delta} - \left(\frac{1-\alpha}{n}\right)^{\delta}\right]$$

A similar type of reasoning shows that $i \geq k+1$ does not want to deviate rejecting the offer, since

$$t_i \geq t_{k+1} \geq \left[\left(\frac{\alpha}{k+1} + \frac{1-\alpha}{n}\right)^{\delta} - \left(\frac{1-\alpha}{n}\right)^{\delta}\right] = \left[\left(\frac{\alpha}{i - \sum_{j<i} I_j} + \frac{1-\alpha}{n}\right)^{\delta} - \left(\frac{1-\alpha}{n}\right)^{\delta}\right]$$

Hence, equation (B2.1) describes actual best responses. Finally, note that in equilibrium $k = \lfloor (1-\lambda)\, n \rfloor$, because the incumbent does not need to capture $\lfloor (1-\lambda)\, n \rfloor$ outlets, and hence the first $\lfloor (1-\lambda)\, n \rfloor$ offers can be equal to zero. Then, the lowest positive transfer must satisfy

$$t_{k+1} \geq \left[\left(\frac{\alpha}{\lfloor (1-\lambda)\, n \rfloor + 1} + \frac{1-\alpha}{n}\right)^{\delta} - \left(\frac{1-\alpha}{n}\right)^{\delta}\right]$$

Finally, note that a strictly positive bribe to the outlets that it is not necessary to capture is a weakly dominated strategy, hence I focus on the equilibrium where all the first t_k bribes are zero. ■

Proof of Proposition 2.1.

To see this, it is enough to take the limit of (2.6) for n going to infinity. Note, however, that it is possible to rewrite K as follows:

$$
\begin{aligned}
K &= \lceil \lambda n \rceil \left[\left(\frac{\alpha}{\lfloor (1-\lambda)\, n \rfloor + 1} + \frac{1-\alpha}{n} \right)^{\delta} - \left(\frac{1-\alpha}{n} \right)^{\delta} \right] \\
&= \frac{\lceil \lambda n \rceil}{n} \left[n \left(\frac{\alpha}{\lfloor (1-\lambda)\, n \rfloor + 1} + \frac{1-\alpha}{n} \right)^{\delta} - n \left(\frac{1-\alpha}{n} \right)^{\delta} \right] \\
&= \frac{\lceil \lambda n \rceil}{n} \left[n \left(\frac{\alpha}{\frac{\lfloor (1-\lambda) n \rfloor + 1}{n}} + 1 - \alpha \right)^{\delta} \frac{1}{n^{\delta}} - n\,(1-\alpha)^{\delta} \frac{1}{n^{\delta}} \right] \\
&= \frac{\lceil \lambda n \rceil}{n} \frac{1}{n^{\delta-1}} \left[\left(\frac{\alpha}{\frac{\lfloor (1-\lambda) n \rfloor}{n} + \frac{1}{n}} + 1 - \alpha \right)^{\delta} - (1-\alpha)^{\delta} \right]
\end{aligned}
$$

To calculate $\lim_{n\to\infty} K$, I use the fact that $\lim_{n\to\infty} \frac{\lceil \lambda n \rceil}{n} = \lambda$ and $\lim_{n\to\infty} \frac{\lfloor (1-\lambda) n \rfloor}{n} = 1 - \lambda$.

Moreover, note that $\lim_{n\to\infty} \frac{1}{n^{\delta-1}} = 0$ when $\delta > 1$. Hence,

$$
\begin{aligned}
\lim_{n\to\infty} K &= \lambda 0 \left[\left(\frac{\alpha}{1-\lambda+0} + 1 - \alpha \right)^{\delta} - (1-\alpha)^{\delta} \right] \\
&= 0
\end{aligned}
$$

■

Proof of Corollary 2.1.

The fact that, for $\delta > 1$, media capture will be observed for sufficiently high competition follows directly from Proposition 2.1.

The increasing effect of n on K requires the additional constraint on δ. To see this, I treat n as a continuous variable and look at the derivative of K with respect to n, making sure that it is positive at least as long as $\lfloor (1-\lambda)\, n \rfloor = 0$, i.e. as long as the politician has to capture the whole market if he wants to stay in power. Define

$$
K(n) = n \left[\left(\alpha + \frac{1-\alpha}{n} \right)^{\delta} - \left(\frac{1-\alpha}{n} \right)^{\delta} \right]
$$

Taking the derivative,

$$\frac{\partial K}{\partial n} = \left(\alpha + \frac{1-\alpha}{n}\right)^{\delta} - \left(\frac{1-\alpha}{n}\right)^{\delta} + \frac{(1-\alpha)\delta}{n}\left[\left(\alpha + \frac{1-\alpha}{n}\right)^{\delta-1} - \left(\frac{1-\alpha}{n}\right)^{\delta-1}\right] \tag{B2.2}$$

$$= \left(\alpha + \frac{1-\alpha}{n}\right)^{\delta}\left(1 - \frac{(1-\alpha)\delta}{\alpha n + 1 - \alpha}\right) - \left(\frac{1-\alpha}{n}\right)^{\delta}(1-\delta)$$

Note that the second part of (B2.2) is surely positive because $\delta > 1$. A sufficient condition for $\frac{\partial K}{\partial n} > 0$ is $1 - \frac{(1-\alpha)\delta}{\alpha n+1-\alpha} \geq 0$, which simplifies to $\alpha \geq \frac{\delta-1}{n-1+\delta}$.

Suppose now that parameters are such that the condition above does not hold. Then, K is increasing in n when

$$\left(\frac{\frac{1-\alpha}{n}}{\frac{\alpha n+1-\alpha}{n}}\right)^{\delta} > \frac{\frac{(1-\alpha)\delta}{\alpha n+1-\alpha} - 1}{\delta - 1}$$

This simplifies to

$$\left(\frac{1-\alpha}{\alpha n + 1 - \alpha}\right)^{\delta} > \frac{(1-\alpha)\delta - \alpha n - 1 + \alpha}{(\alpha n + 1 - \alpha)(\delta - 1)} \tag{B2.3}$$

As the LHS of (B2.3) is smaller than 1, it tends to 0 as δ increases. Moreover, the RHS is increasing in δ. To see this, note that

$$\frac{\partial RHS}{\partial \delta} = \frac{(1-\alpha)(\alpha n + 1 - \alpha)(\delta - 1) - (\alpha n + 1 - \alpha)\left[(1-\alpha)\delta - (\alpha n + 1 - \alpha)\right]}{((\alpha n + 1 - \alpha)(\delta - 1))^2}$$

$$= \frac{(\alpha n + 1 - \alpha)\alpha n}{((\alpha n + 1 - \alpha)(\delta - 1))^2} > 0$$

As a consequence, it is necessary to find an upper bound of δ such that (B2.3) is always true. As this is analytically very complicated, I just show that, for $\delta = 2$, (B2.3) is satisfied, and as a consequence it is satisfied for every $\delta \in (1, 2]$. Replacing in (B2.3) I obtain:

$$\left(\frac{1-\alpha}{\alpha n+1-\alpha}\right)^2 > \frac{(1-\alpha)-\alpha n}{(\alpha n+1-\alpha)}$$

$$\frac{(1-\alpha)^2}{\alpha n+1-\alpha} > 1-\alpha-\alpha n$$

$$(1-\alpha)^2 > (1-\alpha-\alpha n)(1-\alpha+\alpha n)$$

$$(1-\alpha)^2 > (1-\alpha)^2-(\alpha n)^2$$

that is always true.

As a consequence, for $\delta \in (1,2]$ K is strictly increasing in n as long as the politician needs to capture the whole market, and the total cost goes to zero as n increases. Hence, if R is such that capture is profitable for sufficiently low competition but it becomes too costly as n increases and the politician has to capture the whole market, then we observe a non-monotonic effect of competition on media capture. ■

Proof of Proposition 2.2.

The result follows directly from Proposition 2.1 and equation (2.8). ■

Proof of Lemma 2.3.

If there is one free (say i) and (at least) one captured (say j) outlet, then the captured one has the incentive to undercut the free one, since rationally ignorant voters would just pick the cheapest one. Hence, outlet i can either choose a positive price and sell in equilibrium to interested readers only, or set $p_i = 0$ getting a readership $r_i = \alpha + \frac{1-\alpha}{2}$.

If the optimal price is strictly positive, it must be the $\text{argmax}_{p_i} p_i\left(1-p_i\right)\alpha+\frac{1}{16c}\left[\left(1-p_i\right)\alpha\right]^2$, since she will sell to interested voters only and their demand, as a consequence of the assumption about the individual reservation price, is:

$$D(p_i) = \begin{cases} 0 & \text{if } p_i \geq 1 \\ \alpha & \text{if } p_i = 0 \\ (1 - p_i)\,\alpha & \text{if } p_i \in (0,1) \end{cases}$$

Note that, after few manipulations, the objective function can be written as

$$h(p_i) := \alpha \left[\left(\frac{\alpha}{16c} - 1 \right) p_i^2 - \left(\frac{\alpha}{8c} - 1 \right) p_i + \frac{\alpha}{16c} \right] \tag{B2.4}$$

so the function is quadratic in p_i. Moreover, note that $h(p_i) = 0 \; \forall p_i \geq 1$ and that $\lim_{p_i \to 0} h(p_i) = \frac{\alpha^2}{16c} < \frac{\left(\alpha + \frac{1-\alpha}{2}\right)^2}{16c}$. A necessary condition for an interior solution is the concavity of (B2.4), hence $\frac{\alpha}{16c} - 1 < 0$ or $c > \frac{\alpha}{16}$. However, this is not sufficient. In fact, if $\frac{\alpha}{8c} - 1 > 0$, i.e. if $c < \frac{\alpha}{8}$, then $h(p_i)$ would reach its maximum for a $p_i < 0$, which of course is ruled out by the constraints. Hence, if $\frac{\alpha}{16} < c < \frac{\alpha}{8}$, then the only admissible solution is $p_i = 0$, since the function would be strictly decreasing for $p_i \in [0, 1)$. Moreover, note that if $c < \frac{\alpha}{16}$, then $h(p_i)$ is a convex function in $p_i \geq 0$. There are two possibilities, in this case. Either the minimum of the function is in $p_i \geq 1$ or it belongs to the interval $(0, 1)$. In the first case, $h(p_i)$ would be strictly decreasing in $(0, 1)$, and as a consequence the solution of the maximization problem is $p_i = 0$. The same is true also for the second case, since the function would be strictly convex in $[0, 1]$ and as a consequence the maximum can only be on a corner. Since $h(0) > 0 = h(1)$, then the unique solution is $p_i = 0$.

Finally, consider the case of $n = 1$. In case of capture, her objective function will be

$$h'(p_i) = p_i (1 - p_i)(1 - \alpha) + \frac{1}{16c} \left[(1 - p_i)(1 - \alpha) \right]^2$$

and, for the same argument as above, $p_i^* = 0$ if $c \leq \frac{1-\alpha}{8}$ that is greater than $\frac{\alpha}{8}$ in the interesting case of $\alpha < \frac{1}{2}$. If she publishes the signal, her objective function is

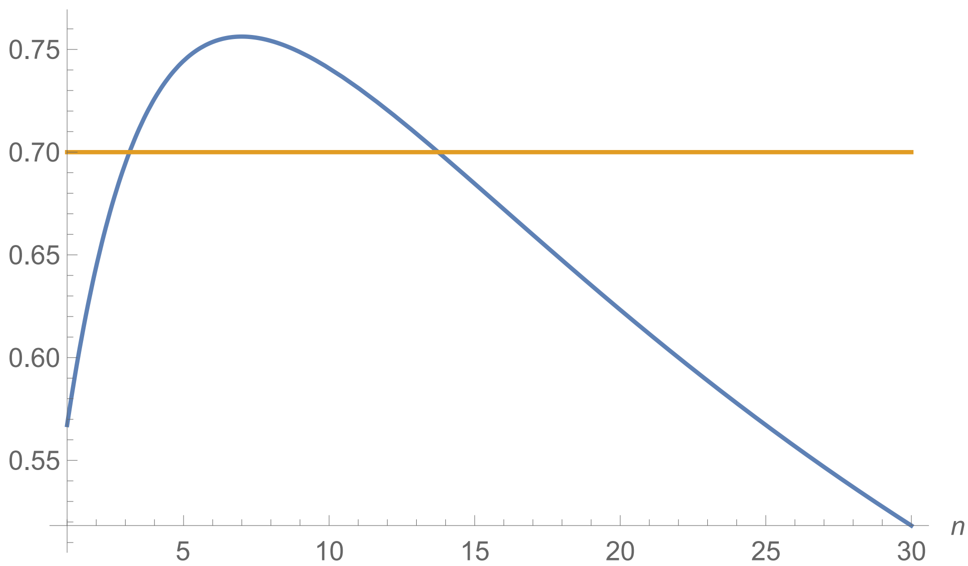
$$h''(p_i) = p_i (1 - p_i) + \frac{1}{16c} \left[(1 - p_i) \right]^2$$

and, for the same argument as above, $p_i^* = 0$ if $c \leq \frac{1}{8}$ that is greater than $\frac{\alpha}{8}$. Hence, the assumption of $c \leq \frac{\alpha}{8}$ ensures that the optimal price is 0 in every possible market configuration and irrespective of the capturing decision. ■

C2 n as a real number

Figure C2.1 shows what happens if n is treated as a real number, rather than as a natural one. This is obviously an approximation, since n is the number of outlets and hence it is discrete by nature. However, it provides a different way of showing the non-monotonicity of K in n. It increases as long as increasing the number of outlets dominates the effect of reducing their outside options, and then it starts decreasing converging to 0 in the limit. As a consequence, media capture occurs when there is a too small or a too large a number of outlets.

Figure C2.1: n as a real number



Plot of K (blue) and R (orange) as a function of n setting $\delta = 2$, $\alpha = 0.45$ and $R = 0.7$.

D2 Data sources

Variables used in the empirical analysis and their sources:

Media freedom score: 41 minus the Environment B score from the Freedom of the Press (Freedom House);

EU: dummy =1 if a country belongs to the European Union on the 25 August 2016;

DTTlaunch: year in which the country started the DTT adoption process. Source: Yearbook 2015;

ASO: year in which the country completed the DTT switch over process. Source: Yearbook 2015;

DTTpen: primary DTT receiver / number of households *100, using country specific tables from 2006-2015 Yearbooks. 0 for years before the DTTlaunch year. Note: no country specific tables for 2014 Yearbook, so I use other data (in practice, DTTraw2013 and HH2013) to work that out. For earlier Yearbooks (2003-2005), I calculated DTTpen using the table "digital TV reception in Europe", vol 2, divided by the number of households, with linear interpolation when necessary and all rounded to the first decimal number;

HH: number of households in thousands, from country specific tables, 2004-2015 Yearbooks. Note: no country specific tables for 2014 Yearbook, so I use the average between YB2013 and YB2015 households to compute HH in 2013 (i.e. linear interpolation);

DTTraw: number of DTT households in thousands from country specific tables and using always the most recent observation available (i.e. taking them from the "market trends" graphs for all the previous years). Note that there are no "market trends" country tables for YB2014, so I use the YB2013 market trend graph for DTTraw 2009. Again, =0 before DTT launch year. For DTTraw 2001-2003 I used the tables T.7.12 (YB2003, vol 2 pag 48) and T.7.6 (YB2003, vol. 2, pag. 42);

Penhh: measure of DTT penetration, equal to dttraw/hh*100. In practice, it calculates the DTT penetration using only the most recent observation available; GDP per capita: GDP per capita in PPP in current international dollars, from UNESCO Institute for Statistics, demographic and socio economic indicators. Used in logs in the regression. Available at: `http://data.uis.unesco.org/Index.aspx?queryid=234`;

Corruption: Corruption Perception Index from Transparency International. Available at `http://www.transparency.org` (recent years) and from Teorell, J. et al

(2015) (older observations);

GDP per capita: GDP per capita in PPP from the UNESCO Institute for Statistics database;

Election: dummy =1 if election year in the lower chamber or in the national assembly. Data collected manually from electionguide.org, provided by the International Foundation for Electoral Systems (IFES).

Herfindahl Audience: Herfindahl index calculated using the annual average of the daily audience market share of the top eight TV channels per country per year. I take the data from the most recent yearbook where available. In particular, Aud1, Aud2 etc denote the yearly average daily audience market share of the highest, second ... eight channel in term of audience in that year (hence, the identity of the channel may change). I use only channels listed in the table. Hence, if less than 8 channels have a listed market share, I put a 0 on the remaining channels (irrespective of whether the row "Others" is different from zero). Most of the observations for Belgium are missed, since data are divided between French and Flemish community. Luxembourg and Malta are also coded differently and hence missed.

TV Advert: total expenditures in television advertising in millions of Euro (when not available directly, I use the expenditures in dollars and convert them to Euro using the year-specific average exchange rate).

TerrPen97: terrestrial television penetration (as a fraction of TV household) in 1997. Coded using table T.2.16 of the 1998 EAO Statistical Yearbook.

References for data sources

Barro, R. J. and Lee, J. W. (2013) A new data set of educational attainment in the world, 1950-2010, *Journal of Development Economics*, 104, 184-198.

European Audiovisual Observatory (1998-2015) *Yearbook. Television, cinema, video and on-demand audiovisual services in Europe*, 1998-2015 Editions, Strasbourg.

Freedom House, Freedom of the Press database (n.d.) https://freedomhouse.org/report-types/freedom-press.

Freedom House, Freedom in the World database (n.d.) https://freedomhouse.org/report-types/freedom-world.

Institute for Democracy and Electoral Assistance (n.d.) database, http://www.idea.int/vt/viewdata.cfm.

International Foundation for Electoral System (n.d.) database, electionguide.org.

Johnson, J. W. and Wallack, J. S. (2010) *ELECTORAL SYSTEMS AND THE PERSONAL VOTE. Update of database from "Particularism Around the World"*, http://dss.ucsd.edu/~jwjohnso/espv.html.

Teorell, J., Dahlberg, S., Holmberg, S., Rothstein, B., Hartmann, F. and Svensson, R. (2015) *The Quality of Government Standard Dataset, version Jan15*, University of Gothenburg: The Quality of Government Institute, http://www.qog.pol.gu.se.

UNESCO Institute for Statistics (n.d.) database, http://data.uis.unesco.org/.

Chapter 3

The Newsroom Dilemma

3.1 Introduction

On April 18, 2013, the *New York Post* plastered its cover page with a picture of two men under the headline "BAG MEN: Feds seek these two pictured at Boston Marathon." The Post was hinting that the duo was responsible for the Boston Marathon bombings and had carried the bombs in their bags. They were innocent, and the Post was wrong. 16-year-old Salaheddin Barhoum and 24-year-old Yassine Zaimi later filed a lawsuit, and the New York Post's credibility was damaged. Similarly, in September 2008, *Bloomberg* incorrectly reported that United Airlines was filing for bankruptcy. Before Bloomberg issuing a correction, United Airlines' stock price nosedived 75 percent.

Media critics often cite such examples to argue that competitive pressures in the modern digital environment have pushed outlets towards early release of less accurate information (Cairncross 2019).[1] Matt Murray, Editor-in-Chief of the *Wall Street Journal*, acknowledged in a recent interview that the Internet had created both time and competitive pressures. However, part of the pressure, he noted, "is to stay true to what has worked and works (really) well, which is reporting verified facts." In a similar vein, some media scholars argue that the fears surrounding the effect of competition may be overblown (Knobel 2018, Carson 2019).

[1]This, of course, is a cause of concern for modern democracies. Media outlets, through fact-checking and investigative journalism, deliver revelations that have a profound impact on the society and its institutions. For instance, *The Hindu*'s Bofors scam exposé in India in 1987 brought the topic of political corruption to center stage and lead to the defeat of the government in power in 1989. More recently, the *New York Times*' exposé on sexual abuse in Hollywood and corporate America has reignited discussions on gender discrimination in the workplace.

In this paper, we discuss why competition among media outlets might not privilege speed over accuracy. We consider the implications of competition on audience welfare and information dissemination. We argue that two opposing forces determine the resolution of the speed-accuracy tradeoff: preemption and reputation. While preemption pushes outlets towards speed, reputation gives media outlets a reason to engage in careful, detailed reporting.

We build a two-period model in which two career-concerned media outlets compete against one another and fear preemption. There is a topic on which the outlets may publish stories. Both outlets receive an initial informative signal about the topic. They may research the topic further at a cost, which depends upon their ability. We model research as generating a perfectly-informative signal about the topic. There is a scoop value associated with being the first outlet to publish a story on the topic. In addition to valuing scoops, outlets also care about their reputations. Reputation depends upon an audience's inference about the outlet's ability to research.

Our model yields three main results. The first two speak to the changes in the media landscape brought about by the Internet. The last result deals with how a source disseminates information to media outlets facing the speed-accuracy tradeoff.

Effect of the Internet. One effect of the Internet has been to increase competitive pressures. The Internet has reduced barriers to entry and contributed to a 24-hour news cycle where reporters are always on deadline. Consequently, pressure on media outlets to be the first to publish have increased.

In our model, while competition can push media outlets to publish more quickly, it can also have the opposite effect – to push outlets to research stories more thoroughly. We find that in more competitive environments, it is easier for outlets to build reputation. This effect increases outlets' willingness to hold back on stories and research them thoroughly. Importantly, our argument relies upon the assumption that the audience does not observe the amount of time outlets spend researching stories but they do observe which outlet publishes first. Knowing the sequence of publication rather than the amount of research, allows for additional observational learning with competition. Consequently, it gives better outlets a reason to differentiate when facing competition.[2]

[2] We discuss in detail the new media studies literature in Section 3.1.1 and show some anecdotal support for our main finding.

We show that when there is a high scoop value, competition drives media outlets to publish more quickly; in contrast, when there is a low scoop value, competition drives media outlets to research stories more. Therefore, the model suggests that breaking news-type stories such as those on terrorist attacks, malfeasance of senior government officials or adverse economic shocks, will suffer particularly from problems of accuracy in the Internet age. In contrast, outlets do better research on non-urgent stories that do not influence immediate decision-making. Examples include: revelations of sexual abuse by Hollywood executives, how terrorist organizations work, and illegal data hacking that is used to influence public opinion.[3]

A second effect of the Internet has been to improve what quickly-released stories look like. Journalists can quickly "contact people, access government records, file Freedom of Information Act requests, and do searches" (Knobel 2018). Similarly, Chan (2014) argues that "digitization brings better access to sources and data." At the same time, however, the cost of doing in-depth research has not changed much. For instance, one would not expect the cost of conducting interviews and building trustworthy sources to have changed significantly. We model such an effect as improving the quality of the initial signal without changing the cost of research.

We find that a *better* initial signal can *reduce* the welfare of the audience. When initial signal becomes better, the audience is less able to attribute correct information by the media outlets to their ability to conduct in-depth research. The audience instead assign it to better initial signal of the outlets that is due to better technology. Thus, reputational concerns get diluted and timing pressures become more salient, making the media outlets move towards speed. Moving towards speed reduces overall welfare only if a significant proportion of audience values better reporting. However, it improves welfare if the audience does value early reporting. It is easy to map the above examples from the previous paragraphs to the relevant situation for audience welfare.

Information dissemination by a source. Our model is also useful for determining

[3]The first story was publishished in both the *New York Times* and the *New Yorker*. `https://www.newyorker.com/news/news-desk/from-aggressive-overtures-to-sexual-assault-harvey-weinsteins-accusers-tell-their-stories`. The second story appeared on the *New York Times* after the reporters researched for more than a year and a half. `https://www.nytimes.com/interactive/2018/04/04/world/middleeast/isis-documents-mosul-iraq.html`. The third story broke out in *The Guardian*. `https://www.theguardian.com/us-news/2015/dec/11/senator-ted-cruz-president-campaign-facebook-user-data`.

how a strategic source shares its information with competing media outlets. Notably, it helps explain why politically-motivated sources may share rumours with multiple outlets to get "unverified facts" out to the audience.

Our model predicts that a source who is merely interested in getting potentially incorrect information out without further research can exploit the time pressures that competing media outlets face. We show that when media outlets are intrinsically driven to explore issues, it is better to share information with all the media outlets to get the information out quickly. More intrinsically motivated media outlets are more likely to do further research independent of the competition. However, by sharing with all the media outlets and creating competitive pressures, additional time pressure can be created. Thus, politicians with propaganda may still hold media outlets hostage even without explicitly capturing or buying them off.

There are, however, situations when such a source shares information only with one media outlet for the quick release of information. The source is likely to do so when media outlets are not intrinsically motivated, and the information is not urgent. When the information is not urgent, there is a general tendency for competing media outlets to investigate further independent of their intrinsic motivation. In this situation, sharing information with just one outlet gets the information out more quickly.

It is worth emphasizing that our model generally covers settings that have elements of preemption and career concerns. For instance, competing researchers working to solve similar problems and hoping to convince a market about their ability face a similar newsroom dilemma. Technology firms face a speed-accuracy tradeoff as they build products and technology to match consumer preferences. Our main results have a natural interpretation in these situations. Notably, better research in competitive environments requires that the initial research idea is not too well-developed.

3.1.1 Stylized facts and new media studies literature

The speed vs. accuracy tradeoff is commonly recognized in the media studies literature. The BBC Academy website observes that "every journalist has to resolve the conflicting demands of speed and accuracy. [...] If you are working on a breaking news story, it is important to remember that first reports may often be confused and misleading. [...]

That is why it is important to weight the facts you have."

The terms of this tradeoff hinge on the surrounding environment. The literature highlights two critical determinants of the rise of "speed-driven journalism" in the modern digital environment. The first one is increasing competitive pressure. Lionel Barber, the Editor of *Financial Times*, points out, "Technology has (also) flattened the digital plain, creating the illusion that all content is equal. It has made it possible for everyone to produce and distribute content that looks equally credible". Thus, outlets cannot only count on their pre-existing reputation to attract readers, and being the first to break the news is increasingly important. Rosenberg and Feldman (2008) note, "Why do experienced journalists telecast unscreened material in volatile situations? Because they can, and because they are driven by powerful, rush-to-report hard instinct, the one commanding them to beat or at least keep astride of the competition and not be left behind".

The second is the 24-hour news cycle (Lee 2014, Starbird et al. 2018), which leads to the possibility of being preempted at any point in time. Newspapers used to have editions making it possible to verify information up until the night before publication, almost without fear of someone else breaking the news. That is no longer the case. As Howard Kurtz from *Washington Post* describes, "In the last year, the pendulum has swung in our newsroom to putting things on the Web almost immediately [...] everybody wants it now-now-now. [...] But the sacrifice (clearly) is in the extra phone calls and the chance to briefly reflect on the story that you are slapping together" (Rosenberg and Feldman 2008).

Importantly, however, reputational concerns remain relevant. *Reuters Handbook of Journalism* states "Reuters aims to report facts, not rumors. Clients rely on us to differentiate between fact and rumor, and our reputation rests partly on that". Note that reputation is based on the ability to check the facts before releasing them, which is also how we model it. Knobel (2018) summarizes her interviews with the editors by saying that they realize that readers can be induced to pay for quality journalism. She quotes Rex Smith, editor of the *Albany Times Union*, "What can separate great journalism from everything else is our commitment to the journalism of verification and watchdog reporting. It will give us credibility that other organizations do not have."

Some new literature from media studies paints a more positive image of the future of watchdog reporting. While not exactly the same as reporting accurate stories, watchdog reporting, which includes investigative journalism and fact-checking, takes time. We show here the data from Knobel's study in support of our theoretical results in Table 3.1. The table shows an increasing share of accountability reporting among a sample of 9 US newspapers for 1991-2011. 2001 in her sample marks the year that the Internet and social media took off in a big way, and became an essential source of news for the audience. We can see how almost all newspapers have increased their accountability reporting since then. The increase is visible for both deep and simple accountability reporting, and across newspaper groups. While the increase may be due to several reasons, her data together with the interviews hint at similar path to that which we outline in this paper.

Table 3.1: Deep (first row) and simple (second row) accountability reporting (as a % of total front-page stories in April) in a sample of 9 newspapers in the US for 1991-2011 in five-year gaps

Newspaper group	Newspaper	1991	1996	2001	2006	2011	Average
Large	*Wall Street Journal*	1.28	2.33	5.88	5.26	4.85	4.03
		30.77	22.09	23.53	22.11	27.18	25.06
	Washington Post	1.51	3.55	4.23	2.72	7.74	3.80
		25.63	27.41	31.92	37.50	36.13	31.43
	New York Times	0.34	0.93	4.35	5.43	3.19	2.46
		10.51	9.29	18.26	19.57	28.72	15.82
Metropolitan dailies	*Albany Times Union* (NY)	6.35	1.22	3.45	4.12	3.61	3.64
		47.62	23.17	28.74	17.53	36.14	26.37
	Denver Post	0.00	4.85	1.80	3.06	5.13	2.92
		23.33	22.33	28.83	29.59	43.59	28.96
	Minneapolis Star Tribune	2.46	1.15	1.83	2.86	5.00	2.68
		31.97	36.78	22.02	34.29	41.00	32.89
	Atlanta Journal-Constitution	1.20	0.00	1.06	1.75	11.84	2.30
		14.97	11.11	13.30	30.70	48.68	20.52
Small	*Bradenton Herald* (FL)	0.93	1.61	1.14	1.27	1.44	1.26
		19.44	33.87	32.95	21.52	19.42	24.16
	Lewiston Tribune (ID)	0.00	0.00	0.00	0.00	1.45	0.32
		22.22	15.25	40.74	28.33	23.19	25.80
Average		1.26	1.81	2.92	3.25	4.46	2.69
		21.52	19.78	24.46	27.26	32.59	24.94

Source: *The Watchdog Still Barks: How Accountability Reporting Evolved for the Digital Age.* Knobel (2018). The author analyzed the content of every front-page story that was published in the month of April (randomly selected) in five-year gaps starting 1991 in a select sample of 9 newspapers. The stories chosen for deep and simple categories involved the following procedure. First, the author eliminated stories that were breaking news. Second, she eliminated stories that had no relation to public policy or politics. In all, she analyzed 1,491 stories in depth using content analysis. Simple accountability reports/stories are those that took a few hours or days to complete, relying on straightforward reporting such as interviews or reviewing published documents. Deep accountability reports/stories are those that took weeks or months to develop and would have remained secret without the journalists' work.

The model we build tries to combine these insights into a unified analysis of the speed-accuracy tradeoff and the competing forces that determine its direction.

3.1.2 Contributions to related economics literature

We primarily contribute to the literature on media competition and quality of news by explicitly modeling the newsroom dilemma. The newsroom dilemma, or the speed-accuracy tradeoff, is surprisingly understudied in the field despite agreement among media scholars on its importance. One exception is Andreottola and De Moragas (2017). They look at the political economy impact of a similar speed-accuracy tradeoff and find that competition leads to a release of less accurate information. Our paper differs because we explicitly model the reputational concerns of media outlets. We identify conditions where the additional information transmitted by the presence of competitors overcomes the preemption concerns.

Some new literature has also started exploring theoretically the effect of the Internet on the media landscape. In Angelucci and Cagé (2019), for instance, the authors show that an Internet-driven drop in the advertising revenues leads to a smaller newsroom, decrease in prices and a move towards "soft" information. Similarly, Armstrong (2005) looks at the relative effect of advertising-only with a subscription-based funding mechanism on journalistic quality. All of these papers and others (Ellman and Germano 2009, Gentzkow 2014) build on two-sided market models (Rochet and Tirole 2003, Rochet and Tirole 2006) and are concerned with pricing decisions. We do not explicitly model advertising and pricing concerns. We instead subsume them under either preemption or reputational concerns.

Some recent papers that do not look at pricing explicitly but explore the political consequences of new media or of media competition are Sobbrio (2014), Allcott and Gentzkow (2017), Barrera et al. (2017), Chen and Suen (2016), Perego and Yuksel (2018) and Vaccari (2018). For instance, Chen and Suen (2016) look at media competition and endogenous attention allocation. They do not have the speed-accuracy tradeoff, but they show that increased competition reduces outlets' investment in reporting quality, increasing the overall influence of the media industry. Perego and Yuksel (2018) and Vaccari (2018) look at the distortive effect of competition on information provision by biased outlets, and as a consequence on the level of information voters can acquire. In both cases, competition can increase distortions. In this paper we abstract from outlets' political motivations to focus on their incentives to provide good quality journalism.

Focusing more on the reputation-building and signaling in media markets, Gentzkow and Shapiro (2006) model media bias and reputation building, showing that competition reduces bias. The model explores an entirely different tradeoff looking at the content of the reporting directly, rather than the timing. Also, the "positive" effect of reputation comes from a different channel – with competition the reader is more and more likely to learn the actual state eventually. Our model does not have this feature as the audience eventually learn the actual story, and competition does not affect the revelation incentives in this way. Gentzkow and Shapiro (2008) later provide an outline of a model that may incorporate reputation-building incentives like ours but they do not consider preemption. Shapiro (2016) shows that reputational concern for unbiasedness may induce journalists to report evidence as ambiguous even when it is not. Preemption concerns and endogenous choice of research are not considered there.

Our modelling strategy shares some features with Hafer et al. (2018) and Hafer et al. (2019). Like us, they have a two period model where competing outlets can acquire information about a politically relevant state of the world and choose when to release it. However, we do not focus on media bias and on the possibility of claiming credit for a story, but rather on the trade off between time pressure and quality of journalism. See Prat and Strömberg (2013) and Strömberg (2015) for recent developments in the political economy of media literature, and other related papers.

We also contribute to the literature on strategic information release. We differentiate from Guttman (2010) and Guttman et al. (2014) by adding reputational concerns and endogenizing the information acquisition choice. Therefore, our results are driven by completely different incentives. Relatedly, Aghamolla (2016) looks at a model of (anti-)herding between financial analysts with endogenous information acquisition. While observational learning is critical in such herding models, reputation building drives such incentives in our model. Observational learning is relevant for the audience in our model because it signals the type of the outlet. Gratton et al. (2017) look at a model in which a sender strategically releases a stream of information to influence perceptions about herself. They show that better sender types release the information earlier and expose themselves to scrutiny. This is in contrast with our model, where better outlets release information later. Preemption concerns drive the incentives in our model, which produces our different result.

Finally, we also contribute to the literature on preemption games and R&D races by adding reputational concerns. Preemption games have long been studied in economics (Fudenberg et al. 1983, Fudenberg and Tirole 1985), but our paper contributes to the more recent literature on preemption games with private information (Hopenhayn and Squintani 2011, Hopenhayn and Squintani 2015, Bobtcheff et al. 2016). It is worth noting that Bobtcheff et al. (2016) have a similar "separating" result for different types of firms, but in a set up without reputation. Here we point out that reputation, combined with actions that partially reveal an opponent's type, can be a different force leading to separating strategies in preemption games.

3.2 A model of the newsroom dilemma

We build a simple two-period model indexed by $t = 1, 2$ featuring three players: two strategic media outlets i, j and a fixed mass of audiences. We also consider a version with just one media outlet.

State of the world. The state of the world ω is binary and unknown to the players. Formally, $\omega \in \Omega := \{a, b\}$ with common prior $\Pr(\omega = a) = \frac{1}{2}$. Ω pertains to the topic on which the media outlets are digging a story, and the relevant information for the audience. This could be, for instance, who is responsible for a terrorist attack, whether a senior government official is involved in corruption or not, who is an appropriate candidate to vote for in the election, etc.

Media outlets. Initially, each outlet privately observes a signal s^i about the state of the world in $t = 1$. We call this the story that the outlets have. We assume that s^i is free and i.i.d. conditional on the state. Its precision is $\Pr(s = \omega|\omega) = \pi \in \left(\frac{1}{2}, 1\right)$. The two outlets decide simultaneously at this stage whether to publish their signals, or conduct further research. The decision d^i for outlet i in $t = 1$ is, therefore, to choose from $\{pub, res\}$ where pub is publish immediately, i.e. in $t = 1$, and res is do more research and then publish in $t = 2$.

Publishing is equivalent to endorsing a particular state of the world (independent of whether published in $t = 1$ or 2). When an outlet publishes its story it sends a message $m \in M = \{\tilde{a}, \tilde{b}\}$ where $\tilde{\omega}$ means endorsing state ω.

Conducting further research (and then publishing in $t = 2$) is costly. In particular, there is a type specific cost of research that perfectly reveals the true state of the world in $t = 2$. Outlets can be of two types, high or low quality, depending on how efficient they are at digging into stories, and this is the private information of each individual outlet. Formally, the type of outlet i is $\theta^i \in \{h, l\}$ with a common prior $\Pr(\theta^i = h) = q = \frac{1}{2}$. The types are independent.

$\theta = l$ faces an infinite cost of conducting research. The low quality outlet never digs stories further and chooses $d = pub$ in $t = 1$. The cost c for the high quality outlet is private information of that outlet, and is story-specific. It comes from a uniform distribution F with support $[-\varepsilon, \bar{c}]$ and is drawn independently for each high quality outlet. ε is greater than zero but small to capture the idea that some high quality outlets may still want to conduct research even in the absence of other rewards.[4] We assume $\bar{c} \geq 2$ so that the support of the distribution F is sufficiently large.

Finally, the assumption on q is just for analytic convenience. A generic $q \in (0, 1)$ would not alter the results, qualitatively. We show this case in Appendix C3.

Audience. The audience enters the game when one or both of the outlets publish their story, and their story is revealed (i.e. m). They only rationally form beliefs about the types of the outlets. They enter with the knowledge of the priors and an understanding of the competition between the outlets. Other than this, the precise information of the audience at the time of belief formation is denoted by the set $\mathcal{I}$.

We assume that the audience observes the sequence of publication but not the actual time of publication, or whether the outlets conducted research. The sequence, as distinguished from the timing, shows whether the outlets moved sequentially or simultaneously. Under this assumption, the audience will be able to determine the actual

[4] Interviews with editors often confirm such motivations; often they feel a sense of responsibility in their positions. For instance, Knobel quotes Marcus Brauchli, *Washington Post*'s former editor, "Doing investigative journalism is in the *Post*'s DNA and has been as long as any of us have been around in journalism." Similarly, Kevin Riley, the Editor of the *Atlanta Journal-Constitution* explains, "People want us to do this. They don't think anyone else will if we don't."

time of publication (i.e. $t = 1, 2$) only if the outlets moved sequentially. It can be summarized by $\tilde{t}^i \in \{\mathrm{I}, \mathrm{II}, \varnothing\}$, which shows whether outlet i was first, second, or it moved simultaneously with j. This assumption is discussed in more details in Section 3.2.2 and its implications are described in the main analysis (section 3.3).

In addition, after both the outlets publish their stories, the state is revealed exogenously. If $m^i = \omega$, then outlet i is said to be right, or R. Otherwise, the outlet is wrong, denoted by W. We call this the outcome O of verification. The audience sees the outcome. Therefore, the information of the audience $\mathcal{I}$ at the end of the game is denoted by a tuple $(O^i_{\tilde{t}}, O^j_{\tilde{t}})$ that consists of four pieces of information, i.e. the position of each outlet in the sequence of publication and their outcomes. Using $\mathcal{I}$, the audience updates its beliefs about each outlet's type. Denote the posterior belief about $\theta = h$ by $\gamma(\mathcal{I})$ when the information held by the audience is $\mathcal{I}$.

Payoffs. Currently, we do not illustrate the payoffs of the audience as they only form beliefs. We will, however, place more structure on its preferences at a later stage and explain the source of outlets' payoffs. For the time being, we only focus on the outlets' payoffs, which are composed of three elements.

1. The first is a scoop value v to the first outlet publishing the story. It captures the preemptive nature of the media market, highlighted for example in Besley and Prat (2006). v can be interpreted as the mass of audience that is drawn to the first media outlet breaking the story.

2. The second is a reputation value of γ^i or the audience's posterior on the quality of outlet i calculated after revelation of the true state. This captures the extent to which the outlets care about their reputation. For instance, future audience of the outlets might depend on their reputations. We assume that reputation enters linearly in the outlets' payoffs. Importantly, the audience cares about whether the outlet is high or low type, not about c. A new c is drawn for every new story and only the high type has the ability to conduct further research.

3. The third is the cost c that the high type outlet chooses to pay if it does research in period 1.

Timing. The timing of the game can now be summarized as follows:

0. Nature draws ω, θ^i and θ^j. θ is privately observed by each outlet. ω is unobserved.

1. At $t = 1$ each outlet privately observes s^i. A cost c of digging into the story is drawn from a uniform distribution $F[-\varepsilon, \bar{c}]$ for the high type.

2. The outlets simultaneously decide $d^i \in \{pub, res\}$ and if $d^i = pub$ then also choose m. As stated before, this is a relevant decision only for the high type. The low type always chooses pub.

3. If both outlets publish, the game ends. Otherwise, the game goes to period 2.

4. At $t = 2$, the state is revealed to every outlet that chose $d^i = res$. Those who did not publish in $t = 1$, publish now by choosing m.

5. Once both the outlets have published, the state ω is revealed to the audience. They observe $\mathcal{I}$ and update beliefs on the type of each outlet. Payoffs are realized.

3.2.1 Solution concept and equilibria selection

The solution concept we use is the Perfect Bayesian Nash Equilibrium in pure strategies. We focus on equilibria where outlets optimally follow the signal they receive, i.e they endorse the state that is more likely to be the true one given their signal. We call such equilibria *signal-based* equilibria.[5] For the rest of the paper, we use "equilibrium" and "signal-based equilibrium" interchangeably.

3.2.2 Discussion of assumptions

Before proceeding to the analysis, it is worth discussing our assumptions in detail.

The first assumption we make is regarding what the audience observes about the timing of the game. The fact that the audience only observes the content of what was published (i.e, m) and the sequence of publication (i.e., $\tilde{t}$ but not the actual t) captures the idea that it is unaware of how much the outlets researched story. We believe this is a realistic assumption in that the amount of research is hardly observable from outside

[5]This means that we ignore equilibria where outlets choose to endorse one particular state to signal their type. Those equilibria may exist, but we argue that they do not make much sense given the environment we are considering. Alternatively, we can assume that signals are hard information, but the reader cannot infer the level of precision: the result would be exactly the same.

the newsroom. Of course, the amount of research conducted maps in a probabilistic way into the accuracy of a story, which the audience can check more easily. We allow for such a possibility by letting the audience observe whether the story is true or false.

The important consequence of this assumption is that player i's decision to publish/not publish can potentially convey information about player j's type. For example, if the two outlets move sequentially and only a high type is expected to conduct research, moving later is a signal of the first outlet being a low type. We show how relaxing this assumption changes our result in Section 3.3.4.

The second assumption we make is about who possesses stories on a topic. In reality, competing media outlets are often unaware of whether their competitors are also exploring the same story. We assume that both of the media outlets are aware that their competitor also possesses the story. Doing so pushes the incentives of the outlets the most towards speed. Still, we show that more research is possible under competition. Including such a possibility further adds to the complications of the model.

The third assumption we make is that outlets build a reputation on their consistent types, and not on the cost of digging into each independent story. Given that different outlets usually have different expertise, it is reasonable to assume that they face different costs when exploring different stories. For instance, *The Wall Street Journal* is a business-centric daily and has invested in building sources and methods for dealing with business stories (such as avoiding lawsuits when potentially sensitive corporate information is published). However, in general, some outlets have a culture of researching while others do not. Their type captures this.

We also make a few assumptions for tractability reasons. First, we do not allow for the outlets to "sit on information" or wait for a period before publishing.[6] Second, we assume that the audience correctly finds out the state at the end of the game. Third, we assume that the media outlet correctly finds out the state upon choosing to research.

3.2.3 Preliminary observations and strategies

We start with a few simplifying observations. All the proofs are in Appendix A3.

[6] We can show that for a sufficiently high v and relevant off-path beliefs, the outlets never choose to wait.

Observation 3.1 *Suppose there are reputational gains in matching the state. If an outlet decides to publish in $t = 1$, it follows its signal s, i.e sends $m = s$. If an outlet decides to do research and then publish in $t = 2$, it follows the outcome of research.*

Observation 3.1 follows from the fact that in $t = 1$ the most informative signal is s. Therefore, the most likely state is the one given by the signal. This is a standard result in this type of environment and follows from the flat priors on the state. Moreover, in $t = 2$ the outlet choosing to research has learned the actual state and therefore, publishes it (independent of what the original signal s stated). Thus, as long as there is a gain in matching the state, each outlet follows its last signal, which is also the most informative one.

There is also a useful result arising from our particular signal structure and flat prior over the state.

Lemma 3.1 *If each outlet follows its last signal when publishing, the following results hold:*

1. *The probability of matching the state after only s is π.*
2. *Regardless of whether i decides to publish or research, from its point of view the expected probability of player j matching the state without research is π.*

Lemma 3.1 will be helpful in writing the incentive compatibility conditions for the players. Doing so will require each outlet to consider whether the other will do research and the subsequent probability of matching the state.

It is useful to define precisely the strategies we will focus our attention on. Note first that the only relevant and meaningful decision that deserves our attention is the one of the high type outlet in period 1. From the the outlet's point of view, this will be a threshold strategy where the threshold is defined on the cost c of research. The high outlet conducts research if the realized cost c is less than some threshold c_D (where subscript D represents the case of a two firm duopoly).[7] From the other outlet's (and the audience's) point of view, define σ^i, the conjectured probability that outlet i chooses

[7]Similarly, the case of single firm monopoly is denoted by a threshold c_M and in general, by a subscript M.

to research further in $t = 1$, conditional on outlet i being a high type. Therefore,

$$\sigma^i = \Pr(c \leq c_D) = F(c_D) = \begin{cases} 0 & c_D < -\varepsilon \\ \frac{c_D+\varepsilon}{\bar{c}+\varepsilon} & -\varepsilon \leq c_D \leq \bar{c} \\ 1 & c_D > \bar{c} \end{cases}$$

We are now ready to move to the equilibrium analysis arising in different market configurations.

3.3 Competition leads to better reporting

3.3.1 Newsroom dilemma with a single firm: Monopoly

Let us start with the simplest case: there is a single media outlet and its type is known to the audience.

Proposition 3.1 *If there is one media outlet and θ is known to the audience, then the high quality outlet conducts research with probability $F(0) = \frac{\varepsilon}{\bar{c}+\varepsilon}$.*

In this case, none of the aforementioned incentives are at play. There is obviously no preemption risk and there is nothing to do in terms of reputation. Every type of outlet gets $v + \mathbb{1}\{\theta = h\}$ so it is pointless to pay any cost for researching. The outlet is driven to research only because of its intrinsic motivation.

The case of monopoly with unknown type is more interesting. Proposition 3.2 summarizes the main result.

Proposition 3.2 *If there is one media outlet and θ is not known to the audience, there exists a unique equilibrium in which the high quality outlet conducts research in $t = 1$ if*

$$c \leq (1 - \pi)(\gamma(R) - \gamma(W)) := c_M$$

where $\gamma(R)$ and $\gamma(W)$ are the audience beliefs about the outlet's quality after it gets the state right and wrong respectively. As a consequence, $\sigma^* = F(c_M) = \frac{c_M(\sigma^*)+\varepsilon}{\bar{c}+\varepsilon}$.

Note that the preemption risk is absent in this case as there is only one outlet. v does not play any role in the threshold above. But a high outlet is incentivized to do research to build a reputation for being a high quality. However, this reputation cannot be based on the observation of sequence or timing. The only relevant thing that the audience observes is whether the outlet is right or wrong, i.e. whether $m = \omega$ or not after the state is verified. Therefore, if the high outlet is expected to choose to research with probability σ, the two relevant belief updates are

$$\gamma(R) = \frac{\sigma + (1-\sigma)\pi}{\sigma + (1-\sigma)\pi + \pi} \text{ and } \gamma(W) = \frac{(1-\sigma)(1-\pi)}{(1-\sigma)(1-\pi) + (1-\pi)} = \frac{1-\sigma}{2-\sigma}$$

from Bayes' rule. Now, it must be that the additional cost c of choosing $d = res$ is more than compensated by the expected reputational gains from endorsing the correct state. This is captured in c_M of Proposition 3.2.

3.3.2 Newsroom dilemma with two firms: Duopoly

The main effect of competition is the introduction of preemption risk. When preemption is relevant and reputation building is not, then the equilibrium where the high quality outlet conducts research becomes even rarer than in Proposition 3.1. Proposition 3.3 below highlights this.

Proposition 3.3 *If there are two media outlets and θ is known to the audience, there exists a unique symmetric equilibrium in which the high quality outlets conduct research with probability $\sigma_D^* = F\left(-\frac{v}{2}\right)$.*

Intuitively, there is nothing to gain from conducting research in terms of reputation as θ is known. The only reason to investigate further is if there is an intrinsic motivation to do so. But now there is a preemption risk that reduces the incentives to investigate. However, if v is sufficiently small relative to the intrinsic motivation (i.e. if $v < 2\varepsilon$), there will still be some high outlets willing to investigate.

The case of competition plus hidden types is the most interesting one. In this case, both the preemption and reputation building concerns are simultaneously relevant and interact with each other. Before we present the key proposition, we discuss how the audience updates beliefs in this environment. Recall that the audience observes both the outcome of verification $\mathcal{O} \in \{R, W\}$ and the sequence of publication $\tilde{t} \in \{\text{I}, \text{II}, \varnothing\}$ for both i and j. Suppose now that a high quality outlet chooses to research with probability σ^i. Then, for a given conjectured level of σ^i and σ^j, the relevant audience's on-path beliefs need to be defined for the following events:

$$\{(R_\varnothing, R_\varnothing), (R_\varnothing, W_\varnothing), (W_\varnothing, W_\varnothing), (W_\varnothing, R_\varnothing), (R_\text{I}, R_\text{II}), (W_\text{I}, R_\text{II}), (R_\text{II}, R_\text{I}), (R_\text{II}, W_\text{I})\}$$

where the first outcome-sequence element in each information set is outlet i's and the second is outlet j's.[8]

It can be shown that there are three relevant set of events for belief updating. The first is when both the outlets get the state correct and they publish simultaneously.

$$\gamma^i(R_\varnothing, R_\varnothing) = \frac{\sigma^i\sigma^j + (1-\sigma^i)(2-\sigma^j)\pi^2}{\sigma^i\sigma^j + (2-\sigma^i)(2-\sigma^j)\pi^2} := \gamma^i(\varnothing)$$

Here the audience is unable to determine the actual timing of publication. It cannot distinguish as to whether both conducted research (which happens only if both are high types) or both published immediately (either because they are both low types, or because there is only one high type and it faced a high cost, or because both are high types but they faced high costs). With some abuse of notation, we denote the updated belief under "no information about timing" event by $\gamma(\varnothing)$.

The second is when the audience is able to determine that outlet i moved in $t = 1$.

$$\gamma^i(R_\varnothing, W_\varnothing) = \gamma^i(W_\varnothing, W_\varnothing) = \gamma^i(W_\varnothing, R_\varnothing) = \gamma^i(R_\text{I}, R_\text{II}) = \gamma^i(W_\text{I}, R_\text{II}) = \frac{1-\sigma^i}{2-\sigma^i} := \gamma^i(1)$$

This, of course happens when i moves first and j moves second (independent of whether i gets the state correct or not). But the audience is also able to understand it when the outlets move simultaneously and at least one of them gets the state incorrect (since researching further perfectly reveals the state). Here the only uncertainty for the audience is whether the outlet is a high quality one that faced a high cost or a low quality one. We denote the updated belief under the "published in period 1" event by

[8] Note that it never happens that an outlet moves second in the sequence and gets the state incorrect. Any outlet that moves second has conducted research and matches the state perfectly. Therefore, any event with W_II does not occur on-path.

$\gamma(1)$. Observe how in these events the presence of a competitor conveys to the reader some additional information about the type of each outlet.

Finally, the third is when the audience is able to determine that outlet i moved in $t=2$.

$$\gamma^i(R_{\mathrm{II}}, R_{\mathrm{I}}) = \gamma^i(R_{\mathrm{II}}, W_{\mathrm{I}}) = 1 := \gamma^i(2)$$

This only happens when outlet i moves second and gets the state correct, which in turn is only possible if it is a high quality outlet. Therefore, the updated belief under "published in $t=2$" event is $\gamma(2) = 1$.[9]

Using these updated beliefs, a high quality outlet's incentive compatibility can be written as follows. For any given conjectured σ^j and audience's beliefs, a high quality outlet i with cost c^i chooses to research further if

$$\overbrace{\frac{1}{2}\left[\sigma^j\left(\frac{v}{2}+\gamma^i(\varnothing)\right)+(1-\sigma^j)\gamma^i(2)\right]+\frac{1}{2}\gamma^i(2)-c^i}^{\text{expected payoff from research}} \geq$$
$$\underbrace{\frac{1}{2}\left[\sigma^j(v+\gamma^i(1))+(1-\sigma^j)\left(\frac{v}{2}+\pi^2\gamma^i(\varnothing)+(1-\pi^2)\gamma^i(1)\right)\right]+\frac{1}{2}\left(\frac{v}{2}+\pi^2\gamma^i(\varnothing)+(1-\pi^2)\gamma^i(1)\right)}_{\text{expected payoff from publication}}$$

which further simplifies to

$$c^i \leq \frac{1}{2}\left[\left(\gamma^i(\varnothing)-\gamma^i(1)\right)\left(\sigma^j-(2-\sigma^j)\pi^2\right)+(2-\sigma^j)\left(1-\gamma^i(1)\right)\right]-\frac{1}{2}v := c_D^i \quad (3.1)$$

Proposition 3.4 then follows:

Proposition 3.4 *If there are two media outlets and θ is not known to the audience, there exists a unique and symmetric equilibrium where $\sigma^{i*} = \sigma^{j*} := \sigma^* = F(c_D)$ such that*

$$c_D = \frac{1}{2}\left[\left(\gamma(\varnothing)-\gamma(1)\right)\left(\sigma^*-(2-\sigma^*)\pi^2\right)+1\right]-\frac{1}{2}v$$

where $\gamma(\varnothing) = \frac{(\sigma^*)^2+(1-\sigma^*)(2-\sigma^*)\pi^2}{(\sigma^*)^2+(2-\sigma^*)^2\pi^2}$ and $\gamma(1) = \frac{1-\sigma^*}{2-\sigma^*}$.

Looking now at the cost threshold c_D of Proposition 3.4, we can see the negative effect of v. If preemption concerns are very salient (i.e. v is high), then separation happens for a smaller range of c making research less likely. On the other hand, the positive

[9] This is off path if, in equilibrium, every high quality outlet chooses not to research. However, we assume that off path beliefs are $\gamma(2) = 1$ in this case as well.

side of the condition is given by the expected reputational gains of matching the state (and publishing second).

3.3.3 Competition may lead to better reporting

The comparison between monopoly and duopoly when reputation building is relevant (Propositions 3.2 and 3.4) provides interesting insights.

Lemma 3.2 *The reputational gains are always higher in duopoly than in monopoly.*

The reason lies in the availability of additional information in the case of duopoly. First, the presence of two outlets allows the audience to compare their contents, i.e. which states outlets i and j endorsed. Second, it allows the audience to observe the sequence of publication of the two outlets. Together, these two factors allow outlet i to publish after outlet j, match the state correctly, and signal its type more easily. In turn, this makes outlet i more willing to pay the cost of research. However, the additional preemption concerns in duopoly counterbalance this positive information effect, and makes c_D decreasing in v. The two effects combined yield our first main result pertaining to the effect of Internet-driven competition on reporting.

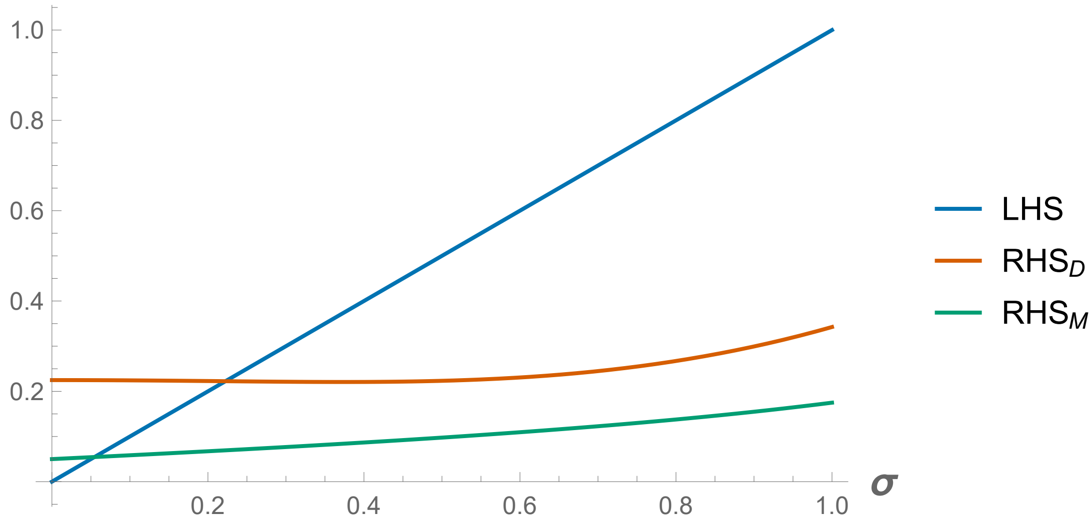
Proposition 3.5 *There exists a nonempty interval of v values where $\sigma_D^* > \sigma_M^*$.*

Basically, what Proposition 3.5 says is that there is a nonempty set of parameters where research is more likely in duopoly than in monopoly. Therefore, competition may lead to better reporting.

A good way to illustrate Proposition 3.5 is Figure 3.1. The orange line is $F(c_D)$, the green line is $F(c_M)$ and the blue one is the 45 degree line. The equilibrium probability of research is given by the point of intersection of $F_c(c_D)$ and $F_c(c_M)$ with the 45 degree line. It is clear that $\sigma_D^* > \sigma_M^*$ for sufficiently small v.

Intuitively, reputational gains in monopoly are given by the increased probability of getting the state right. In duopoly, the audience can use one extra piece of information

Figure 3.1: Equilibrium σ_D^* and σ_M^* for when $\pi = .6$, $v = .3$, $\bar{c} = 2$ and $\varepsilon = .1$.



– the action of the other outlet, which includes the outcome of verification and the sequence of publication. Hence, competition induces a trade off between those two forces pushing in opposite directions. Importantly, this trade off is not obvious. The main point of Proposition 3.5 is precisely to point out that, contrary to the wisdom of the crowd in media studies literature, competition does not necessarily lead to a faster release of less accurate information.

3.3.4 The role of audience's information

The previous results relied critically on what the audience observes from the competition, or simply the "transparency". To build further intuition, here we analyze how changing the transparency affects our result. In general, the effect of transparency on the possibility that competition induces better reporting is non-monotonic. To see why, consider the two other possibilities – nothing about the timing is observable and the timing of research is fully observable. Our original assumption lies in the middle of this increasing transparency spectrum. Of course, the content of publication is always visible to the audience, i.e. the audience observes m.

Unobservable timing or zero transparency. Suppose the audience observes neither the timing of publication nor the sequence of publication. It simply consumes the content of the outlet publishing the story. In this case, the behavior of the monopolist is exactly as before. Hence, $c_M = (1-\pi)(\gamma(R) - \gamma(W))$ does not change. In the case of duopoly, however, the endorsement of the other outlet does not matter anymore in the updating. The audience considers each outlet separately because nothing about the timing is observed. Therefore, $\gamma(R, .) = \gamma(R)$ and $\gamma(W, .) = \gamma(W)$. The consequence is summarized in the following corollary.

Corollary 3.1 *If neither time nor the sequence of publication are observable,*

$$c'_D = c_M - \frac{1}{2}v$$

and therefore, $c'_D < c_M$ for every strictly positive v.

Intuitively, there are no additional reputational gains because it is not easier to "look good" in the presence of a competitor. In fact, the reputational part of the cost threshold is exactly the same. But the additional risk of preemption pushes c_D down.

Observable timing or full transparency. If the timing of publication is observable, the monopolist can fully differentiate itself by publishing in period 2. This is possible because the audience can now perfectly distinguish between period 1 and 2, and therefore, is fully aware of whether research was conducted or not. Moreover, this is true in duopoly as well. In fact, the actual content of the publication does not matter for the reputation-building, and differentiation is driven entirely by the timing. As a consequence, the logic applies as before. The reputational part of the threshold is the same, but preemption concerns reduce the incentives to investigate and conduct research.

Corollary 3.2 *If the timing of publication is observable,*

$$c''_M = 1 - \gamma(1) \ \text{ and } \ c''_D = 1 - \gamma(1) - \frac{1}{2}v$$

where $\gamma(1) = \frac{1-\sigma}{2-\sigma}$. Therefore, $c''_D < c''_M$ for every strictly positive v.

Note that now the cost thresholds are bigger than in the previous information environments. This is so because now maximum distinction is possible between the two outlets. Therefore, the actual levels of reputational benefits are also higher. This is captured in the belief updating,

$$\gamma(1) = \frac{1-\sigma}{2-\sigma} \text{ but now } \gamma(2) = 1.$$

It is worth emphasizing that both of these extreme transparency assumptions are somehow problematic. Completely unobservable timing clashes with the idea of a scoop value, or more generally with the preemptive nature of the media market. If the audience has no understanding of when the publication happened, there is nothing to gain from being first. There are only gains from ultimate publication. This is obviously not true in reality. Completely observable timing, on the other hand, implies that the reader perfectly understands exactly how much research went into an article. Therefore, the whole differentiation happens on the time dimension, rather than on the truthfulness of the story. Again, this hardly seems true in reality.

3.4 Stories and the effect of better initial information

We are now in a position to discuss what kinds of stories are susceptible to more speed-driven journalism and what aren't. To do so, we place more restrictions on audience preferences.

Let there be a unit mass of audience. The audience decides on whether to take an action or not. Let this action be denoted by $\alpha \in \{a, b\}$ and interpreted as "matching the state". The audience seeks out the information published by the outlets and consumes their content to the extent it wants to match its action to the story. Examples include decisions on who to vote or to form opinions. Note that the audience also has an option not take the action at all and therefore, not opt for any outlet.

For any given story, a fraction u of this audience requires the information urgently. The preferences of members of the urgent audience is given by

$$V_u = \begin{cases} 1 \text{ if deciding today and } \alpha = \omega, \\ 0 \text{ if deciding today and } \alpha \neq \omega, \\ -k \text{ if remaining undecided or deciding tomorrow} \end{cases}$$

where ‘today” happens for the audience when the first outlet publishes its content. So, when the outlets publish sequentially, there is a clear notion of today and tomorrow. However, when the outlets publish simultaneously, today happens at that time period.[10]

The remaining audience, a fraction $1 - u$, is patient. Its preference, on the other hand is given by

$$V_{1-u} = \begin{cases} 1 \text{ if } \alpha = \omega, \\ -k \text{ if } \alpha \neq \omega, \\ 0 \text{ if remaining undecided.} \end{cases}$$

Observe that the patient audience does not care about when it makes the decision. Taking an accurate decision matters more. Notably, the difference between the urgent and the patient audience lies in how they value making a wrong decision. Assuming k to be sufficiently large, it matters more to the urgent audience to make a decision as soon as the first outlet publishes. For the patient audience, on the other hand, it matters more to not make an incorrect decision. Therefore, it chooses to remain uninformed (i.e. does not consume the content) if it is unsure about whether the story has been researched or not. Specifically, when the outlets publish simultaneously, the patient audience prefers to remain undecided and uninformed than to take the wrong action.[11]

Given these preferences, the audience picks its most preferred outlet. Type u audience always consumes content from the first outlet to publish the story, and $1 - u$ picks the second one, if there is one. When both of the outlets publish simultaneously, then only u types are available. This audience chooses one of the outlets randomly. Therefore, u is akin to v, or the scoop value from the previous analysis. In addition, we assume that the entire mass of audience is available for reputation building.

u is story-specific and when the outlets get a story they also learn perfectly the value of u. The idea is that those stories with a relatively high u are more urgent than others. These could include, for example, information about whether a company has

[10]Today and tomorrow are essentially defined along $t = 1$ and 2. They both happen before the true story in eventually revealed.

[11]We treat each of the two subgroups that compose the audience as a single entity, as the preferences of their members are identical. If they have different information, e.g. because half of the group consumes one outlet and half the other, and outlets endorse different states, we assume they choose the action by tossing a fair coin. Note that subgroup payoffs depend on the decision of the subgroup, not on the collective decision of the audience as a whole.

gone bankrupt, or whether the police caught the terrorists, etc.

First, observe that nothing changes relative to the monopoly case discussed in Proposition 3.2. As there is no sense of time order, the audience preference for urgency does not alter the equilibrium. However, now the duopoly case looks different. Noting that the belief updating remains the same, the new condition for outlet i conducting research becomes

$$\overbrace{\frac{1}{2}\left[\sigma^j\left(\frac{u}{2}+\gamma^i(\varnothing)\right)+(1-\sigma^j)(1-u+\gamma^i(2))\right]+\frac{1}{2}(1-u+\gamma^i(2))-c^i}^{\text{expected payoff from research}} \geq$$
$$\underbrace{\frac{1}{2}\left[\sigma^j(u+\gamma^i(1))+(1-\sigma^j)\left(\frac{u}{2}+\pi^2\gamma^i(\varnothing)+(1-\pi^2)\gamma^i(1)\right)\right]+\frac{1}{2}\left(\frac{u}{2}+\pi^2\gamma^i(\varnothing)+(1-\pi^2)\gamma^i(1)\right)}_{\text{expected payoff from publication}},$$

which simplifies to

$$c^i \leq \frac{1}{2}\left[\left(\gamma^i(\varnothing)-\gamma^i(1)\right)\left(\sigma^j-(2-\sigma^j)\pi^2\right)+(2-\sigma^j)\left(1-\gamma^i(1)\right)-\sigma^j(1-u)\right]+1-\frac{3}{2}u := \bar{c}^i_D. \tag{3.2}$$

Observe in condition (3.4) that the incentives to research have increased. By researching and being the second one to publish the story, the outlet gets an additional $1-u$ readers on top of building a perfect reputation. Said another way, this dilutes preemption concerns as both the first and the second mover have their respective markets. Therefore, we first need to check if a symmetric and unique equilibrium $\bar{c}_D$ exists à la Proposition 3.4.

Proposition 3.6 *Let there be a fraction u of audience available to the first outlet publishing and let $\bar{c} \geq 2.5$. If there are two media outlets and θ is not known to the audience, there exists a unique and symmetric equilibrium where $\bar{\sigma}^{i*} = \bar{\sigma}^{j*} := \bar{\sigma}^* = F(\bar{c}_D)$ such that*

$$\bar{c}_D = \frac{1}{2}\left[\left(\gamma(\varnothing)-\gamma(1)\right)\left(\bar{\sigma}^*-(2-\bar{\sigma}^*)\pi^2\right)-\bar{\sigma}^*(1-u)\right]+\frac{3}{2}(1-u)$$

where $\gamma(\varnothing) = \frac{(\bar{\sigma}^*)^2+(1-\bar{\sigma}^*)(2-\bar{\sigma}^*)\pi^2}{(\bar{\sigma}^*)^2+(2-\bar{\sigma}^*)^2\pi^2}$ *and* $\gamma(1) = \frac{1-\bar{\sigma}^*}{2-\bar{\sigma}^*}$.

Note that while v was unbounded, $u \in [0,1]$. But an increase in the fraction of urgent audience u still has a negative effect on $\bar{c}_D$ and decreases $\bar{\sigma}^*$. Therefore, a high fraction of impatient audience pushes the outlets towards speed. The next proposition compares the probabilities of research in the no-competition monopoly case with the duopoly case on the basis of u.

Proposition 3.7 *There exists an interior* u*,* $\bar{u} \in (0, 1)$ *such that*

- *for stories with* $u < \bar{u}$*,* $\bar{c}_D > c_M$ *so that research by high outlets in duopoly is more likely than in monopoly* ($\bar{\sigma}_D > \sigma_M$)*;*
- *for stories with* $u > \bar{u}$*,* $\bar{c}_D < c_M$ *so that research by high outlets in duopoly is less likely than in monopoly* ($\bar{\sigma}_D < \sigma_M$)*; and*
- *for stories with* $u = \bar{u}$*,* $\bar{c}_D = c_M$ *so that research by high outlets in duopoly is equally likely as in monopoly* ($\bar{\sigma}_D = \sigma_M$)*.*

We can therefore see that competitive environments are better for research on non-urgent topics. A good example is the recent *New York Times* exposé on sexual abuse in Hollywood. It is reasonable to believe that sexual abuse in the movie industry does not directly impact a large fraction of society. Yet, it was an important finding that will have a long-run impact as women come forward and demand justice, and organizations respond. On the flip side, investigations and research on urgent topics is less likely in competitive environments. The example of terrorist attacks fits perfectly in this setting. In fact, after the Boston Marathon Bombings in April 2013 there was much confusion in the media and articles were published without fact-checking. The intuition is simple: when a large fraction of the audience seeks information quickly, outlets compete to be the first one to publish the news.

We can now also make assessments about the audience's welfare.[12] The audience's welfare V is defined as follows

$$
\begin{aligned}
V = & \left[\left(\frac{1}{2}\right)^2 + 2\frac{1}{4}(1-\bar{\sigma}^*) + \left(\frac{1}{2}\right)^2(1-\bar{\sigma}^*)^2\right]\pi u + \\
& + 2\frac{1}{4}\bar{\sigma}^*\left[1 + (1-\bar{\sigma}^*)\right](1 - u + \pi u) + \left(\frac{1}{2}\right)^2(1-(1-\bar{\sigma}^*)^2)u \\
= & \frac{(4-(\bar{\sigma}^*)^2)}{4}\pi u + \frac{1}{2}\bar{\sigma}^*(2-\bar{\sigma}^*)(1-u) + \frac{1}{4}(1-(1-\bar{\sigma}^*)^2)u.
\end{aligned}
$$

[12]Note that even if the audience knows that outlets may be publishing without research, it is still better to listen to the outlets rather than to follow the priors in decision-making.

The first term is the probability that the two outlets move together but do not research further, i.e. they publish in $t = 1$. As a result, the probability of matching the state is π and only fraction u of the audience gets this payoff. The second term is the probability that the outlets move sequentially, in which case the fraction $1-u$ match the state, but fraction u only match it with probability π. Finally, the third is when both outlets move together in $t = 2$ after researching further. In this case, they match the state perfectly but fraction $1-u$ does not receive this payoff.

As discussed in Section 3.1, another important effect of the Internet has been to make it easier to conduct preliminary research. Emails and social media make it particularly easy to share pictures, video and text from any part of the world. One way to interpret it is as an increase in π or the precision of s. This, Knobel (2018) argues, should lead to better reporting. We show below that that is not necessarily true. Our next proposition shows that the overall effect of an increase in π on V is dependent on the kind of story u being explored.

Proposition 3.8 *There exists an interior u, $\bar{u}^V \in (0,1)$, such that if $u < \bar{u}^V$ an increase in precision π of initial signal s decreases the overall welfare V.*

The intuition for this somewhat surprising result is easy. The equilibrium probability of research falls as precision π increases. This is because a higher π reduces the reputational gain that comes with separation. The audience attributes correctly matching the state more to better initial information that comes costlessly due to better technology rather than actual research. Preemption concerns, therefore, become more salient and push the outlets towards speed. In turn, it hurts the average audience if it is composed of more patient types, i.e. u is low and then π increases.

Formally, the welfare of the urgent audience increases with an increase in π.

$$\frac{\partial V|_u}{\partial \pi} = \frac{(4-\bar{\sigma}^2)}{4} + \left[-\pi\frac{2\bar{\sigma}}{4} + 2\frac{1-\bar{\sigma}}{4}\right]\frac{\partial \bar{\sigma}}{\partial \pi} > 0$$

because $\frac{(4-\bar{\sigma}^2)}{4} > \frac{1-\bar{\sigma}}{2}$. And the welfare of the patient audience reduces due to an increase in π,

$$\frac{\partial V|_{1-u}}{\partial \pi} = (1-\bar{\sigma})\frac{\partial \bar{\sigma}}{\partial \pi} < 0.$$

When the fraction of urgent audience is low enough, an increase in π hurts an average audience member. Better preliminary research is good news for the audience only if separation does not happen and is not desired. However it also discourages separation, which hurts the audience when it is desired.

3.5 Information dissemination by a source

We now turn back to our original model and discuss the case of a strategic source. We can use our model to determine how a source can share information with media outlets.

In general, our strategic source's preferences are summarized by the following objective function,

$$\mathbb{1}\{\text{publication in } t = 1\} + \mu \Pr(\text{matching the state}).$$

Therefore, the source has a preference for speed vs. accuracy. The parameter $\mu \geq 0$ captures the weight that the source places on accurate information from at least one outlet *vis-à-vis* having at least one outlet publishing in period 1. For instance, a concerned citizen or an employee in a firm witnessing some wrongdoing might have a high preference for accuracy. On the flip side, a politically-motivated source who merely wants to get some potentially incorrect information out quickly will have a low preference for accuracy. We want to determine whether a source wants to share information with one or both the outlets to fulfill her objective.

In line with our model, we will assume that if the source shares a story with both of the outlets, both are aware that the other also possesses the same story. Therefore, the information is shared "publicly".[13] But when the source shares information with just one outlet, we will assume that the other is unaware. This allows the outlet with a story to effectively behave as a monopolist from our analysis in Section 3.3.1. In addition, we assume that the source possesses a story of a fixed precision π. She makes her decision about who to share the story with at the beginning of the game before time 0. The type of the outlet is still each outlet's private information; the source does not have this information when making her decision.

[13] One may imagine a politician revealing some negative evidence about a competitor on Twitter as an example. It is common for news outlets to pick up this information and relay it, either as is or after further fact-checking and investigations.

First, we make a simple observation that follows from our analysis of monopoly and duopoly. (In what follows, we drop the star notation for convenience with an understanding that we are talking about equilibrium values.)

Corollary 3.3 *The equilibrium probability of research by a high outlet in monopoly is* $\sigma_M > 0$ *while in duopoly is* $\sigma_D \geq 0$.

Corollary 3.3 is an important one. It highlights that while in monopoly the probability of research is always positive; in duopoly it might be zero if v is sufficiently high. This corollary will help us outline the behavior of a source who is aware of how high v is associated with her story.

Second, we write down the expected utility of the source for the equilibrium research probabilities that will be induced in the following subgame. The expected payoff from sharing information with one outlet is

$$\frac{1}{2}(1 + \mu\pi) + \frac{1}{2}[\sigma_M \mu + (1 - \sigma_M)(1 + \mu\pi)] \tag{3.3}$$

The first term reflects what the source gets if she gives the story to a low quality outlet, and the second term is for giving it to a high quality outlet. Similarly, the expected payoff from sharing information with both the outlets is

$$\frac{1}{4}(1+\mu\pi)+\frac{1}{4}[1+\mu(\sigma_D+(1-\sigma_D)\pi)]2+\frac{1}{4}[(1-\sigma_D)^2(1+\mu\pi)+2\sigma_D(1-\sigma_D)(1+\mu)+\sigma_D^2\mu]. \tag{3.4}$$

Again, the first term reflects the source's payoff from facing two low type outlets. The second is the payoff from facing one high type and one low type outlet. Note that in this case the story is always published in the first period, but the high outlet matches the state only if it does research. The third term is the payoff from facing two high type outlets. Here, the possible situations are that neither researches; one researches, or both research. The following lemma helps simplify the source's optimal response for a given σ_M and σ_D.

Lemma 3.3 *The source's best response can be summarized as follows:*

- *The source prefers to share the story with both the outlets unambiguously for any* $\mu \geq 0$ *if* $\frac{\sigma_D^2}{2} \leq \sigma_M \leq \frac{\sigma_D(4-\sigma_D)}{2}$.

- *Otherwise, the source prefers to share the story with both outlets if*

$$\mu(1-\pi)(2\sigma_M - \sigma_D(4-\sigma_D)) \leq 2\sigma_M - \sigma_D^2$$

The lemma shows that there is a range of equilibrium σ_M and σ_D for which the source always prefers to send information to both the outlets independent of μ. Interestingly, this region lies around the $\sigma_D = \sigma_M$ line. Therefore, the lemma shows that for σ_M and σ_D close to each other there is reason to prefer both outlets. To understand why, let us break this down into two further statements.

First, there are parameters where one outlet alone is more likely to research than when it is competing with another (i.e. $\sigma_M > \sigma_D$) and μ is very large, and yet the source prefers to share the story with two outlets. This happens because a lower σ_D is compensated by a higher probability of investigation from more firms. But this requires σ_D and σ_M to be close to each other. To see this, let us compare the total probability of research (and matching the state) from sharing the story with one vs. both the outlets. When shared with one it is equal to $\frac{1}{2}\sigma_M$. When shared with both it is given by

$$\frac{1}{4}[\sigma_D^2 + 2\sigma_D(1-\sigma_D)] + \frac{2}{4}\sigma_D = \sigma_D - \frac{\sigma_D^2}{4}.$$

Therefore, despite $\sigma_M > \sigma_D$ the source shares the story with both outlets if $\sigma_D - \frac{\sigma_D^2}{4} \geq \frac{1}{2}\sigma_M$. This condition simplifies to give us our upper bound

$$\sigma_M \leq \frac{\sigma_D(4-\sigma_D)}{2}.$$

Second, there are parameters where one firm is less likely to research than two (i.e. $\sigma_M < \sigma_D$) and μ is very low, and yet the source prefers to share the story with two outlets. This, on the other hand, happens because competition between two firms ensures the story comes out quicker despite each independent outlet researching with a higher probability. To see this, we now compare the total probabilities of the story being published in $t = 1$ under the two scenarios. When shared with one, this probability is equal to $\frac{1}{2} + \frac{1}{2}(1-\sigma_M) = 1 - \frac{\sigma_M}{2}$. When shared with both it is given by

$$\frac{1}{4} + \frac{2}{4} + \frac{1}{4}[(1-\sigma_D)^2 + 2\sigma_D(1-\sigma_D)] = 1 - \frac{\sigma_D^2}{4}.$$

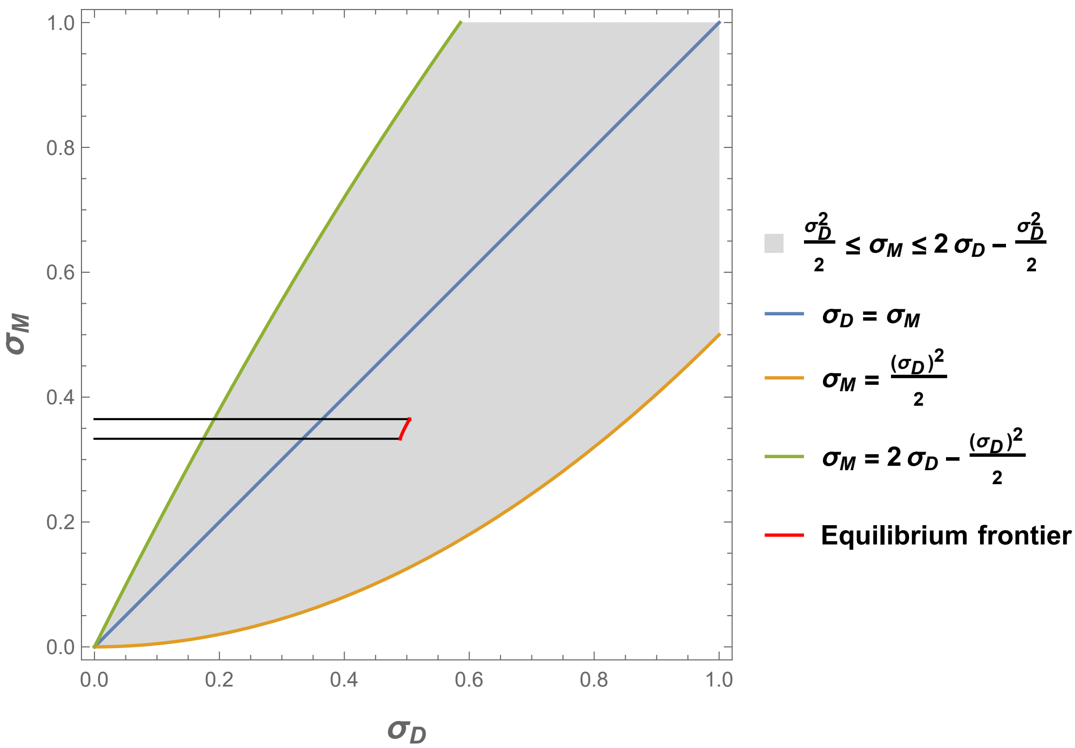
So, now despite $\sigma_M < \sigma_D$ the source shares the story with both outlets if $1 - \frac{\sigma_D^2}{4} \geq 1 - \frac{\sigma_M}{2}$. This condition simplifies to

$$\sigma_M \geq \frac{\sigma_D^2}{2},$$

giving us our lower bound. But note again that for this argument to work σ_D and σ_M should not be too different from each other. When this is the case, then what the source does depends on her preference μ (captured in the second bullet point of Lemma 3.3).

In Figure 3.2, the shaded gray region shows the combinations of σ_D and σ_M where the source always prefers to share stories with both outlets. The region is enclosed between $\sigma_M = \frac{\sigma_D(4-\sigma_D)}{2}$ (green) and $\sigma_M = \frac{\sigma_D^2}{2}$ (orange), which includes $\sigma_M = \sigma_D$ (blue).

Figure 3.2: Equilibria in $\sigma_D - \sigma_M$ space and the behavior of the source.



We now look at possible equilibria that can arise in the $\sigma_D - \sigma_M$ space relative to the source's preferences. We begin by plotting an equilibrium frontier for a given $\bar{c}$ and ε.

Definition 2 (Equilibrium frontier) *The equilibrium frontier is given by the combination of equilibrium σ_D and σ_M generated by varying $\pi \in [.5, 1]$ for $v = 0$ and a fixed $\bar{c}$ and ε.*

The equilibrium frontier, therefore, shows the maximum equilibrium value that σ_D can take for any equilibrium σ_M (since σ_D is decreasing in v from Corollary 3.3 and we are setting $v = 0$). As proved in Lemma 3.2, when $v = 0$, $\sigma_D > \sigma_M$. Therefore, the frontier lies to the right of the 45 degree line. In addition, note that it is upwards sloping. The positive slope is a result of the fact that both σ_M and σ_D are decreasing functions of π.[14] A north-east movement along the frontier arises due to a decrease in π. Figure 3.2 plots the equilibrium frontier for $\bar{c} = 2$ and $\varepsilon = 1$ in red.[15]

Once we have the equilibrium frontier, it is easy to see the set of all possible equilibrium values that might arise for different parameter ranges. Particularly, increasing v is a leftward movement from the frontier along the same σ_M. For v sufficiently high, $\sigma_D = 0$ while $\sigma_M > 0$ (see Corollary 3.3). We are now left with comparing these equilibrium values with what the source wants.

Proposition 3.9 *For a source with preferences given by $\mu \geq 0$,*

- *there exists an $\varepsilon > 0$ small enough and two thresholds $\bar{v} > \underline{v}$ such that if $v < \underline{v}$ the source sends the story to both if $\mu \geq \frac{\sigma_D^2 - 2\sigma_M}{\sigma_D(4-\sigma_D) - 2\sigma_M} \frac{1}{(1-\pi)}$, if $\bar{v} \geq v \geq \underline{v}$ the source always sends the story both, and if $v > \bar{v}$ the source sends the information to both if $\mu \leq \frac{2\sigma_M - \sigma_D^2}{2\sigma_M - \sigma_D(4-\sigma_D)} \frac{1}{(1-\pi)}$;*
- *there exists an $\varepsilon > 0$ large enough and a threshold $\bar{\bar{v}}$ such that if $v \leq \bar{\bar{v}}$ the source sends the story to both, and if $v > \bar{\bar{v}}$ the source sends the information to both if $\mu \leq \frac{2\sigma_M - \sigma_D^2}{2\sigma_M - \sigma_D(4-\sigma_D)} \frac{1}{(1-\pi)}$.*

Our third main result follows by setting $\mu = 0$ in the above proposition. It pertains to the situation where the source only cares about getting the story out quickly independent of whether it is accurate or not. Political actors are often interested in doing so to highlight their achievements or to bring out potentially damaging information

[14] The proofs have been omitted from the main text for the sake of brevity.

[15] We choose a high value of ε for graphical representation only. When ε is low, the range of σ_M and σ_D is also small, and it becomes difficult to clearly see the equilibria graphically.

about their competitors. Twitter and other social media platforms are one way to communicate such stories, which are then picked up by media outlets and relayed to the public without further research.

Corollary 3.4 *When the source does not care about accuracy, i.e.* $\mu = 0$*,*

- *there exists an* $\varepsilon > 0$ *small enough and* $\bar{v}$ *such that for* $v < \bar{v}$*, the source sends the story to one outlet, and sends to two in all other cases, and*
- *there exists an* $\varepsilon > 0$ *large enough such that the source sends the story to both outlets.*

The proof of both Proposition 3.9 and its corollary is by construction. The idea is that when ε is small, (at least a part of) the frontier lies below the orange line in Figure 3.2. Therefore, there arise two thresholds on v where only the middle part lies between the two curves. For a ε high enough, there is only one threshold on v as depicted in the figure. One can easily get the result of Corollary 3.4 by setting $\mu = 0$ in Proposition 3.9.

Consider the intuition for the case of $\mu = 0$. When the intrinsic motivation to conduct research is high then independent of whether only one outlet has the story or both, the outlets are more likely to conduct research. This is, however, not something a $\mu = 0$ source desires. By sending to both, she is able to create preemption risk as well (even for a low v). This improves on the situation of sending to one as the outlets are more driven towards speed. On the other hand, when intrinsic motivation is low, outlets are less likely to research. Now, the source does not always want to share the story with both. Notably, when v is low the source wants to share information with just one. Sending to both risks the outlets trying to separate by doing research, thereby increasing the overall probability of research. However, again when v is high, the source is happy to share the story with both as preemption concerns will become salient for the outlets.[16]

[16] The intuition for the general case presented in Proposition 3.9 is similar but it is not easy to make sharp predictions like we could with $\mu = 0$ case. However, some additional predictions can be made by choosing specific μ values. For instance, when v is very high so that $\sigma_D = 0$, the source prefers to send to one outlet only if $\mu > 2$.

3.6 Conclusion

There have been increasing concerns in the past decade about how the Internet has altered the incentives of media outlets. Notably, media critics have argued that increasing competition in the Interent era has pushed outlets towards speed-driven journalism. Our model shows that conventional wisdom about the effect of competition and the modern digital environment on the media market should be taken *cum grano salis*. We prove that competition in itself may make it easier for high quality outlets to engage in more research-driven journalism to separate themselves from the low quality outlets. For this to happen, it must be that the action of one of the outlets is somehow informative about the type of the other. This result and intuition finds support in some of the new media studies literature such as in Knobel (2018) and Carson (2019).

It is, however, worth emphasizing the importance of a "sophisticated" audience in generating the better-reporting result. We need the audience to place importance on the accuracy of stories, and not always seek quick information. Gentzkow and Shapiro (2008) suggest that scoop value is usually not too high in the media markets. But at the same time, some media scholars have argued that the audience usually seeks information earlier on social media. Similarly, our model shows the importance of the audience observing the *sequence* of publication. This might also be an issue if technology perfectly "flattens the digital plain" (see Section 3.1.1).

Our paper is one of the first to incorporate preemption and reputation concerns in a single model by thinking of a natural setting where both incentives play a role. However, there is further scope for research here. For instance, one may expand the model to include news media bias. Bias and the speed-accuracy tradeoff can interact in interesting ways. If bias makes reputational gains less salient (e.g. because future readership does not depend on reputation) then it should push toward speed. On the other hand, if bias implies a less informative publication and hence a smaller "scoop value", then it may actually push toward accuracy.

Our model also produces important testable predictions about how the modern digital environment has altered the media landscape. First, we should see better reporting of non-urgent issues in the Internet-age as the outlets try to build a reputation on such stories. Second, the effect of the Internet on the reporting of breaking-news type stories is ambiguous. It might improve because of better source information but might deteriorate because of more time pressure.

Appendices

A3 Proofs from the main text

Proof of Observation 3.1.

Suppose that the outlet chooses $d = pub$. Without loss of generality, suppose that $s^i = a$. It is easy to see that $\Pr(\omega = a|s^i = a) > \Pr(\omega = b|s^i = a)$ because

$$\frac{\pi\frac{1}{2}}{\pi\frac{1}{2} + (1-\pi)\frac{1}{2}} > \frac{(1-\pi)\frac{1}{2}}{\pi\frac{1}{2} + (1-\pi)\frac{1}{2}}$$

which is true because $\pi > \frac{1}{2}$. ■

Proof of Lemma 3.1.

First part. Without loss of generality, suppose that $s^i = a$. Then, if i chooses to publish, it will endorse state a, i.e. send message $m = a$. by Bayes' rule,

$$\Pr(\omega = a|s^i = a) = \frac{\pi\frac{1}{2}}{\pi\frac{1}{2} + (1-\pi)\frac{1}{2}} = \pi$$

as claimed.

Second part. We are interested in the probability that j matches the state from choosing $d = pub$ when i has received a signal s^i. This is equal to

$$\Pr(s^j = a|s^i)\Pr(\omega = a|s^j = a \,\text{and}\, s^i) + \Pr(s^j = b|s^i)\Pr(\omega = b|s^j = b \,\text{and}\, s^i) \quad \text{(A3.1)}$$

Note that, for a generic s^j, by Bayes' rule we have that $\Pr(s^j|s^i) = \frac{\Pr(s^j \,\text{and}\, s^i)}{\Pr(s^i)}$ and

$$\Pr(\omega = s^j|s^j \,\text{and}\, s^i) = \frac{\Pr(s^j \,\text{and}\, s^i|\omega = s^j)\Pr(\omega = s^j)}{\Pr(s^j \,\text{and}\, s^i)}$$

As a consequence, (A3.1) can be simplified to

$$\frac{\Pr(s^j = a \,\text{and}\, s^i|\omega = a)\Pr(\omega = a)}{\Pr(s^i)} + \frac{\Pr(s^j = b \,\text{and}\, s^i|\omega = b)\Pr(\omega = b)}{\Pr(s^i)} \quad \text{(A3.2)}$$

However, since signals are independent conditional on the state,

$$\Pr(s^j \text{ and } s^i|\omega = s^j) = \Pr(s^j|\omega = s^j)Pr(s^i|\omega = s^j)$$

Moreover, $\Pr(s^j|\omega = s^j) = \pi$. Hence, (A3.2) becomes

$$\pi\frac{\Pr(s^i|\omega = a)\Pr(\omega = a) + \Pr(s^i|\omega = b)\Pr(\omega = b)}{\Pr(s^i)} = \pi$$

as claimed. ■

Proof of Proposition 3.2.

Suppose that a high type outlet chooses $d = res$ with probability σ. Reminding ourselves from the main text that

$$\gamma(R) = \frac{\sigma + (1-\sigma)\pi}{\sigma + (1-\sigma)\pi + \pi}$$

$$\gamma(W) = \frac{(1-\sigma)(1-\pi)}{(1-\sigma)(1-\pi) + (1-\pi)} = \frac{1-\sigma}{2-\sigma}$$

from Bayes' rule and using the fact that a low type outlet always chooses *pub*.

A high type outlet optimally chooses *res* if

$$\gamma(R) - c \geq \pi\gamma(R) + (1-\pi)\gamma(W) \implies c \leq (1-\pi)(\gamma(R) - \gamma(W)) := c_M$$

In equilibrium the conjectured σ must be equal to the actual one, hence it must be that

$$\sigma^* = F(c_M(\sigma^*)) = \frac{c_M(\sigma^*) + \varepsilon}{\bar{c} + \varepsilon}. \tag{A3.3}$$

We need to check if such a fixed point exists. To do so, three observations are in order. First, note that both the LHS and RHS of the above are continuous in σ^*. Second, LHS$(\sigma^* = 0) = 0 <$ RHS$(\sigma^* = 0) = \frac{\varepsilon}{\bar{c}+\varepsilon}$ (as $c_M = 0$ at $\sigma^* = 0$). Third, LHS$(\sigma^* = 1) = 1 >$ RHS$(\sigma^* = 1) = F(\frac{1-\pi}{1+\pi})$. Therefore, the above is true.

Finally, we need to check for the uniqueness of the fixed point. Note that

$$\frac{\partial \text{RHS}}{\partial \sigma^*} = \frac{1-\pi}{\bar{c}+\varepsilon}\left[\frac{\pi(1-\pi)}{(\sigma^*+(1-\sigma^*)\pi+\pi)^2}+\frac{1}{(2-\sigma^*)^2}\right] > 0,$$

but the sign of

$$\frac{\partial^2 \text{RHS}}{\partial (\sigma^*)^2} = \frac{2(1-\pi)}{\bar{c}+\varepsilon}\left[-\frac{\pi(1-\pi)^2}{(\sigma^*+(1-\sigma^*)\pi+\pi)^3}+\frac{1}{(2-\sigma^*)^3}\right]$$

is not clear immediately. $\frac{\partial^2 \text{RHS}}{\partial (\sigma^*)^2} > 0$ requires

$$-\pi(1-\pi)^2(2-\sigma^*)^3+(\sigma^*+(1-\sigma^*)\pi+\pi)^3 > 0 \tag{A3.4}$$

It is easy to see that the LHS of (A3.4) is strictly increasing in σ^* for all $\pi \in (0.5, 1]$. Moreover, the LHS of (A3.4) when we substitute $\sigma^* = 0$ is $-1+2\pi > 0$. As a consequence, the RHS of (A3.3) is strictly increasing and convex. Combined with the above, it means that there is only one fixed point in the $[0, 1]$ interval. ■

Proof of Proposition 3.3.

If θ is known, then by choosing *pub* in $t = 1$ a high quality outlet receives a payoff of

$$\frac{1}{2}\frac{v}{2}+\frac{1}{2}\left[v\sigma+\frac{v}{2}(1-\sigma)\right]+\mathbb{1}\{\theta = h\},$$

where σ is the (symmetric) probability that the high quality competitor engages in more research. By instead choosing *res* and publishing in $t = 2$ a high type outlet gets a payoff of $\frac{1}{2}\sigma\frac{v}{2}+\mathbb{1}\{\theta = h\}-c$. Comparing the two, each outlet is willing to investigate if $c \leq -\frac{v}{2}$. As a consequence, $\sigma_D^* = F\left(-\frac{v}{2}\right)$ in symmetric equilibrium. Research happens with positive probability when $-\frac{v}{2} > -\varepsilon$, which can be rearranged to $v < 2\varepsilon$. ■

Proof of Proposition 3.4.

We complete this proof in several steps. To begin with, we conjecture that whenever an outlet chooses to publish, it is optimal to endorse the state suggested by the signal. This will be verified at the end of the proof.

Step 1: We begin by showing that in any signal-based equilibria outlets' period 1 decisions on whether to research or publish is described by a threshold on c. This follows from the discussion in the text. Let σ^i and σ^j be the conjectured strategies. Then equation (3.1) defines the threshold c_D^i for outlet i.

$$c^i \leq \frac{1}{2}\left[\left(\gamma^i(\varnothing) - \gamma^i(1)\right)\left(\sigma^j - (2-\sigma^j)\pi^2\right) + (2-\sigma^j)\left(1-\gamma^i(1)\right)\right] - \frac{1}{2}v := c_D^i \quad (3.1)$$

where $\gamma(\varnothing) = \frac{\sigma^i\sigma^j + (1-\sigma^i)(2-\sigma^j)\pi^2}{\sigma^i\sigma^j + (2-\sigma^i)(2-\sigma^j)\pi^2}$ and $\gamma(1) = \frac{1-\sigma^i}{2-\sigma^i}$. The problem is identical for player j.

Step 2: Next, we show that for any σ^j there is only one σ^i that solves the equilibrium fixed point for player i.

Given that cost is uniformly distributed in $[-\varepsilon, \bar{c}]$ and that, in equilibrium the conjectured probability of investigation must be equal to the actual probability, the equilibrium levels of σ^i and σ^j must be the solutions of

$$\sigma^i = F(c_D^i(\sigma^i, \sigma^j)) \text{ and } \sigma^j = F(c_D^j(\sigma^j, \sigma^i))$$

where

$$F(c_D^i(\sigma^i,\sigma^j)) = \begin{cases} 0 & c_D^i(\sigma^i,\sigma^j) < -\varepsilon \\ \frac{c_D^i(\sigma^i,\sigma^j)+\varepsilon}{\bar{c}+\varepsilon} & -\varepsilon \leq c_D^i(\sigma^i,\sigma^j) \leq \bar{c} \\ 1 & c_D^i(\sigma^i,\sigma^j) > \bar{c} \end{cases}$$

and

$$f(c_D^i(\sigma^i,\sigma^j)) = \begin{cases} 0 & c_D^i(\sigma^i,\sigma^j) < -\varepsilon \\ \frac{1}{\bar{c}+\varepsilon} & -\varepsilon \leq c_D^i(\sigma^i,\sigma^j) \leq \bar{c} \\ 0 & c_D^i(\sigma^i,\sigma^j) > \bar{c} \end{cases}$$

We want to show that, for every σ^j, there is only one σ^i that solves $\sigma^i = F(c_D^i(\sigma^i, \sigma^j))$.

1. The LHS is linear, with slope equal to 1, starting at 0 and ending at 1.

2. As $c_D^i(\sigma^i = 1, \sigma^j) < 1 < \bar{c}$, the RHS evaluated at $\sigma^i = 1 < 1 =$ LHS at $\sigma^i = 1$;

3. The RHS evaluated at $\sigma^i = 0$ is greater than or equal to zero.

4. For any σ^j, both LHS and RHS are continuous in σ^i.

Hence, they cross at least once and there is at least one solution to this fixed point problem.

To show that they cross only once, we need to show that the slope of the RHS is never above 1. First, note that the slope of the RHS is either 0 or $f(c_D^i)\frac{\partial c_D^i}{\partial \sigma^i}$. Second, $\frac{\partial \gamma^i(\varnothing)}{\partial \sigma^i} = \frac{(\sigma^j-(2-\sigma^j)\pi^2)\pi^2(2-\sigma^j)}{(\sigma^i\sigma^j+(2-\sigma^i)(2-\sigma^j)\pi^2)^2}$, whose sign depends on the sign of $(\sigma^j-(2-\sigma^j)\pi^2)$ and $\frac{\partial \gamma^i(1)}{\partial \sigma^i} = \frac{-1}{(2-\sigma^i)^2} < 0$. Using these we can write $\frac{\partial c_D^i}{\partial \sigma^i} = \frac{1}{2}\left[\frac{(\sigma^j-(2-\sigma^j)\pi^2)^2\pi^2(2-\sigma^j)}{(\sigma^i\sigma^j+(2-\sigma^i)(2-\sigma^j)\pi^2)^2} + \frac{2-\pi^2(2-\sigma^j)}{(2-\sigma^i)^2}\right]$ where both terms are always positive. Third, we can show that the sign of $\frac{\partial^2 c_D^i}{\partial(\sigma^i)^2}$ is ambiguous, but $\frac{\partial^3 c_D^i}{\partial(\sigma^i)^3} \geq 0$. As a consequence, the second derivative is always increasing in σ^i and the first derivative is convex in σ^i. So, $\frac{\partial c_D^i}{\partial \sigma^i}|_{\sigma^i=1} > \frac{\partial c_D^i}{\partial \sigma^i}|_{\sigma^i=0}$, and c_D^i reaches its steepest point around $\sigma^i = 1$. Therefore, it is enough to show that $\frac{\partial c_D^i}{\partial \sigma^i}|_{\sigma^i=1} \leq 1$. This requires

$$2(\sigma^j+(2-\sigma^j)\pi^2)^2 \geq (\sigma^j-(2-\sigma^j)\pi^2)^2\pi^2(2-\sigma^j)+(\sigma^j+(2-\sigma^j)(1-\pi^2)(\sigma^j+(2-\sigma^j)\pi^2)$$

which further simplifies to

$$(\sigma^j + (2-\sigma^j)\pi^2)^2(2-\sigma^j-2+\sigma^j) \geq -4\sigma^j(2-\sigma^j)^2\pi^4.$$

This latter condition is always verified (strictly for positive σ^j, weakly when $\sigma^j = 0$).

Now, combining the above with the fact that $c_D^i(\sigma^i = 1, \sigma^j) < 1$, implies that they cannot cross more than once.

Step 3: Third, we show that if an equilibrium exists, it is unique for $\bar{c} \geq 2$.

Define $\hat{\sigma}^i(\sigma^j)$ the optimal σ^i for a given σ^j. In equilibrium, it must be that

$$\hat{\sigma}^i(\hat{\sigma}^j(\sigma^i)) = \sigma^i \tag{A3.5}$$

Rearranging, the equilibrium is the solution of $\hat{\sigma}^i(\hat{\sigma}^j(\sigma^i)) - \sigma^i = 0$. Differentiating with respect to σ^i, we obtain $\frac{\partial \hat{\sigma}^i}{\partial \hat{\sigma}^j}\frac{\partial \hat{\sigma}^j}{\partial \sigma^i} - 1 = 0$. For the equilibrium to be unique (conditional on its existence), it is now sufficient to show that the LHS is negative. This implies that only one fixed point of (A3.5) can be found. This happens when $\frac{\partial\hat{\sigma}^i}{\partial\hat{\sigma}^j}$ and $\frac{\partial\hat{\sigma}^j}{\partial\sigma^i}$ are between -1 and 1. As the players are identical, it is enough to show that this holds for one of them.

To show the above, begin by noting that $\sigma^i(\sigma^j)$ is implicitly defined by the unique solution of $\sigma^i - F(c^i_D(\sigma^i,\sigma^j)) = 0$. (Going forward we drop the ˆ notation with an understanding that we are concerned with optimal responses.) By implicit function theorem,

$$\frac{\partial\sigma^i}{\partial\sigma^j} = \frac{\frac{\partial F(c^i_D)}{\partial\sigma^j}}{1-\frac{\partial F(c^i_D)}{\partial\sigma^i}} \tag{A3.6}$$

Consider first the denominator of (A3.6). From Step 2, we know that it is always positive. Moreover, it will be smaller the bigger is $\frac{\partial F(c^i_D)}{\partial\sigma^i}$. On the other hand, it is the biggest when $\frac{\partial F(c^i_D)}{\partial\sigma^i}$ is zero. When $\frac{\partial F(c^i_D)}{\partial\sigma^i}$ is non-zero, it is linear and increasing in $\frac{\partial c^i_D}{\partial\sigma^i}$. As this reaches its maximum for $\sigma^i = 1$, we simply replace it and look for a maximum with respect to σ^j.

$$\begin{aligned}
max_{\sigma^j}\frac{\partial c^i_D}{\partial\sigma^i}|_{\sigma^i=1} &= \frac{1}{2}\left[\frac{(\sigma^j-(2-\sigma^j)\pi^2)^2\pi^2(2-\sigma^j)}{(\sigma^j+(2-\sigma^j)\pi^2)^2}+2-\pi^2(2-\sigma^j)\right]\\
&= \frac{1}{2}\left[2-\frac{4\sigma^j(2-\sigma^j)^2\pi^4}{(\sigma^j+(2-\sigma^j)\pi^2)^2}\right]\\
&= 1
\end{aligned}$$

where the second equality is a rearrangement and the third one follows from the fact that this is maximized for $\sigma^j = 0$.

As a consequence, $max_{\sigma^i,\sigma^j}\frac{\partial F(c^i_D)}{\partial\sigma^i} = \frac{1}{\bar{c}+\varepsilon}$ and the smallest the denominator can be is $\frac{1}{\bar{c}+\varepsilon}$.

Second, consider the numerator. $\frac{\partial F(c^i_D)}{\partial\sigma^j}$ is either zero or $\frac{1}{\bar{c}+\varepsilon}\frac{\partial c^i_D}{\partial\sigma^j}$. Further, note that

$$\frac{\partial c^i_D}{\partial\sigma^j} = \frac{1}{2}\left[\frac{\partial\gamma^i(\varnothing)}{\partial\sigma^j}(\sigma^j-(2-\sigma^j)\pi^2)+\gamma^i(\varnothing)+\pi^2(\gamma^i(\varnothing)-\gamma^i(1))-1\right]. \tag{A3.7}$$

Finding the overall maximum and minimum is complicated, so we look for sufficient conditions. We start out by looking at $\frac{\partial\gamma^i(\varnothing)}{\partial\sigma^j}$. After few algebric manipulations, we derive

$$\frac{\partial\gamma^i(\varnothing)}{\partial\sigma^j} = \frac{2\sigma^i\pi^2}{(\sigma^i\sigma^j+(2-\sigma^i)(2-\sigma^j)\pi^2)^2}$$

Its sign is positive, but it is hard to determine the maximum. We proceed as follows. First, note that

$$\frac{\partial^2\gamma^i(\varnothing)}{\partial(\sigma^j)^2} = \frac{-4(\sigma^i - (2-\sigma^i)\pi^2)\sigma^i\pi^2}{(\sigma^i\sigma^j + (2-\sigma^i)(2-\sigma^j)\pi^2)^3}$$

whose sign is ambiguous. However,

$$\frac{\partial^3\gamma^i(\varnothing)}{\partial(\sigma^j)^3} = \frac{12(\sigma^i - (2-\sigma^i)\pi^2)^2\sigma^i\pi^2}{(\sigma^i\sigma^j + (2-\sigma^i)(2-\sigma^j)\pi^2)^4}$$

which is positive. This implies that (for any σ^i) $\frac{\partial\gamma^i(\varnothing)}{\partial\sigma^j}$ is a convex function in σ^j which is maximized either at $\sigma^j = 0$ or at $\sigma^j = 1$. By substitution,

$$\begin{aligned}\frac{\partial\gamma^i(\varnothing)}{\partial\sigma^j}|_{\sigma^j=0} &= \frac{\sigma^i}{2\pi^2(2-\sigma^i)^2}\\ \frac{\partial\gamma^i(\varnothing)}{\partial\sigma^j}|_{\sigma^j=1} &= \frac{2\sigma^i\pi^2}{(\sigma^i + (2-\sigma^i)\pi^2)^2}\end{aligned}$$

Still we are left to determine the maximum possible value of $\frac{\partial\gamma^i(\varnothing)}{\partial\sigma^j}$ because the comparison is not straightforward. But we can show that for every π, $max_{\sigma^i}\frac{\partial\gamma^i(\varnothing)}{\partial\sigma^j}|_{\sigma^j=0} > max_{\sigma^i}\frac{\partial\gamma^i(\varnothing)}{\partial\sigma^j}|_{\sigma^j=1}$. To prove this, first see that

$$max_{\sigma^i}\frac{\partial\gamma^i(\varnothing)}{\partial\sigma^j}|_{\sigma^j=0} = \frac{1}{2\pi^2}$$

But to get $max_{\sigma^i}\frac{\partial\gamma^i(\varnothing)}{\partial\sigma^j}|_{\sigma^j=1}$,

$$\frac{\partial}{\partial\sigma^i}\left(\frac{\partial\gamma^i(\varnothing)}{\partial\sigma^j}|_{\sigma^j=1}\right) = \frac{\partial}{\partial\sigma^i}\left(\frac{2\sigma^i\pi^2}{(\sigma^i + (2-\sigma^i)\pi^2)^2}\right) = \frac{2\pi^2(\sigma^i + (2-\sigma^i)\pi^2) - 4(1-\pi^2)\sigma^i\pi^2}{(\sigma^i + (2-\sigma^i)\pi^2)^3} \tag{A3.8}$$

Note that the relevant expression in (A3.8) is always positive for $\sigma^i \leq \frac{2\pi^2}{1-\pi^2}$. For a sufficiently high π, this includes the whole range of values of σ^i. Hence, the function is maximised at $\sigma^i = 1$, and

$$max_{\sigma^i}\frac{\partial\gamma^i(\varnothing)}{\partial\sigma^j}|_{\sigma^j=1} = \frac{2\pi^2}{(1+\pi^2)^2}.$$

But now it is easy to see that $\frac{1}{2\pi^2} \geq \frac{2\pi^2}{(1+\pi^2)^2}$ requires $1 + 2\pi^2 - 3\pi^4 \geq 0$, which is always true for $\pi \in (0.5, 1]$. Therefore, our claim of $max_{\sigma^i}\frac{\partial\gamma^i(\varnothing)}{\partial\sigma^j}|_{\sigma^j=0} > max_{\sigma^i}\frac{\partial\gamma^i(\varnothing)}{\partial\sigma^j}|_{\sigma^j=1}$ is true.

However, for low π, we have that $argmax_{\sigma^i} \frac{\partial \gamma^i(\varnothing)}{\partial \sigma^j}|_{\sigma^j=1} = \frac{2\pi^2}{1-\pi^2} \in [0,1]$. In particular, this happens for $\pi^2 \leq \frac{1}{3}$. Even in this case, it is easy to show that $\frac{1}{2\pi^2} \geq \frac{2\pi^2\left(\frac{2\pi^2}{1-\pi^2}\right)}{\left((1-\pi^2)\left(\frac{2\pi^2}{1-\pi^2}\right)+2\pi^2\right)^2}$ requires $\pi^2 \leq \frac{2}{3}$, i.e. it is always the case in the range of parameters of interest. As a consequence, we have that $max_{\sigma^i} \frac{\partial \gamma^i(\varnothing)}{\partial \sigma^j}|_{\sigma^j=0} > max_{\sigma^i} \frac{\partial \gamma^i(\varnothing)}{\partial \sigma^j}|_{\sigma^j=1}$. Since we want $\frac{\partial \gamma^i(\varnothing)}{\partial \sigma^j}$ as big as possible, we can set it as $\frac{1}{2\pi^2}$ for our sufficiency conditions.

Given this, the lowest value the numerator of $\frac{\partial \sigma^i}{\partial \sigma^j}$ from (A3.6) can be found by making the relevant replacement from above to (A3.7). Therefore,

$$min_{\sigma^i,\sigma^j} \frac{\partial F(c_D^i)}{\partial \sigma^i} \geq \frac{1}{\bar{c}+\varepsilon}\frac{1}{2}\left[\frac{1}{2\pi^2}(-2\pi^2)-1\right] = \frac{-1}{\bar{c}+\varepsilon}.$$

To see this, note that $min_{\sigma^i,\sigma^j}(\sigma^j-(2-\sigma^j)\pi^2) = -2\pi^2$, $min_{\sigma^i,\sigma^j}\gamma^i(\varnothing) \geq 0$, $min_{\sigma^i,\sigma^j}(\gamma^i(\varnothing)-\gamma^i(1)) \geq 0$. Therefore, our first sufficient condition for the uniqueness of the equilibrium is

$$\frac{-\frac{1}{\bar{c}+\varepsilon}}{1-\frac{1}{\bar{c}+\varepsilon}} > -1,$$

which simplifies to $\bar{c} \geq 2$, as assumed.

Looking now at the upper bound, again by replacing in (A3.7) note that

$$\begin{aligned} max_{\sigma^i,\sigma^j} \frac{\partial F(c_D^i)}{\partial \sigma^i} &\leq \frac{1}{\bar{c}+\varepsilon}\frac{1}{2}\left[\frac{1}{2\pi^2}(1-\pi^2)+1+\pi^2-1\right] = \\ &= \frac{1}{2(\bar{c}+\varepsilon)}\left[\frac{1-\pi^2}{2\pi^2}+\pi^2\right] \end{aligned}$$

To see this, note that $max_{\sigma^i,\sigma^j}(\sigma^j-(2-\sigma^j)\pi^2) = 1-\pi^2$, $max_{\sigma^i,\sigma^j}\gamma^i(\varnothing) \leq 1$, $max_{\sigma^i,\sigma^j}(\gamma^i(\varnothing)-\gamma^i(1)) \leq 1$. Therefore, our second sufficient condition for the uniqueness of the equilibrium is

$$\frac{\frac{1}{2(\bar{c}+\varepsilon)}\left[\frac{1-\pi^2}{2\pi^2}+\pi^2\right]}{1-\frac{1}{\bar{c}+\varepsilon}} < 1$$

The numerator is maximised at $\pi = \frac{1}{2}$, hence the condition simplifies to $\bar{c}+\varepsilon > \frac{15}{16}$. Again, this is satisfied for $\bar{c} \geq 2$.

Step 4: Fourth, we show that a symmetric equilibrium where $\sigma^{i*} = \sigma^{j*} = \sigma^*$ always exists. Therefore, it is also unique among the set of signal-based equilibria.

Because of symmetry, the equilibrium must be the fixed point of

$$\sigma^i = \sigma^j = \sigma^* = F(c_D(\sigma^*)) \tag{A3.9}$$

where from (3.1)

$$c_D(\sigma^*) = \frac{1}{2}\left[\left(\frac{(\sigma^*)^2 + (1-\sigma^*)(2-\sigma^*)\pi^2}{(\sigma^*)^2 + (2-\sigma^*)^2\pi^2} - \frac{1-\sigma^*}{2-\sigma^*}\right)(\sigma^* - (2-\sigma^*)\pi^2) + 1\right] - \frac{1}{2}v$$

Looking at (A3.9), note that both LHS and RHS are continuous on the $[0,1]$ interval. Moreover, RHS($\sigma^* = 0$) $\geq 0 =$ LHS(σ^*) and $RHS(\sigma^* = 1) < 1 = LHS(\sigma^* = 1)$. As a consequence, there exists a solution in the $[0,1]$ interval. From the previous steps, we know that this solution is unique.

Step 5: Finally, we show that in the symmetric equilibrium it is optimal to endorse the state suggested by the most informative signal.

Assume that player j behaves as in the equilibrium described above. Now, by endorsing the wrong state in period 2 player i shifts beliefs from $\gamma^i(2) = 1$ to $\gamma^i(1)$ if it is the only one publishing in that period, and from $\gamma^i(\varnothing)$ to $\gamma^i(1)$ if both outlets publish in period 2. In both cases, sticking to the correct state is at weakly dominant.

If outlet i chooses to publish in period 1, by endorsing the least likely state outlet i is indifferent if it is the only one to publish in that period. If instead outlet j publishes in period 1 as well, the expected reputation of outlet i by endorsing the state suggested by the signal is $\pi^2\gamma^i(\varnothing) + (1-\pi^2)\gamma^i(1)$. By endorsing the opposite state, the expected reputation is $\pi\gamma^i(1) + (1-\pi)\left[\pi\gamma^i(\varnothing) + (1-\pi)\gamma^i(1)\right]$. Again, the former is strictly bigger than the latter because $\gamma^i(\varnothing) \geq \gamma^i(1)$. ■

Proof of Lemma 3.2.

To show this, we compare the cost threshold in monopoly and duopoly shutting down the preemption concerns, i.e. assuming $v = 0$. We want to show that in this case $c_D > c_M$. This would require

$$\frac{1}{2}\left[(\gamma(\varnothing) - \gamma(1))(\sigma - (2-\sigma)\pi^2) + 1\right] > (1-\pi)(\gamma(R) - \gamma(W)) \tag{A3.10}$$

Observe that $\gamma(1) = \gamma(W) = \frac{1-\sigma}{2-\sigma}$. Moreover, define $\gamma(\varnothing) - \gamma(1) := A$. We can now rearrange equation (A3.10) so that it becomes

$$\frac{1}{2}\left[A\sigma + 1\right] > (1-\pi)(\gamma_R - X) + \frac{1}{2}A(2-\sigma)\pi^2 \tag{A3.11}$$

Now, after the relevant substitutions A can be simplified as $A = \frac{\sigma^2}{(2-\sigma)(\sigma^2+(2-\sigma)^2\pi^2)}$. As a consequence,

$$\frac{\partial A}{\partial \sigma} = \frac{2\sigma(2-\sigma)(\sigma^2+(2-\sigma)^2\pi^2) - \sigma^2(\sigma^2+3(2-\sigma)^2\pi^2 - 2\sigma(2-\sigma))}{((2-\sigma)(\sigma^2+(2-\sigma)^2\pi^2))^2} \tag{A3.12}$$

Signing (A3.12) is not easy in its current form. However, it is clear that $lim_{\sigma\to 0}\frac{\partial A}{\partial \sigma} = 0$. Moreover, we can rearrange A in a more tractable way. In particular, $A = \frac{1}{(2-\sigma)(1+\pi^2B^2)}$ where $B = \frac{2-\sigma}{\sigma}$. Since $B > 0$ and $\frac{\partial B}{\partial \sigma} = -\frac{2}{\sigma^2} < 0$, it is now easy to see that

$$\frac{\partial A}{\partial \sigma} = \frac{1+\pi^2B - 2\pi^2B\frac{\partial B}{\partial \sigma}(2-\sigma)}{((2-\sigma)(1+\pi^2B^2))^2} > 0.$$

The sign of $\frac{\partial^2 A}{\partial \sigma^2}$ is even more complicated, but as A is defined over just two parameters, $\sigma \in [0,1]$ and $\pi \in (0.5, 1]$, we can prove graphically that $\frac{\partial^2 A}{\partial \sigma^2} > 0$. In particular, Figure A3.1 shows that $\frac{\partial^2 A}{\partial \sigma^2}$ (the orange plane) is always strictly above the zero (blue plane) for the entire set of relevant parameters.

It is now straightforward to see that in equation (A3.11) $\frac{\partial \text{LHS}}{\partial \sigma} > 0$ and $\frac{\partial^2 \text{LHS}}{\partial \sigma^2} > 0$ so the LHS is strictly increasing and convex. Moreover, $\frac{\partial RHS}{\partial \sigma} > 0$.

To complete the proof, we show that LHS($\sigma = 0$) >RHS($\sigma = 1$) for all $\pi \in (0.5, 1)$. This requires

$$\frac{1}{2} > \frac{1-\pi}{1+\pi} + \frac{1}{2}\frac{\pi^2}{1+\pi^2}$$

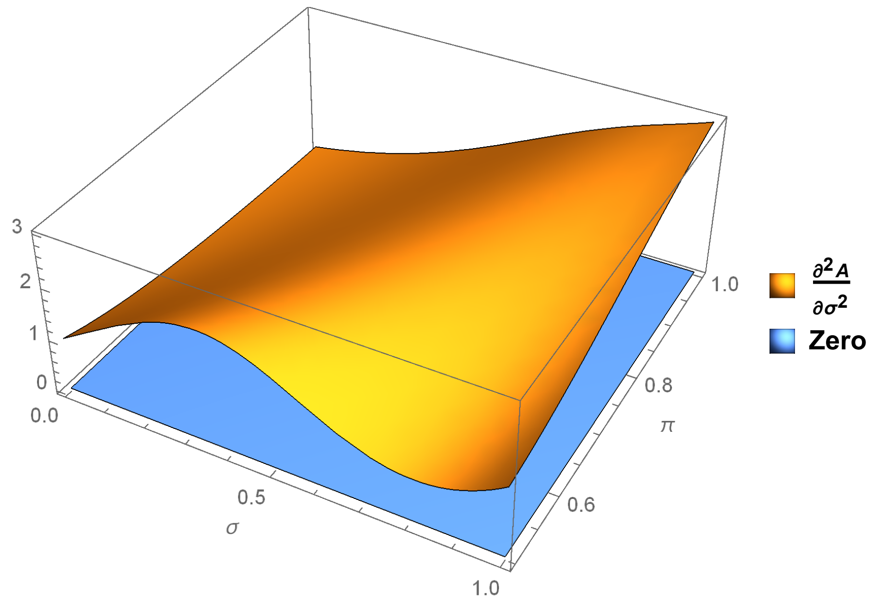
which further simplifies to

$$1 - 3\pi + 2\pi^2 - 2\pi^3 < 0$$

Noticing that the LHS of the above is strictly decreasing in π, and it remains negative for both $\pi = \frac{1}{2}$ and $\pi = 1$, completes the proof. ■

Proof of Proposition 3.5.

Figure A3.1: Proof of Lemma 3.2: Proving $\frac{\partial^2 A}{\partial \sigma^2} > 0$.



Proof of Lemma 3.2: Proving $\frac{\partial^2 A}{\partial \sigma^2} > 0$. Orange plane: $\frac{\partial^2 A}{\partial \sigma^2}$, blue plane: $0.\sigma + 0.\pi$ in $\pi - \sigma$ space.

This follows directly from the strict inequality of equation (A3.10) and the fact that v only reduces its LHS, without affecting the RHS. ■

Proof of Corollary 3.1.

The behavior of the monopolist is unchanged with respect to Section 3.3.1. Looking at the duopoly case, by Bayes' rule

$$\gamma^i(R,.) = \frac{(1-\sigma^i)\pi + \sigma^i}{(1-\sigma^i)\pi + \sigma^i + \pi} = \gamma^i(R)$$

$$\gamma^i(W,.) = \frac{1-\sigma^i}{2-\sigma^i} = \gamma^i(W)$$

Therefore, the cost threshold for research is given by

$$\begin{aligned}
&\frac{1}{2}\left[\sigma^j\left(\frac{v}{2}+\gamma^i(R)\right)+(1-\sigma^j)\gamma^i(R)\right]+\frac{1}{2}\gamma^i(R)-c \geq \\
&\frac{1}{2}\left[\sigma^j\left(v+\pi\gamma^i(R)+(1-\pi)\gamma^i(W)\right)+(1-\sigma^j)\left(\frac{v}{2}+\pi\gamma^i(R)+(1-\pi)\gamma^i(W)\right)\right]+ \\
&+\frac{1}{2}\left(\frac{v}{2}+\pi\gamma^i(R)+(1-\pi)\gamma^i(W)\right),
\end{aligned}$$

which simplifies to

$$c \leq (1-\pi)(\gamma^i(R) - \gamma^i(W)) - \frac{1}{2}v := c'_D \tag{A3.13}$$

Note that the first part of (A3.13) is the same as c_M, and the only term that changes is $-\frac{1}{2}v$, making it smaller than c_M.

In terms of existence and uniqueness of the equilibrium in this set up, note that σ^{i*} and σ^{j*} are the solution of the same fixed point problem, i.e.

$$\sigma^* = F(c'_D(\sigma^*))$$

where $c'_D = c_M - \frac{1}{2}v$. The same logic of the proof of Proposition 3.2 applies here as well. Hence the equilibrium exists and it is unique and symmetric. ■

Proof of Corollary 3.2.

Consider first the case of monopoly. Here, only the high quality outlet can publish in period 2, and this is observable. As a consequence,

$$\gamma(2) = 1$$

$$\gamma(1) = \frac{1-\sigma}{2-\sigma}$$

The monopolist chooses to investigate when $c \leq 1 - \gamma(1) := c''_M$.

In duopoly, the beliefs are updated the same way. Each outlet is considered independently and only the timing matters. The threshold is, therefore, given by

$$\begin{aligned}\frac{1}{2}\left[\sigma^j\left(\frac{v}{2}+1\right)+(1-\sigma^j)\right]+\frac{1}{2}-c \geq & \frac{1}{2}\left[\sigma^j\left(v+\gamma^i(1)\right)+(1-\sigma^j)\left(\frac{v}{2}+\gamma^i(1)\right)\right]+ \\ & +\frac{1}{2}\left(\frac{v}{2}+\gamma^i(1)\right).\end{aligned}$$

It follows then that $c''_D = 1 - \gamma^i(1) - \frac{1}{2}v = c''_M - \frac{1}{2}v < c''_M$ as claimed.

In terms of existence and uniqueness, note that σ^* is the solution of

$$\sigma^* = F(c''(\sigma^*))$$

The RHS is continuous on the $[0,1]$ interval and, irrespective of the market structure, it is either strictly increasing and convex or flat. Moreover, RHS$(\sigma^* = 0) \geq$LHS$(\sigma^* = 0)$ and RHS$(\sigma^* = 1) < 1 =$LHS$(\sigma^* = 1)$ since $\bar{c} > 1$. ■

Proof of Proposition 3.6.

We proceed in steps as outlined in Proposition 3.4. We drop the bars from σ for convenience.

Step 1: We begin by showing that in any signal-based equilibria outlets' period 1 decision on whether to research or publish is described by a threshold on c. This follows from the discussion in the text. Let σ^i and σ^j be the conjectured strategies. Then equation (3.2) defines the threshold c^i_D for outlet i.

$$c^i \leq \frac{1}{2}\left[\left(\gamma^i(\varnothing) - \gamma^i(1)\right)(\sigma^j - (2-\sigma^j)\pi^2) + (2-\sigma^j)\left(1-\gamma^i(1)\right) - \sigma^j(1-u)\right] + 1 - \frac{3}{2}u := \bar{c}^i_D \tag{3.2}$$

where $\gamma(\varnothing) = \frac{\sigma^i\sigma^j + (1-\sigma^i)(2-\sigma^j)\pi^2}{\sigma^i\sigma^j + (2-\sigma^i)(2-\sigma^j)\pi^2}$ and $\gamma(1) = \frac{1-\sigma^i}{2-\sigma^i}$. The problem is identical for player j.

Step 2: Next, we show that for any σ^j there is only one σ^i that solves the equilibrium fixed point for player i.

All of the definitions from Proposition 3.4 remain unaltered.

We want to show that, for every σ^j, there is only one σ^i that solves $\sigma^i = F(\bar{c}^i_D(\sigma^i, \sigma^j))$.

1. The LHS is linear, with slope equal to 1, starting at 0 and ending at 1.
2. Now, $\bar{c}^i_D(\sigma^i = 1, \sigma^j) = c^i_D(\sigma^i = 1, \sigma^j, v = 0) + (1-u)(1 - \frac{\sigma^j}{2})$, where each term is less than or equal to 1. But since $\bar{c} \geq 2.5$, therefore $\bar{c}^i_D(\sigma^i = 1, \sigma^j) < \bar{c}$. As a result, the RHS evaluated at $\sigma^i = 1 < 1 =$ LHS at $\sigma^i = 1$;
3. The RHS evaluated at $\sigma^i = 0$ is greater than or equal to zero.
4. For any σ^j, both LHS and RHS are continuous in σ^i.

Hence, they cross at least once and there is at least one solution to this fixed point problem.

Further, note that $\bar{c}_D^i$ behaves the same way as c_D^i with respect to σ^i. Therefore, the rest of the proof in this step is as before.

Step 3: Third, we show that if an equilibrium exists, it is unique for $\bar{c} \geq 2.5$.

Other than changing the relevant definitions to include σ, nothing changes in this step until we evaluate $\frac{\partial \bar{c}_D^i}{\partial \sigma^j}$

$$\frac{\partial \bar{c}_D^i}{\partial \sigma^j} = \frac{1}{2}\left[\frac{\partial \gamma^i(\varnothing)}{\partial \sigma^j}(\sigma^j - (2-\sigma^j)\pi^2) + \gamma^i(\varnothing) + \pi^2(\gamma^i(\varnothing) - \gamma^i(1)) - (2-u)\right]. \tag{A3.14}$$

Again, the rest of the proof remains unaltered until we find the first sufficient condition. The lowest value of the numerator of $\frac{\partial \sigma^i}{\partial \sigma^j}$ from (A3.6) can be found by making the relevant replacement from above to (A3.14). Therefore,

$$min_{\sigma^i,\sigma^j} \frac{\partial F(\bar{c}_D^i)}{\partial \sigma^j} \geq \frac{1}{\bar{c}+\varepsilon}\frac{1}{2}\left[\frac{1}{2\pi^2}(-2\pi^2) - (2-u)\right] = \frac{-1}{\bar{c}+\varepsilon}\left(\frac{3-u}{2}\right).$$

Therefore, our new first sufficient condition for the uniqueness of the equilibrium is

$$\frac{-\frac{1}{\bar{c}+\varepsilon}\left(\frac{3-u}{2}\right)}{1-\frac{1}{\bar{c}+\varepsilon}} > -1,$$

which simplifies to $\bar{c} \geq \frac{5-u}{2}$. The highest value possible of $\frac{5-u}{2}$ is 2.5 at $u = 0$, which is assumed.

Looking now at the upper bound, again by replacing in (A3.14) we get

$$\begin{aligned} max_{\sigma^i,\sigma^j} \frac{\partial F(\bar{c}_D^i)}{\partial \sigma^i} &\leq \frac{1}{\bar{c}+\varepsilon}\frac{1}{2}\left[\frac{1}{2\pi^2}(1-\pi^2) + 1 + \pi^2 - 2 + u\right] = \\ &= \frac{1}{2(\bar{c}+\varepsilon)}\left[\frac{1}{2\pi^2} + \pi^2 - \frac{3}{2} + u\right]. \end{aligned}$$

Therefore, our second new sufficient condition for the uniqueness of the equilibrium is

$$\frac{\frac{1}{2(\bar{c}+\varepsilon)}\left[\frac{1}{2\pi^2} + \pi^2 - \frac{3}{2} + u\right]}{1-\frac{1}{\bar{c}+\varepsilon}} < 1$$

The numerator is maximised at $\pi = \frac{1}{\sqrt{2}}$, hence the condition simplifies to $\bar{c}+\varepsilon > \frac{u+2}{2}$. Again, this is satisfied for $\bar{c} \geq 2.5$ since $\frac{5-u}{2} > \frac{u+2}{2}$ for $u \in [0,1]$.

Step 4: Fourth, we show that a symmetric equilibrium where $\sigma^{i*} = \sigma^{j*} = \sigma^*$ always exists. Therefore, it is also unique among the set of signal-based equilibria.

Because of symmetry, the equilibrium must be the fixed point of

$$\sigma^i = \sigma^j = \sigma^* = F(c_D(\sigma^*)) \tag{A3.15}$$

where from (3.2)

$$c_D(\sigma^*) = \frac{1}{2}\left[\left(\frac{(\sigma^*)^2 + (1-\sigma^*)(2-\sigma^*)\pi^2}{(\sigma^*)^2 + (2-\sigma^*)^2\pi^2} - \frac{1-\sigma^*}{2-\sigma^*}\right)(\sigma^* - (2-\sigma^*)\pi^2) - \sigma^*(1-u)\right] + \\ - \frac{3}{2}(1-u) \tag{A3.16}$$

Looking at (A3.15), note that both LHS and RHS are continuous on the $[0, 1]$ interval. Moreover, RHS$(\sigma^* = 0) \geq 0 =$ LHS(σ^*) and RHS$(\sigma^* = 1) < 1 =$LHS$(\sigma^* = 1)$. As a consequence, there exists a solution in the $[0, 1]$ interval. From the previous steps, we know that this solution is unique.

Step 5: Finally, we show that in the symmetric equilibrium it is optimal to endorse the state suggested by the most informative signal.

This is true because now there is more incentive to build a reputation. Since reputation requires matching the state, there is even less reason to not endorse the state suggested by the most informative equilibrium. ■

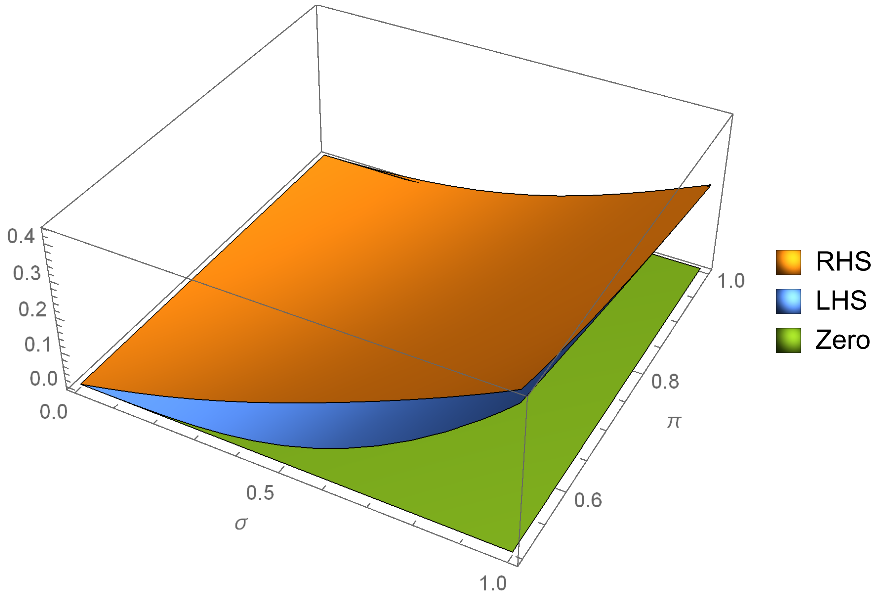
Proof of Proposition 3.7.

We drop the bars for convenience. First, note that $\bar{c}_D$ is strictly decreasing in u. This is so because it can be rearranged as

$$\bar{c}_D = \frac{1}{2}\left[(\gamma(\varnothing) - \gamma(1))(\sigma^* - (2-\sigma^*)\pi^2) - \sigma^*\right] + \frac{3}{2} - u\left(\frac{3}{2} - \frac{\sigma^*}{2}\right)$$

where $\frac{3}{2} - \frac{\sigma^*}{2} > 0$ for any $\sigma^* \in [0, 1]$. Also, c_M and σ_M^* do not change with u.

Figure A3.2: Proof of Proposition 3.7: Proving LHS < RHS.



Proof of Proposition 3.7: Proving LHS < RHS. Orange plane: RHS, blue plane: LHS and green plane: $0.\sigma + 0.\pi$ in the $\pi - \sigma$ space.

Second, consider the case when $u = 1$. We will show that $\bar{c}_D < c_M$. This requires

$$\frac{1}{2}\left[(\gamma(\varnothing) - \gamma(1))(\sigma - (2-\sigma)\pi^2)\right] < (1-\pi)(\gamma(R) - \gamma(W)).$$

Using the terminology introduced in Lemma 3.2, we can rewrite the above as

$$\frac{1}{2}A\sigma < (1-\pi)(\gamma(R) - \gamma(W)) + \frac{1}{2}A(2-\sigma)\pi^2.$$

Now, both the LHS and the RHS of the above equation are only functions of two variables, π and σ. Therefore, we can plot them in a graph (see Figure A3.2) and check that the above is true.

Third, consider the case of $u = 0$. We want to show that $\bar{c}_D > c_M$. This is equivalent to showing that

$$\frac{1}{2}\left[(\gamma(\varnothing) - \gamma(1))(\sigma - (2-\sigma)\pi^2) - \sigma\right] + \frac{3}{2} > (1-\pi)(\gamma(R) - \gamma(W)).$$

We showed in Lemma 3.2 that $c_D(v = 0) > c_M$. It is easy to check that $\bar{c}_D(u = 0) = c_D(v = 0) + 1 - \frac{1}{2}\sigma$ where $1 - \frac{1}{2}\sigma > 0$ for all $\sigma \in [0, 1]$. Therefore, $\bar{c}_D(u = 0) > c_D(v = 0) > c_M$.

Combining the three parts above, our result follows. ■

Proof of Proposition 3.8.

We drop the bars and stars for convenience. Reminding ourselves that

$$V = \frac{(4-\sigma^2)}{4}\pi u + \frac{1}{2}\sigma(2-\sigma)(1-u) + \frac{1}{4}(1-(1-\sigma)^2)u,$$

we first take the first derivative of V with respect to π (we drop the stars and D in what follows for convenience).

$$\begin{aligned}\frac{\partial V}{\partial \pi} &= u\frac{(4-\sigma^2)}{4} + \left[\pi u\frac{(-2\sigma)}{2} + \frac{1}{2}(1-u)2(1-\sigma) + u\frac{(1-\sigma)}{2}\right]\frac{\partial \sigma}{\partial \pi} \\ &= u\frac{(4-\sigma^2)}{4} + \frac{2(1-\sigma) - u(1-\sigma(1-\pi))}{2}\frac{\partial \sigma}{\partial \pi} \end{aligned} \tag{A3.17}$$

Now, we need to show under what conditions $\frac{\partial \sigma}{\partial \pi} < 0$. Reminding that σ is implicitly defined by (A3.15) define

$$K := \sigma - \left[\frac{1}{2}\left[(\gamma(\varnothing)-\gamma(1))(\sigma-(2-\sigma)\pi^2) - \sigma(1-u)\right] + \frac{3}{2}(1-u)\right]\frac{1}{\bar{c}+\varepsilon} - \frac{\varepsilon}{\bar{c}+\varepsilon}$$

Further, using the definitions in the proof of Lemma 3.2, we can rewrite K as

$$K = \sigma - \frac{1}{2(\bar{c}+\varepsilon)}\left[A(\sigma-(2-\sigma)\pi^2) - \sigma(1-u)\right] - \frac{3}{2(\bar{c}+\varepsilon)}(1-u) - \frac{\varepsilon}{\bar{c}+\varepsilon}$$

Differentiating and simplifying, we first obtain

$$\begin{aligned}\frac{\partial K}{\partial \pi} &= -\frac{1}{2(\bar{c}+\varepsilon)}\left[\frac{-2\pi B^2(\sigma-(2-\sigma)\pi^2)}{(2-\sigma)(1+\pi^2B^2)^2} - \frac{2\pi}{1+\pi^2B^2}\right] \\ &= \frac{1}{2(\bar{c}+\varepsilon)}\frac{1+B}{(1+\pi^2B^2)^2} > 0,\end{aligned}$$

and second we obtain

$$\begin{aligned}\frac{\partial K}{\partial \sigma} &= 1 - \frac{1}{2(\bar{c}+\varepsilon)}\left[\frac{\partial A}{\partial \sigma}(\sigma - (2-\sigma)\pi^2) + (1+\pi^2)A - (1-u)\right] \\ &= 1 + \frac{1}{2(\bar{c}+\varepsilon)}(1-u) - \frac{1}{2(\bar{c}+\varepsilon)}\left[\frac{\partial A}{\partial \sigma}(\sigma - (2-\sigma)\pi^2) + (1+\pi^2)A\right] \\ &= 1 + \frac{1}{2(\bar{c}+\varepsilon)}(1-u) - \frac{1}{\bar{c}+\varepsilon}\frac{\partial c_D}{\partial \sigma}\end{aligned}$$

where c_D is the cost threshold we derived in Proposition 3.4.

We can now show that $\frac{\partial c_D}{\partial \sigma} \leq 1$ in the neighborhood of the equilibrium σ. The proof for this is presented in Proposition C3.2 (Appendix C3) for a generic prior q. Therefore, it is also true in our special case of $q = \frac{1}{2}$.

Putting these two facts together and using the Implicit Function Theorem, we can now conclude that $\frac{\partial \sigma}{\partial \pi} < 0$.

Finally, we want to find the condition under which $\frac{\partial V}{\partial \pi} < 0$. From (A3.17), this happens when

$$u\frac{(4-\sigma^2)}{4} < \frac{2(1-\sigma) - u(1-\sigma(1-\pi))}{2}\sigma_\pi,$$

where $\left(-\frac{\partial \sigma}{\partial \pi}\right) := \sigma_\pi > 0$. Now, the LHS of the above equation is linearly increasing in u. Similarly, the RHS is linearly decreasing in u. In addition, for $u = 0$, LHS $<$ RHS and the condition is verified. And for $u = 1$, LHS $>$ RHS. To see this, LHS$(u=1) = 2 - \frac{\sigma^2}{2} > 1$, while RHS$(u=1) = (1-\sigma(1+\pi))\sigma_\pi < 1$. Therefore, $\bar{u}^V$ exists and lies between 0 and 1. ■

Proof of Lemma 3.3.

Comparing the source's expected utility given by expressions in (3.3) and (3.4) and simplifying gives the following condition to prefer two firms:

$$\mu(1-\pi)(2\sigma_M - \sigma_D(4-\sigma_D)) \leq 2\sigma_M - \sigma_D^2 \tag{A3.18}$$

Now we discuss different cases based on possible values of σ_M and σ_D.

Case 1: v is very high so that $\sigma_D = 0$. Substituting in A3.18 gives that the source prefers to send the story to both outlets if

$$\mu \leq \frac{1}{1-\pi} > 1$$

Therefore, if v is very large it is possible that $\mu > 1$ (so that the source cares relatively more about matching the state) and $\sigma_D = 0$ (so that in duopoly no one does research), but still the source prefers to share information with both the outlets. This happens because $\pi > .5$ and the source still cares about getting the information out quickly.

Case 2: v is high enough so that $\sigma_D < \sigma_M$. Now, the RHS of equation (A3.18) is greater than zero. But first, σ_D might not be too small so that in the LHS < 0 i.e. $2\sigma_M \leq \sigma_D(4-\sigma_D)$. In this case, sending to both is always preferred independent of μ. Therefore, sending to both is preferred if

$$\sigma_D < \sigma_M \leq \frac{\sigma_D(4-\sigma_D)}{2}.$$

Second, σ_D might in fact be very small so that on the LHS > 0 i.e. $2\sigma_M > \sigma_D(4-\sigma_D)$. In this case, sending to both is preferred only if

$$\mu \leq \frac{2\sigma_M - \sigma_D^2}{2\sigma_M - \sigma_D(4-\sigma_D)} \frac{1}{(1-\pi)}.$$

Case 3: v is small so that $\sigma_D > \sigma_M$. Again there are two possible situations. First, consider the case in which σ_D is not too large so that the RHS of equation (A3.18) is still positive, i.e. $2\sigma_M \geq \sigma_D^2 \implies \sigma_M \geq \frac{\sigma_D^2}{2}$. Now, in this case we want to see whether the LHS can be negative i.e. if $\sigma_M < \frac{\sigma_D(4-\sigma_D)}{2}$. But this must be true because $\sigma_D > \sigma_M$ and we know that $\frac{\sigma_D(4-\sigma_D)}{2} > \sigma_D$. Therefore, the LHS is negative and the RHS is positive, so the condition outlined in (A3.18) is satisfied. Sending to both is always preferred if

$$\frac{\sigma_D^2}{2} \leq \sigma_M < \sigma_D.$$

Second, σ_D might in fact be very large so that the RHS is negative, i.e. $\sigma_M < \frac{\sigma_D^2}{2}$. Now, it cannot be that the LHS is positive because that requires $\sigma_M > \frac{\sigma_D(4-\sigma_D)}{2}$ which contradicts $\sigma_M < \sigma_D$. Therefore, LHS must also be negative. From condition (A3.18), the source prefers both outlets only if

$$\mu \geq \frac{\sigma_D^2 - 2\sigma_M}{\sigma_D(4-\sigma_D) - 2\sigma_M}\frac{1}{(1-\pi)}.$$

Case 4: v is such that $\sigma_D = \sigma_M := \sigma$. When this is the case, the condition (A3.18) reduces to

$$-\mu(1-\pi)(2-\sigma) < (2-\sigma)$$

which is always true. Therefore, sending to both is preferred.

Our result follows from combining all the above cases. ■

Proof of Proposition 3.9.

The proof is by construction. We have already constructed the equilibrium frontier and the set of all possible equilibria for a given $\bar{c}$ and ε.

We now show what happens as $\varepsilon \to 0$. Consider σ_M first. From Proposition 3.2, observe that as $\varepsilon \to 0$ LHS$(\sigma = 0) = 0 \approx$ RHS$(\sigma = 0) = \frac{\varepsilon}{\bar{c}+\varepsilon} \to 0$ in equation (A3.3). Therefore, for any π the only fixed point equilibrium $\to 0$.

Now, consider σ_D at $v = 0$. Fix a π. We know that as $\varepsilon \to 0$, since $c_D(\sigma = 0) = \frac{1}{2}$, we have that RHS$(\sigma = 0) \to \frac{1}{2\bar{c}}$ in equation (A3.9). But this is strictly greater than LHS$(\sigma = 0) = 0$. Therefore, the equilibrium fixed point $\sigma_D > 0$ and also $\frac{\sigma_D^2}{2} > 0$. Moreover, this is true for any π.

Therefore, in the $\sigma_D - \sigma_M$ space as $\varepsilon \to 0$, the equilibrium frontier lies below the $\sigma_M = \frac{\sigma_D^2}{2}$ line.

Now, let us look at what happens as $\varepsilon \to \infty$. Given that the fixed point is defined by $\sigma^* = \frac{c^*+\varepsilon}{\bar{c}+\varepsilon}$, both σ_M and σ_D approach 1 (without ever being exactly equal to 1). However, because the frontier is defined for $v = 0$ case, the frontier lies close to and to the right of the $\sigma_M = \sigma_D$ line.

Combining the two observations above with Lemma 3.3, we get our proposition. ■

B3 Allowing for sitting on information

In this appendix, we show that allowing outlets to "sit on information" (i.e. just refrain from publishing until period 2 without acquiring the additional signal) does not preclude the equilibrium outlined in Proposition 3.4. We prove it formally for sufficiently low π and then use mathematical simulation to argue that it holds more generally. Uniqueness of such an equilibrium (among signal-based equilibria), however, is not obvious anymore. We make only one change with respect to the model described in Section 3.2. Now $d^i \in \{res, pub, wait\}$, where $d^i = wait$ means that the outlet does not acquire the second signal but still publishes in period 2.

This addition poses some challenges in the tractability of the model because the choice is no longer just between two options and strategies are not necessarily just thresholds in c. However, even in this more complicated setup we can show a few results. First, for sufficiently low π, it is possible to find values of v such that the equilibrium described in Proposition 3.4 exists; waiting is never a best response if the other player never waits and $\sigma_D^* > \sigma_M^*$. Second, we can simulate the model showing that we can assign values to v such that, for the resulting equilibrium σ_D^*, publishing in period 1 is better than waiting and at the same time $\sigma_D^* > \sigma_M^*$.

We begin with the following lemma considering that we are interested in the (candidate) equilibrium strategies described in Proposition 3.4 where $d^i = wait$ is never played in equilibrium.

Lemma B3.1 *It is always possible to find off path beliefs such that, for sufficiently high v, $Eu^i(d^i = wait) \leq Eu^i(d^i = pub)$.*

Proof of Lemma B3.1.

Note that $\gamma(W_{\text{II}}, .)$ is off path in the equilibrium we are considering. For any $\gamma(\varnothing)$ and $\gamma(1)$ as defined above, the expected utility from choosing $d^i = wait$ is

$$\frac{1}{2}\sigma^j \left(\frac{v}{2} + \pi\gamma(\varnothing) + (1-\pi)\gamma(1)\right) + \left(\frac{1}{2}(1-\sigma^j) + \frac{1}{2}\right)(\pi + (1-\pi)\gamma(W_{\text{II}}, .)) \quad \text{(B3.1)}$$

On the other hand, the expected utility from publishing immediately is given by

$$\frac{1}{2}\sigma^j\left(v+\gamma(1)\right)+\left(\frac{1}{2}(1-\sigma^j)+\frac{1}{2}\right)(\frac{v}{2}+\pi^2(\gamma(\varnothing)-\gamma(1))+\gamma(1)) \tag{B3.2}$$

Comparing (B3.1) and (B3.2) and solving for v, we find that $Eu^i(d^i = wait) \leq Eu^i(d^i = pub)$ when

$$v \geq \sigma^j\pi(\gamma(\varnothing)-\gamma(1))-(2-\sigma^j)\left[\pi^2\gamma(\varnothing)+(1-\pi^2)\gamma(1)-\pi-(1-\pi)\gamma(W_{\mathrm{II}},.)\right] \tag{B3.3}$$

Therefore, it is possible to find v and $\gamma(W_{\mathrm{II}},.)$ such that the above condition is satisfied.

■

This makes intuitive sense as a sufficiently high scoop value should always deter sitting on information. From now on, we set $\gamma(W_{\mathrm{II}},.) = 0$ and we define $\bar{v} := \sigma^j\pi(\gamma(\varnothing)-\gamma(1))-(2-\sigma^j)\left[\pi^2\gamma(\varnothing)+(1-\pi^2)\gamma(1)-\pi\right]$.

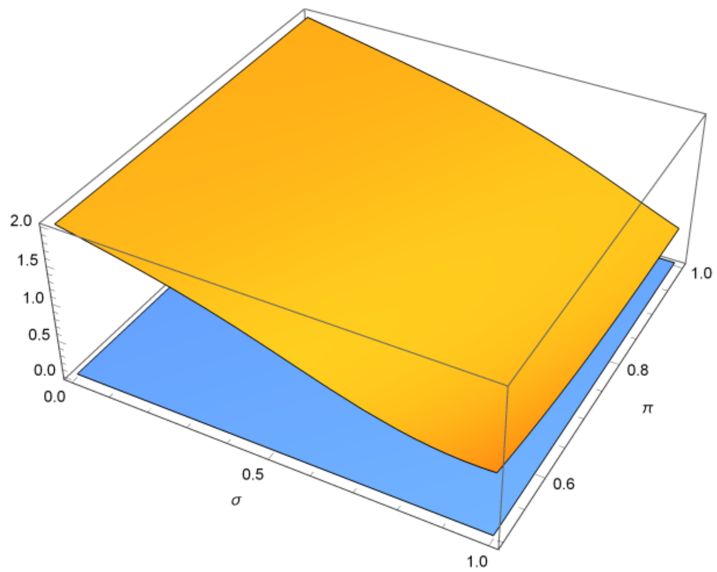
We can now move to the main proposition.

Proposition B3.1 *For sufficiently low π, it is possible to find values of v such that the equilibrium described in Proposition 3.4 exists. In such an equilibrium, waiting is never a best response if the other player follows the equilibrium strategies and $\sigma^*_D > \sigma^*_M$.*

Proof of Proposition B3.1.]

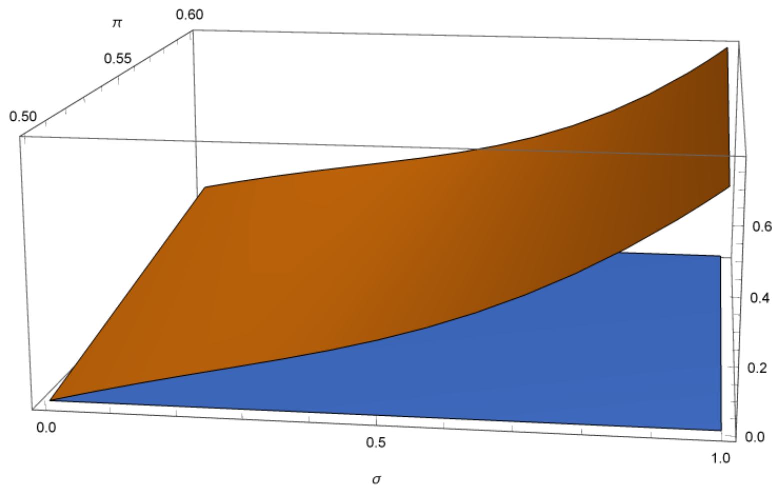
Suppose that player j always publishes when low type and chooses just between publishing or researching when high type. Moreover, suppose that the audience conjectures that both players use the equilibrium strategies described by Proposition 3.4. For this to be an equilibrium in the new setup, it is sufficient to prove that, given the correct audience's beliefs updating, for every σ, $Eu^i(d^i = wait) \leq Eu^i(d^i = pub)$. To show this, first we prove through Figure B3.1 that $\frac{\partial\bar{v}}{\partial\pi} > 0$. Moreover, Figure B3.2 shows that there exists a range of π such that $argmax_{\sigma}\bar{v}(\pi) = 1$. In the figure, it happens for $\pi \in [.5, .6]$. As a consequence, for every $v \geq \bar{v}(\sigma = 1, \pi \in [0.5, 0.6])$ it is true that, for every σ, $Eu^i(d^i = wait) \leq Eu^i(d^i = pub)$. In other words, if the audience conjectures an equilibrium where no types and no players choose to wait and the choice for the high type is just between publishing and researching described by a threshold strategy

Figure B3.1: Proof of Proposition B3.1: Proving $\frac{\partial \bar{v}}{\partial \pi} > 0$.



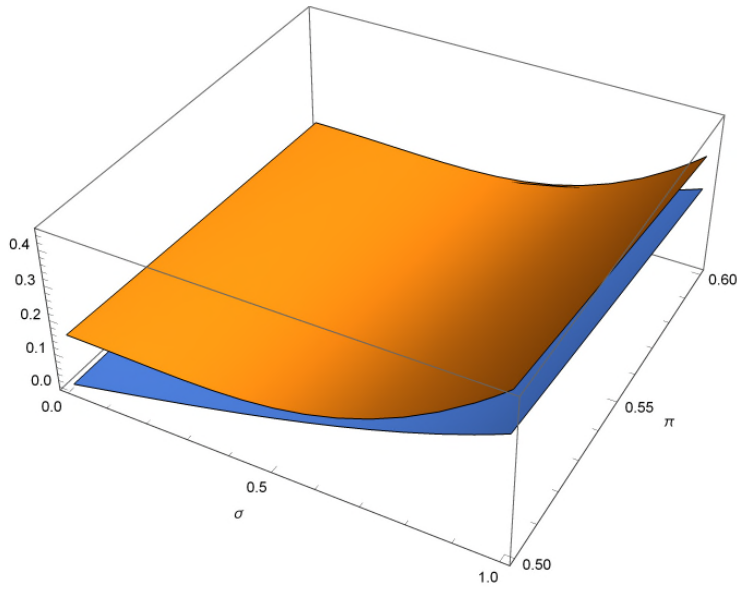
Proof of Proposition B3.1: Proving $\frac{\partial \bar{v}}{\partial \pi} > 0$. Orange plane: $\frac{\partial \bar{v}}{\partial \pi}$, blue plane: $0.\sigma + 0.\pi$ in the $\pi - \sigma$ space.

Figure B3.2: Proof of Proposition B3.1: Proving $argmax_{\sigma}\bar{v}(\pi) = 1$.



Proof of Proposition B3.1: Proving $argmax_{\sigma}\bar{v}(\pi) = 1$. Orange plane: $\bar{v}(\pi)$, blue plane: $0.\sigma + 0.\pi$ in the $\pi - \sigma$ space for $\pi \in [0.5, 0.6]$.

Figure B3.3: Proof of Proposition B3.1: Proving $c_D(\bar{v}(\sigma = 1)) > c_M$.



Proof of Proposition B3.1: Proving $c_D(\bar{v}(\sigma = 1)) > c_M$. Orange plane: $c_D(\bar{v}(\sigma = 1))$, blue plane: c_M in the $\pi - \sigma$ space for $\pi \in [0.5, 0.6]$.

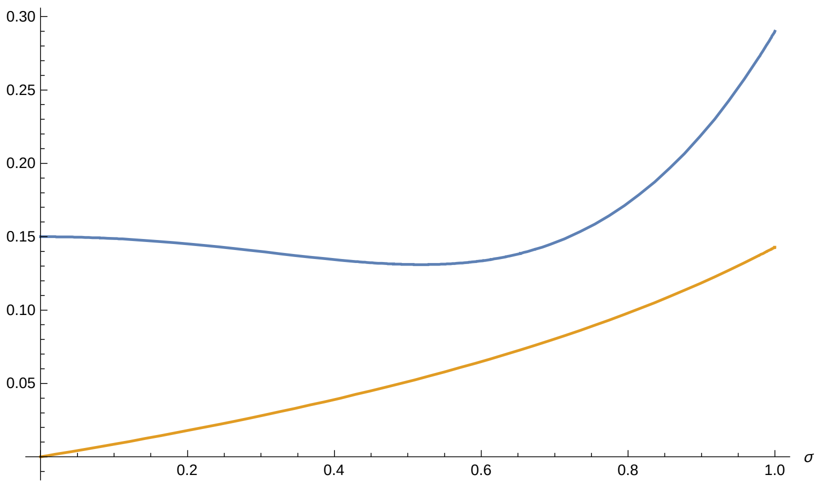
on c, behaving in this way is an equilibrium strategy for the outlets.

Finally, Figure B3.3 plots c_M and $c_D(\bar{v}(\sigma = 1))$ for sufficiently small π, proving that we can still increase v from $\bar{v}(\sigma)$ maintaining the necessary condition for $\sigma_D^* > \sigma_M^*$, i.e. $c_D \geq c_M$. ■

When $\pi > 0.6$, we can show the existence of our candidate equilibrium through mathematical simulations. Consider for example the following set of parameters: $\pi = 0.75$, $v = 0.7$, $\bar{c} = 2$, $\varepsilon = 0.1$. In this case, the equilibrium described in Proposition 3.4 (assuming it still exists) gives a solution $\sigma_D^* = 0.118219$.[17] Suppose now that player i expects player j to never wait and choose to research (if it is high type) with probability 0.118219. Further, suppose the audience think that both outlets never wait and research (if they are high types) with probability 0.118219. In this case, $Eu^i(d^i = wait) = 0.754218$ and $Eu^i(d^i = pub) = 0.841236$. Hence, there is no incentive to choose waiting instead of publishing, and the meaningful choice is just between

[17] We simulated the model with Mathematica. The code is available upon request.

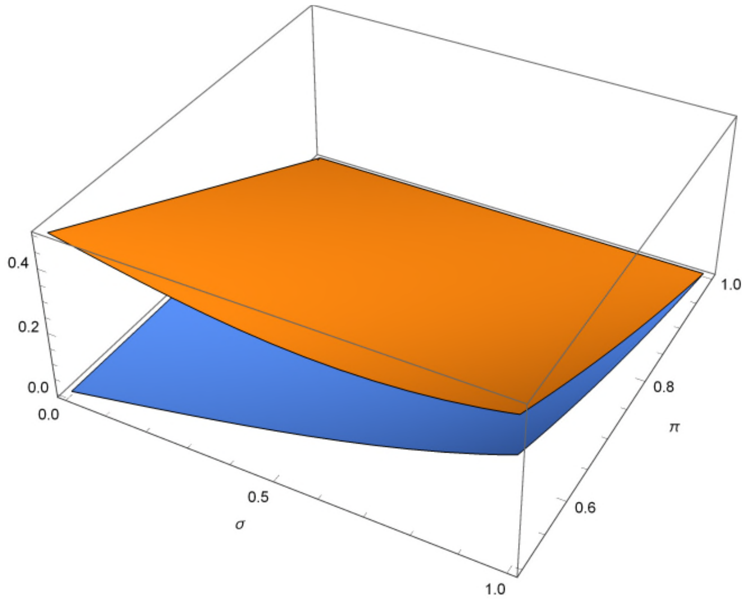
Figure B3.4: $c_D > c_M$ for $\pi = 0.75$ and $v = 0.7$.



$c_D > c_M$ for $\pi = 0.75$ and $v = 0.7$. Orange line: c_M, blue line: c_D as a function of σ)

researching and publishing. The solution to this problem is the same as that described by Proposition 3.4. Finally, Figure B3.4 shows that, for $\pi = 0.75$ and $v = 0.7$, it is still true that $c_D \geq c_M$ for every σ. More generally, Figure B3.5 plots c_M and c_D (in the $\pi - \sigma$ space) by replacing v with the corresponding $\bar{v}$. Still, c_D is above c_M throughout the entire range of parameters of our model.

Figure B3.5: $c_D(v = \bar{v} > c_M$ for every combination of σ and π.



$c_D(v = \bar{v}) > c_M$ for every combination of σ and π. Orange plane: $c_D(v = \bar{v})$, blue plane: c_M in the $\pi - \sigma$ space.

C3 Generic prior on the type

This appendix shows that our main results are qualitatively unaffected by the assumption of $Pr(\theta^i = h) = \frac{1}{2}$. In this section, we assume a generic $Pr(\theta^i = h) = q \in (0, 1)$, leaving the rest of the model unchanged. We consider monopoly, duopoly and their comparison for when θ is unknown to the reader.

Monopoly

The proposition of the main result is unchanged in monopoly, as q enters only in the readers' beliefs updating.

Proposition C3.1 *If there is one media outlet and θ is not known to the audience, there exists a unique equilibrium in which the high quality outlet conducts research in $t = 1$ if*

$$c \leq (1 - \pi)(\gamma(R) - \gamma(W)) := c_M$$

where $\gamma(R)$ and $\gamma(W)$ are the audiences' beliefs about the outlet's quality after it gets *the state right and wrong respectively. As a consequence,* $\sigma^* = F(c_M(q)) = \frac{c_M(q,\sigma^*)+\varepsilon}{\bar{c}+\varepsilon}$.

Proof of Proposition C3.1.

Suppose that a high type outlet chooses $d = res$ with probability σ. Reminding ourselves from the main text that by Bayes' rule,

$$\gamma(R) = \frac{q(\sigma + (1-\sigma)\pi)}{q(\sigma + (1-\sigma)\pi) + (1-q)\pi}$$

$$\gamma(W) = \frac{q(1-\sigma)(1-\pi)}{q(1-\sigma)(1-\pi) + (1-q)(1-\pi)} = \frac{q(1-\sigma)}{1-q\sigma}.$$

A high quality optimally chooses res if

$$\gamma(R) - c \geq \pi\gamma(R) + (1-\pi)\gamma(W) \implies c \leq (1-\pi)(\gamma(R) - \gamma(W)) := c_M(q)$$

In equilibrium the conjectured σ must be equal to the actual one, hence it must be that

$$\sigma^* = F(c_M(q, \sigma^*)). \tag{C3.1}$$

We need to check if such a fixed point exists. To do so, three observations are in order. First, note that both the LHS and RHS of the above are continuous in σ^*. Second, LHS$(\sigma^* = 0) = 0 <$ RHS$(\sigma^* = 0) = \frac{\varepsilon}{\bar{c}+\varepsilon}$ (as $c_M(q) = 0$ at $\sigma^* = 0$). Third, LHS$(\sigma^* = 1) = 1 >$ RHS$(\sigma^* = 1) = F\left(\frac{(1-\pi)q}{q+(1-q)\pi}\right)$, so the equilibrium is the solution of $\sigma^* = \frac{c_M(q,\sigma^*)+\varepsilon}{\bar{c}+\varepsilon}$ and LHS and RHS must cross at least once.

Finally, we need to check for the uniqueness of the fixed point. To show this, it is sufficient to prove that the derivative of the RHS with respect to σ is smaller than 1. Note that

$$\frac{\partial \text{RHS}}{\partial \sigma^*} = \frac{1-\pi}{\bar{c}+\varepsilon}\left[\frac{\pi(1-\pi)q(1-q)}{(q(\sigma + (1-\sigma)\pi) + (1-q)\pi)^2} + \frac{q(1-q)}{(1-q\sigma^*)^2}\right] > 0$$

Moreover, we can rewrite the equation as

$$\frac{\partial \text{RHS}}{\partial \sigma^*} = \frac{(1-\pi)q(1-q)}{\bar{c}+\varepsilon}\left[\frac{\pi(1-\pi)}{(q(\sigma + (1-\sigma)\pi) + (1-q)\pi)^2} + \frac{1}{(1-q\sigma^*)^2}\right]$$

It is easy to see that, in the range of parameters of the model, $(1-\pi)q(1-q) \leq \frac{1}{8}$; $\frac{\pi(1-\pi)}{(q(\sigma+(1-\sigma)\pi)+(1-q)\pi)^2} \leq 1$ because $\pi(1-\pi)$ is at most $\frac{1}{4}$ and $q(\sigma+(1-\sigma)\pi)+(1-q)\pi$ is at least $\frac{1}{2}$ (when $\sigma = 0$ and $\pi = \frac{1}{2}$); $\frac{1}{(1-q\sigma^*)^2} \leq 1$. As a consequence,

$$\frac{\partial \text{RHS}}{\partial \sigma^*} < \frac{1}{8}[1+1] < 1$$

and this completes the proof. ■

Duopoly

For the case of duopoly, we look directly at symmetric equilibria, showing that there exists a unique symmetric equilibrium.

Proposition C3.2 *If there are two media outlets and θ is not known to the audience, there exists a unique symmetric equilibrium where $\sigma^{i*} = \sigma^{j*} := \sigma^* = F(c_D(q))$ such that*

$$c_D(q) = \left[(\gamma(\varnothing) - \gamma(1))\left(q\sigma^* - (1-q\sigma^*)\pi^2\right) + 1 - q\right] - \frac{1}{2}v$$

where $\gamma(\varnothing) = \frac{q((\sigma^*)^2 q + (1-\sigma^*)(1-q\sigma^*)\pi^2)}{(q\sigma^*)^2 + (1-q\sigma^*)^2\pi^2}$ *and* $\gamma(1) = \frac{q(1-\sigma^*)}{1-q\sigma^*}$.

Proof of Proposition C3.2.

We focus directly on symmetric equilibria where each high type outlet uses a threshold strategy on c in the decision on whether to publish or investigate. Define σ as the probability (from the point of view of the other players) that a high quality outlet chooses to do research. For the same logic as in Proposition 3.4, the threshold is given by

$$c^i \leq \left[(\gamma(\varnothing) - \gamma(1))\left(q\sigma - (1-q\sigma)\pi^2\right) + (1-q\sigma)(1-\gamma(1))\right] - \frac{1}{2}v := c_D(q) \qquad (3.1)$$

where, by Bayes' rule, $\gamma(\varnothing) = \frac{q(\sigma^2 q + (1-\sigma)(1-q\sigma)\pi^2)}{q^2\sigma^2 + (1-q\sigma)^2\pi^2}$ and $\gamma(1) = \frac{q(1-\sigma)}{1-q\sigma}$.

Given that cost is uniformly distributed in $[-\varepsilon, \bar{c}]$ and that in equilibrium the conjectured probability of investigation must be equal to the actual one, the (symmetric) equilibrium level of σ, if it exists, must be the solution of

$$\sigma = F(c_D(q,\sigma)) \tag{C3.2}$$

where

$$F(c_D(q,\sigma)) = \begin{cases} 0 & c_D(q,\sigma) < -\varepsilon \\ \frac{c_D(q,\sigma)+\varepsilon}{\bar{c}+\varepsilon} & -\varepsilon \leq c_D(q,\sigma) \leq \bar{c} \\ 1 & c_D(q,\sigma) > \bar{c} \end{cases}$$

and

$$f(c_D(q,\sigma)) = \begin{cases} 0 & c_D(q,\sigma) < -\varepsilon \\ \frac{1}{\bar{c}+\varepsilon} & -\varepsilon \leq c_D(q,\sigma) \leq \bar{c} \\ 0 & c_D(q,\sigma) > \bar{c} \end{cases}$$

Note that:

1. The LHS of equation (C3.2) is linear, with slope equal to 1, starting at 0 and ending at 1;
2. RHS($\sigma = 0$) $\geq 0 =$LHS($\sigma = 0$);
3. RHS($\sigma = 1$) $< 1 =$LHS($\sigma = 1$);
4. Both LHS and RHS are continuous in σ.

Hence, they cross at least once and there is at least one solution to this fixed point problem.

To show uniqueness, we can rewrite $c_D(q)$ as

$$c_D = AE + 1 - q - \frac{1}{2}v$$

where $A := \gamma(\varnothing) - \gamma(1) = \frac{q^2(1-q)}{(1-q\sigma)[q^2+\pi^2 B^2]}$, $B := \frac{1-q\sigma}{\sigma}$ and $E := q\sigma - (1-q\sigma)\pi^2$. It is easy to see that $\frac{\partial E}{\partial \sigma} \geq 0$. Moreover, it is also true that $\frac{\partial A}{\partial \sigma} \geq 0$. To see this, note that

$$\frac{\partial A}{\partial \sigma} = \frac{-q^2(1-q)\left[-q(q^2+\pi^2 B^2) + 2\pi^2 B \frac{\partial B}{\partial \sigma}(1-\sigma q)\right]}{((1-q\sigma)\left[q^2+\pi^2 B^2\right])^2} \geq 0$$

because $\frac{\partial B}{\partial \sigma} \leq 0$. However, the sign of E is ambiguous, with $E < 0$ for $\sigma < \frac{\pi^2}{q(1+\pi^2)} :=$ σ^T. We claim that the following two conditions are sufficient for uniqueness:

1. $\frac{\partial c_D(q)}{\partial \sigma} \leq 1$ for $\sigma \leq \sigma^T$;

2. $\frac{\partial^2 c_D(q)}{\partial \sigma^2} \geq 0$ for $\sigma \geq \sigma^T$;

The argument is as follows: as RHS($\sigma = 0$) ≥ 0 =LHS($\sigma = 0$) and RHS($\sigma = 1$) $<$ 1 =LHS($\sigma = 1$), the fixed point is:

1. Only at $\sigma = 0$, as RHS($\sigma = 0$) =RHS($\sigma = \sigma^T$) and below that in between. Moreover, there cannot be any additional crossing point above σ^T because the RHS would be coming from below, and, as it is convex, it cannot be that they cross and RHS($\sigma = 1$) < 1 =LHS($\sigma = 1$).

2. If the solution is not at 0, the first time they cross it must be that the LHS comes from below. There are two sub-cases:

   - If the first crossing point is in $\sigma \leq \sigma^T$, then there cannot be others in the same interval as $\frac{\partial c_D(q)}{\partial \sigma} \leq 1$. Moreover, there cannot be any other crossing point above σ^T because the RHS would be coming from below, and, as it is convex, it cannot be that they cross and RHS($\sigma = 1$) < 1 =LHS($\sigma = 1$).

   - If the first crossing point is above σ^T, it must be unique as a second solution would violate RHS($\sigma = 1$) < 1 =LHS($\sigma = 1$).

We now prove that the two sufficient conditions outlined above apply to our model.

First, a sufficient condition for $\frac{\partial c_D(q)}{\partial \sigma} \leq 1$ for $\sigma \leq \sigma^T$ is $\frac{\partial E}{\partial \sigma} A \leq 1$. This implies $(1 + \pi^2)q^3(1 - q) \leq (1 - q\sigma)(q^2 + \pi^2 B^2)$. As the RHS is decreasing in σ, this condition must hold for the highest possible σ, i.e. for $\sigma = \sigma^T$. Substituting and simplifying, this requires $q(1 - q) \leq \frac{1}{\pi^2(1+\pi^2)}$. The LHS is at most $\frac{1}{4}$ while the RHS is at least $\frac{1}{2}$, hence the condition is always satisfied.

Second, a sufficient condition for convexity of $c_D(q)$ for $\sigma \geq \sigma^T$ is $\frac{\partial^2 A}{\partial \sigma^2} \geq 0$. To show that it is always the case, note that

$$\frac{\partial^2 A}{\partial \sigma^2} = -q^2(1-q)\frac{\frac{\partial^2 D}{\partial \sigma^2}D^2 - 2D\frac{\partial D}{\partial \sigma}^2}{D^4} \tag{C3.3}$$

where $D = (1-q\sigma)\left[q^2 + \pi^2 B^2\right]$, $\frac{\partial D}{\partial \sigma} = -q(q^2+\pi^2B^2) + \pi^2 2B\frac{\partial B}{\partial \sigma}(1-\sigma q) < 0$ and $\frac{\partial^2 D}{\partial \sigma^2} = -q\pi^2 2B\frac{\partial B}{\partial \sigma} + 2\pi^2\left[\left(\frac{\partial B}{\partial \sigma}^2 + \frac{\partial^2 B}{\partial \sigma^2}B\right)(1-\sigma q) - qB\frac{\partial B}{\partial \sigma}\right] > 0$. A sufficient condition for (C3.3) to be positive is $2\frac{\partial D}{\partial \sigma}^2 \geq \frac{\partial^2 D}{\partial \sigma^2}D$.

By substitution, this implies

$$2\left[-q(q^2\sigma^2 + \pi^2(1-q\sigma)^2)\frac{1}{\sigma^2} - 2\pi^2\frac{(1-q\sigma)^2}{\sigma^3}\right]^2 \geq \tag{C3.4}$$
$$\left[q\pi^2 2\sigma\frac{(1-q\sigma)}{\sigma^4} + 2\pi^2(1-q\sigma)\frac{1}{\sigma^4} + \frac{4(1-q\sigma)^2}{\sigma^4}\pi^2 + 2q\pi^2\sigma\frac{(1-q\sigma)}{\sigma^4}\right](1-q\sigma)(q^2+\pi^2 B)$$
$$\sigma^2\left[-q(q^2\sigma^2 + \pi^2(1-q\sigma)^2) - 2\pi^2\frac{(1-q\sigma)^2}{\sigma}\right]^2 \geq$$
$$3\pi^2(1-q\sigma)^2(q^2\sigma^2 + \pi^2(1-q\sigma)^2)$$
$$\sigma^2 q^2(q^2\sigma^2+\pi^2(1-q\sigma)^2)^2 + 4\pi^4(1-q\sigma)^4 + 4\pi^2(1-q\sigma)^2 q\sigma(q^2\sigma^2+\pi^2(1-q\sigma)^2) \geq$$
$$3\pi^2(q^2\sigma^2+\pi^2(1-q\sigma)^2)(1-q\sigma)^2$$

where the second line follows by multiplication of both sides by σ^4 and the third by dividing both sides by 2 and working out explicitly the square on the LHS. Note that σ and q always appear together in the last line of (C3.4). As a consequence, we can redefine $\sigma q := x$ and check whether the condition holds for $x \in [0,1]$ and $\pi \in [0.5, 1]$. We prove this graphically using figure C3.1. It plots the difference between LHS and RHS of (C3.4) for the whole range of possible values of x and π, showing that this difference is always positive. This completes the proof. ■

Monopoly-Duopoly comparison

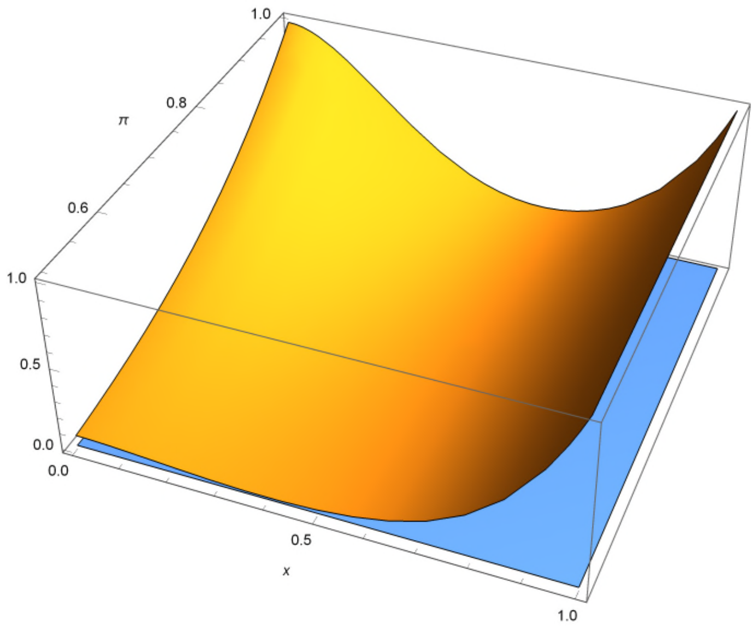
Finally, we show that sufficient conditions for competition leading to more research than monopoly can be found in this set up as well.

Proposition C3.3 *There exists a nonempty interval of v values where $\sigma_D^*(q) > \sigma_M^*(q)$.*

Proof of Proposition C3.3.

A sufficient condition for the proposition to hold is that, for some values of v, $c_D(q) >$

Figure C3.1: Proof of Proposition C3.2: Proving LHS > RHS in (C3.4).



Proof of Proposition C3.2: Proving LHS > RHS in (C3.4). Orange plane: LHS−RHS, blue plane: $0 * x + 0 * \pi$ in the $\pi - x$ space.

$c_M(q)$. Setting $v = 0$, and defining $B = \frac{1-q\sigma}{\sigma}$ note that:

$$\begin{aligned} c_D(q) &= (\gamma(\emptyset) - \gamma(1))(q\sigma - (1-q\sigma)\pi^2) + 1 - q \qquad \text{(C3.5)} \\ &= q\sigma \frac{q^2(1-q)}{(1-q\sigma)(q^2+\pi^2B^2)} - \pi^2(1-q\sigma)\frac{q^2(1-q)}{(1-q\sigma)(q^2+\pi^2B^2)} + 1 - q \\ &= (1-q)\left(\frac{q^3\sigma}{(1-q\sigma)(q^2+\pi^2B^2)} - \frac{\pi^2q^2}{q^2+\pi^2B^2} + 1\right) \\ &= (1-q)\left(\frac{q^2+(1-q\sigma)\pi^2B^2}{(1-q\sigma)(q^2+\pi^2B^2)} - \frac{\pi^2q^2}{q^2+\pi^2B^2}\right) \\ &= \frac{1-q}{1-q\sigma}\left(\frac{q^2+(1-q\sigma)\pi^2B^2}{(q^2+\pi^2B^2)} - \frac{\pi^2q^2(1-q\sigma)}{q^2+\pi^2B^2}\right) \end{aligned}$$

where the first equality follows by substitution and the rest is a series of rearrangements.

Note that, as $q \in (0,1)$, neither $1-q$ nor $1-q\sigma$ are ever 0. Similarly, by substitution,

$$\begin{aligned} c_M(q) &= (1-\pi)\left(\frac{q(\sigma+(1-\sigma)\pi}{q\sigma+q(1-\sigma)\pi+(1-q)\pi)} - \frac{q(1-\sigma)}{1-q\sigma}\right) \\ &= (1-\pi)q\left(\frac{\sigma+(1-\sigma)\pi}{q\sigma(1-\pi)+\pi} - \frac{1-\sigma}{1-q\sigma}\right) \\ &= \frac{(1-\pi)q\sigma(1-q)}{(1-q\sigma)(q\sigma(1-\pi)+\pi)} \end{aligned} \tag{C3.6}$$

As a consequence, by comparison of (C3.5) and (C3.6), $c_D(q) > c_M(q)$ implies

$$\frac{q^2+(1-q\sigma)\pi^2(B^2-q^2)}{(q^2+\pi^2B^2)} > \frac{(1-\pi)q\sigma}{(q\sigma(1-\pi)+\pi)} \tag{C3.7}$$

Note that both LHS and RHS of (C3.7) are decreasing in π. The case of RHS is straightforward. For the LHS, a sufficient condition is

$$(1-\sigma q)2\pi(B^2-q^2)(q^2+\pi^2B^2-2\pi B^2(q^2+(1-q\sigma)\pi^2(B^2-q^2))<0$$

This simplifies to $-\sigma q 2\pi q^2 B^2 - (1-\sigma q)2\pi q^4$ that is always negative.

As a consequence, a sufficient condition for $c_D(q) > c_M(q)$ is LHS($\pi = 1$) >RHS($\pi = 0.5$). By substitution, this implies

$$\frac{q^2+(1-q\sigma)(B^2-q^2)}{(q^2+B^2)} > \frac{q\sigma}{1+q\sigma}$$

After few simplifications and substituting back the value of B, we obtain

$$\sigma^2 q^2 \frac{2q\sigma-1}{\sigma^2} + \frac{(1-q\sigma)^3}{\sigma^2} > 0$$

A sufficient condition for this to hold is

$$1-3q\sigma+2q^2\sigma^2+q^3\sigma^3>0$$

Noticing that $q\sigma$ is bounded between 0 and 1, the condition is always satisfied and this completes the proof. ■

D3 Monopoly with public signal

In this appendix we assume that the audience learns the actual timing with positive probability z in monopoly. This helps us establish that additional learning in our benchmark duopoly model happens not only because the timing is revealed with some probability but also because the audience uses additional information from outlets matching the state. In this set up, the outlet does not know whether the audience has learned the timing or not when taking its decision. Then, the condition for doing research is

$$z\gamma_M(2) + (1-z)\gamma_M(\varnothing) - c \geq z\gamma_M(1) + (1-z)(\pi\gamma_M(\varnothing) + (1-\pi)\gamma_M(1)) \tag{D3.1}$$

Note that $\gamma_M(2) = \gamma(2) = 1$ and $\gamma_M(1) = \gamma(1) = \frac{1-\sigma}{2-\sigma}$. However,

$$\gamma_M(\varnothing) = \frac{\sigma + (1-\sigma)\pi}{\sigma + (1-\sigma)\pi + \pi} \neq \gamma(\varnothing) = \frac{\sigma^2 + (1-\sigma)(2-\sigma)\pi^2}{\sigma^2 + (2-\sigma)^2\pi^2}$$

because in duopoly the audience can learn also from the other player getting the state wrong. Hence, it is confused only if both outlets publish simultaneously and they both get the state right.

For comparison, we can write the duopoly condition for $v = 0$ a bit differently. Define χ the probability that the opponent behaves in a way that reveals the timing to the reader. Note that χ is "artificial" because it is the probability that j does not research when player i does (i.e. $\frac{1}{2}(1-\sigma^j) + \frac{1}{2}$ on the LHS) and vice-versa (i.e. $\frac{1}{2}\sigma$ on the RHS). In such cases, the action of player j is fully revealing of the timing, irrespective of the endorsement. The duopoly condition for research is then

$$\chi\gamma(2) + (1-\chi)\gamma(\varnothing) - c \geq \chi\gamma(1) + (1-\chi)(\pi^2\gamma(\varnothing) + (1-\pi^2)\gamma(1)) \tag{D3.2}$$

Comparing (D3.1) and (D3.2) reveals that they are similar, but not identical. Even if we set $z = \chi$, the difference in $\gamma(\varnothing)$ and in the π^2 term of the RHS is still there. Hence, our result is not just due to the fact that the publication timing of the opponent reveals information about the timing of the other player. The content of the endorsements plays a role as well.

Chapter 4

Ora et guberna. The economic impact of Benedict's rule in medieval England

4.1 Introduction

The claim that "institutions matter" has become overwhelmingly popular in economics in the last few decades, supported by an increasing amount of empirical evidence tying good institutions to better economic performance (e.g. Glaeser et al. 2004, Acemoglu et al. 2019). In this paper we exploit the historical case of Benedictine monasteries in medieval England, at the turn of the Norman conquest, to study the economic effect of institutional differences in societies where only few of the elements that now constitute a democracy are in place.

English Benedictine monasteries were part of the medieval feudal organisation of the society, playing an important role as landlords at a time when ownership and political power were overlapping (Aston 1958). Quite often, monasteries were not only controlling the land where they stood, but also other holdings located in different places. Moreover, this role as big landlords, alongside secular noblemen and bishops, remained salient after the Norman conquest (Finn 1963, Knowles 1966). At the same time, their institutional structure was determined by Benedict's rule, prescribing, for example, the election of their leader. Were monasteries governing their land more or less efficiently than secular landlords? And, if so, for what reason?

In this paper we address these questions exploiting a unique historical setting in which different institutional features can be compared at a very disaggregated level: we focus on the biggest land reshuffling in British history to identify the economic effect of being controlled by a Benedictine monastery (*vis-à-vis* secular landlords). We consistently find a positive effect of Benedictine control on productive capacity.

Exploring different potential channels, we argue that at least part of this effect is due to an important difference in the institutional structure: the Benedictine's ability to elect their leader.

Benedictine monasteries are an interesting case study because they had a very well defined institutional structure: they were abiding to the "Rule" established by Benedict of Nursia[1] in 529 AD. This vast and comprehensive document was (and still is) supposed to regulate monastic life, including its governance. And it did so in a way that scholars defined "monastic democracy" (Judson 1898, Moulin 2016): the head of a monastery (called abbot) was an elective office, elected for life by the community of monks and he had an obligation to consult with the community on important decisions. However, the ultimate decision power was in his own hands only. This was happening at a time where secular landlords had complete discretionary power on their own land. As a consequence, we are comparing economic outcomes of very small and geographically close political units whose governance changes depending on whether the landlord is a secular noblemen, appointed (or heir of a previous landlord) and with substantially absolute power on the land, or a Benedictine monastery, with an elected leader.

Different contributions highlight a higher (Postan 1973, Ekelund et al. 1996, Rost 2017) or a lower (Heldring et al. 2017) level of efficiency in monastic management of lands, with various potential channels. To identify precisely whether a causal effect of Benedictine control can be found, we exploit the unique historical context of the Norman conquest and the change in land ownership that followed, focusing only on holdings that were assigned to monasteries (or to new secular landlords) by the King after the Conquest.

[1]Nursia, *Norcia* in Italian, is a small village in central Italy where Benedict was born around 480 AD.

Our dataset merges information from the Domesday Book,[2] the English Monastic Archive and a variety of other sources for geographic and historical controls. All these data are at holding level, i.e. the smallest available unit.[3] As we have information on the economic performance of these holdings (in terms of productive capacity) and on the landlord name, we can compare the productivity of land controlled by Benedictine monasteries with that of land controlled by secular landlords.

A first, rough comparison can be obtained in a cross sectional regression, where we use as outcome variable the growth rate between 1066 and 1086, i.e. the two data points available in the Domesday Book. A dummy for Benedictine-controlled holdings is the main explanatory variable. We can control for a rich set of potential confounding factors, ranging from the historical importance of the location (its distance to Roman roads and Roman settlements), its geography (latitude, longitude, ruggedness, terrain suitability to pasture and agriculture etc.), its access to marketplaces (distance to markets, London etc.), to other specific characteristics of the individual holding (size, number and dispersion of holdings controlled by the same landlord, productive capacity in 1066). Moreover, we include county or hundred fixed effect, hence effectively comparing Benedictine and non-Benedictine holdings within small and homogeneous geographic units.

As we are worried about selection (i.e. monasteries being able to pick better areas *ex ante*) and about the differential effect of the Norman conquest, we exploit the panel structure of our dataset and the "natural experiment" induced by William's replacement of the Anglo-Saxon elite, causing the re-allocation of most of the holdings. In particular, we focus our attention on the subsample of holdings that were controlled by Anglo-Saxon secular landlords in 1066 and changed owner in 1086. Our "treatment" group is made of holdings controlled by Benedictine monasteries in 1086, and the "control" group is made of all the others, that ended up in the hands of Norman noblemen or bishops. Applying a difference-in-difference approach, we find that Benedictine holdings experienced a 20 years productive capacity growth rate at least 4 percentage points higher than holdings controlled by secular landlords.

[2] An extensive survey of productivity and land allocation of its reign, commissioned by William in 1086. More in Section 4.2.

[3] Most of these holdings are manors, i.e. the smallest administrative unit and the basic unit of inquest of the Domesday Book (Finn 1963).

The effect we identify may be due to several channels, but we argue that the institutional structure played an important role. To support our claim, we provide several tests to take into account and rule out alternative potential explanations. In particular, we can show that Benedictine holdings are not superior in terms of population measured in 1086 or in terms of technology-related variables, such as the number of ploughs and of mills in 1086. Moreover, we show that the distance from a generic Benedictine monastery has no effect on productive capacity, suggesting that, at least in the short-run perspective we are considering, our effect is not due to cultural transmission. Also, it does not seem to be a matter of economies of scale, as we can control for both the size of the holding and the number and dispersion of holdings controlled by the same landlord, nor an effect due to the historical roots of the places. Moreover, we are able to rule out the hypothesis that our effect is due to the comparison of religious with secular institutions. In particular, we show that holdings controlled by Anglo-Saxon nobility in 1066 and passed to bishops in 1086 are not statistically different from those controlled by Anglo-Saxons and then assigned to Norman noblemen in the same period. This is an interesting result because bishops are religious institutions sharing the same cultural and religious roots of Benedictine monasteries but that were (and are) not constrained by Benedict's Rule. On top of that, we show that Benedictine-controlled holdings over-perform Anglo-Saxon controlled holdings as well, so our effect is not due to their better knowledge of the land when compared with the newly arrived Normans. Further, more ancient monasteries are not better rulers than more recent ones, reducing the concern that our effect is due to better/longer investments in human capital. Finally, the effect is substantially unchanged when we exclude holdings that are closer than 5km to a monastery, thus ruling out that it is due to the monks actively working the land.

Overall, our contribution is twofold. First, we are able to identify the short-term effect on productivity of being governed by a Benedictine monastery, in medieval England. We show that this effect is positive and relevant, as it is able to switch the 20-years growth rate of holdings' productive capacity from negative to positive. Secondly, we provide a more general contribution on the effect of institutional arrangements on economic outcomes.

On top of that, we think our results can shred some light on the role of (relatively) democratic selection in contexts where other tools to keep the leader accountable are very weak, hence isolating as much as possible the specific role of democratic selection from other elements (absent at that time) of the democratic process. This is related with the literature on democracy and economic outcomes[4], but we take into account that democracy is a multidimensional concept, composed by several elements. It involves selection of officials, tools that keep them accountable, the rule of law *et cetera*.[5] Empirically, it is very hard to disentangle the elements that constitute a democracy, and to explore their effect separately. Usually, the comparison is between appointed and elected officials (Martinez-Bravo et al. 2014, Burgess et al. 2015, Fetzer and Kyburz 2018), but appointed officials are different from elected ones because they are non-selected (by the voters) and also non-accountable (again with respect to the voters), and it is usually very hard to separate the two elements.[6] In this paper, we compare institutional structures that are fairly similar, being both characterised by very weak accountability mechanisms (i.e. leaders in charge for life with the ultimate decision power in their own hands), with the important difference that abbots are elected and secular landlords are not. Our results suggest that the possibility of selecting a leader, even when accountability is weak, may result in better economic performance.

Related literature Two papers are particularly close to this one: Heldring et al. (2017) and Andersen et al. (2017).

Heldring et al. (2017) looks at the impact of monastic dissolution on the industrial revolution in England. They find that pre-dissolution monastic income is positively correlated with agricultural yields, patenting and the number of gentry in a parish, and negatively with the share of labour force in agriculture. Similarly with the present

[4]Most of the recent empirical literature finds a positive, causal effect of democracy on growth (e.g. Acemoglu et al. 2005 and Acemoglu et al. 2019), with some noteworthy exceptions (Gerring et al. 2005). Moreover, at the local level, democracy is associated with reduction in violence (Fetzer and Kyburz 2018), better long term outcomes (Angelucci et al. 2017) and more efficient public good provision (Acemoglu et al. 2014, Martinez-Bravo et al. 2014, Martinez-Bravo et al. 2017).

[5]The interplay of those elements is not trivial: for example, it has been shown theoretically that there can be a trade-off between accountability and selection (Ashworth et al. 2017). Moreover, unaccountable (non-selected) officials may do better than accountable (non-selected) ones (Maskin and Tirole 2004, Alesina and Tabellini 2007) and a more inclusive decision making structure does not necessarily lead to better decisions (Persico 2004).

[6]Term-limited elected officials have weaker re-election concerns than non term-limited. However, they have more experience than non-selected leaders and they could care about their future career. In our case, we are able to compare selected and non-selected leaders that are both in charge for life.

work, they control for geographic specific factors and hundred fixed effects. However, our research question is different, as we look at the contemporaneous effect of different institutions on economic outcomes. Moreover, they look at a different period and we are able to identify a historical event that changed the institutional structure of land holding in a quasi-experimental fashion.

Andersen et al. (2017) use county-level English data to show the long-run impact of Cistercian monasteries on cultural values and population growth. We use variation at a much finer level (our unit of analysis is the individual holding and we control for hundred fixed effects or even for holding fixed effect) to test the contemporaneous impact of different institutional structures, looking at an earlier historical period. Similar to us, they do not find any effect of Benedictine presence on population growth, but we show that there is a strong and positive effect on productive capacity, visible in the short term (less than 20 years). As the "work-ethic factor" takes time to develop, our results suggest that other mechanisms were at play, such as a different institutional decision making process.

We are also not the first one comparing outcomes of secular and ecclesiastical holdings, but other studies focus on much smaller and less representative samples (McDonald and Snooks 1986) and they usually do not have an identification strategy. Results are very mixed. Heldring et al. (2017) note that "the most obvious explanation for any impact of the Dissolution is that the land held by the Church was initially utilised inefficiently", pointing toward a lower productivity of ecclesiastical holdings. Campbell (1983) finds no significant differences between ecclesiastic and lay holdings, and Campbell (2006) points out that there are two competing effects, going on. On the one hand, "on conventual and collegiate estates inertia rather than enterprise could all too easily rule" (p. 421). On the other hand, "such landlords were also in a uniquely privileged position to develop the management of their estates on a long-term sustainable basis" (p. 421). Ekelund et al. (1996) stresses the superior efficiency of the Cistercian order due to their cheap labour and the use of *conversi*,[7] but they still highlight an interesting trade off between this fact and the efficiency losses related with team production and difficult monitoring.

[7]Note that the period we are considering precedes the creation of the Cistercians and the introduction of *conversi* within monastic communities.

To the best of our knowledge, we are the first one to use the entire Domesday Book for our analysis, and to suggest an identification strategy that controls for geographic proximity and economic determinants, together with the quasi experimental set up given by the post-Conquest re-assignment. Angelucci et al. (2017) look at the long-term effect of farm grants (i.e. basically self-governing municipalities) on political support for more inclusive institutions. They also use data from the Domesday Book, but they focus only on boroughs, while we look at countryside holdings. Moreover, Norman conquest is the starting point of their long-term analysis, while we look at short term impact before and after that event.

Few papers (Acemoglu et al. 2014, Martinez-Bravo et al. 2014, Burgess et al. 2015, Fetzer and Kyburz 2018) look at local variations in the degree of "democracy", but in a fairly different way. In particular, they compare elected and appointed officials.

4.2 Historical background

4.2.1 "Monastic democracy": Benedictine monasteries and the Rule

The Benedictine order was (and still is) composed by the set of monasteries committed to follow the Rule wrote by Benedict of Nursia in the Early middle age (Knowles 1966). They quickly became the most important monastic order in Europe, at least until the Cistercian reform of the XII century. They arrived in England in 597 AD, when they built a monastery in Canterbury. English Benedictine monasteries grew rapidly, as shown in Figure 4.1, and acquired control over a number of holdings: partially because of endowments and partially because of direct assignment of the Anglo-Saxon kings.[8] Importantly, the management of those holdings was highly centralised and kept in the hand of the monastic community until the XII c. (Knowles 1966).

Each individual monastery was an autonomous entity (there was not a "head of the order") run by an abbot: therefore, the common membership to the same order was granted by the adherence to the same rule. As Knowles (1966, p. 101), notes, by the time of the Conquest monastic houses were substantially independent to each other, with nothing like a federation or a formal interdependence into place. Further, the

[8] Aston (1958), Ayton and Davis (1987).

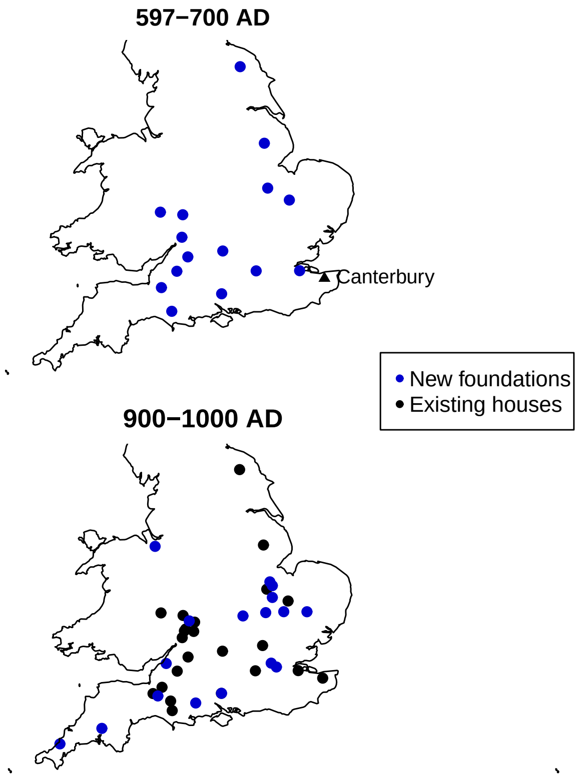


Figure 4.1: Diffusion of Benedictine monasteries in medieval England

internal organisation of a Benedictine monastery became explicitly independent to the interference of the local bishop after the council of Hertford launched by Archbishop Theodore in 672 AD (Lapidge 1995, Dell'Omo 2011). Independence from the King was granted usually by papal provisions.[9]

Even though the main purpose of the Rule was ascetic, it contains detailed and comprehensive instructions for the community of monks (prayer times, kitchen duties etc). Its importance for the efficiency of monastic life has been noted by scholars in management science (Rost and Graetzer 2013, Ehrmann et al. 2013, Rost 2017). Inevitably, some of those rules where dedicated to the way the community was governed. In particular, Chapter 64 explains that the abbot is an elective office (elected for life by the monks) and that majority rule can be used if there is no unanimity in the community. In particular, it states:

> *At the election of an abbot let this principle be always observed, that he be appointed whom the whole community, being of the same mind and in the*

[9] "*Epistulae privilegii*", as witnessed by some records in Beda's *Historia Ecclesiastica Gentis Anglorum* (Dell'Omo 2011).

fear of God, or even a part albeit a small part of the community shall with calmer deliberation have elected.[10]

On top of this, Chapter 3 states that the abbot had an obligation to routinely consult with the senior members of the community.

And if any less important business has to be transacted on behalf of the monastery, let counsel be taken, but with the seniors only, as it is written: 'Do everything with counsel and having so done thou wilt not repent.'[11]

Finally, the whole community, including its most junior members, has to be consulted for important decisions.

As often as any special business has to be transacted in the monastery, let the abbot convoke the whole community [...] And we have thus said that all are to be called to council because it is often to a junior that the Lord reveals what is best.[12]

However, it is important to stress that this does not mean that the community can act as a sort of "parliament". Chapter 3 clarifies that the decision making power lies in the hands of the abbot only:

And having listened to the counsel of the brethren, let him settle the matter in his own mind and do what seems to him most expedient.[13]

On top of this, Chapter 4 clarifies that the monks should "yield obedience in all things to the abbot's precepts, even if he himself act contrary to their spirit".

Obviously, peasants were not allowed to vote or to participate into the political decision process and the accountability of the abbot to the community was limited, as he was expected to be in charge for the rest of his life. However, the set of rules mentioned above seems quite different with respect to the standard feudal institutional arrangements, where a landlord has absolute power on his land, he is not elected and has no

[10]Benedict's Rule, Ch. 64.
[11]Benedict's Rule, Ch. 3.
[12]Benedict's Rule, Ch. 3.
[13]Benedict's Rule, Ch. 3.

obligation to consult with anyone in his decision. Interestingly, Moulin (2016) finds that the appropriate words to summarise this set of rules is “monastic democracy”.

4.2.2 Land ownership in medieval England

The manor is the basic unit of analysis of the Domesday Book (Finn 1963), and it was also the basic unit of the feudal structure[14] of the society, whose importance can be traced back to the Anglo-Saxon era (Aston 1958, McDonald and Snooks 1986, Roffe 2000, Roffe 2007).[15] Landlords had absolute control over their land, granted by the King, or by the Tenant in Chief (Aston 1958). Kings granted land to churchmen or monasteries directly, while other were received from secular landlords as endowments or donations (Ayton and Davis 1987).

Landlords, usually, rented out part of the land. The rest, usually around 30% of the total (Kosminsky 1961), called demesne, was kept under the direct control of the landlord. Peasants had to spend a certain number of days per year working on that land, in a “security for labor” contract that attracted attention from economists (such as North and Thomas 1971, Jones 1972, North and Thomas 1973). As a consequence, the landlord was deriving income from two sources: the output of production from the demesne and the rents he was receiving for the rest of the land (Postan 1973).

According to Kosminsky (1961), manors varied greatly in size and organisation, some of them having no villages at all and other having very big demesnes. Overall, Church manors seemed to be bigger, but not very different from manors of comparable size controlled by secular landlords (Kosminsky 1961, p. 103-113). Postan (1973) points out that Benedictine manors seem to mirror more closely the traditional structure of manors. Swanson (1979) highlights the similarity in terms of manorialism between the holdings of the Church and those of contemporary secular landlords.

While the manor was essentially a unit based on ownership of land, the country was also organised into “administrative” units used by the monarchy to implement royal

[14] Among historians the nature of the pre-Conquest structure of English society has been debated. Nonetheless, as Roffe (2007, p. 150) states: “it is probably true to say that there is now a universal appreciation of a significant role for forms approximating to ‘feudal’ lordship in pre-Conquest society.”

[15] The vast majority of the observations in our sample are manors, or other pieces of land associated with a manor (e.g. a dependency). There are also some “non manorial units” (Palmer 1987) that albeit being not legally part of a manor, they where still constituting holdings with a landlord in charge of them.

laws and to collect taxes, namely *Shires*, renamed *Counties* after the Conquest, and, at a more local level, *Wapentakes*, or *Hundred*.[16]

4.2.3 The Norman Conquest and the Domesday Book

In 1066 William, the duke of Normandy, invaded England and conquered it after the Battle of Hastings. He claimed he was the legitimate heir of the Anglo-Saxon King Edward[17], who was his first cousin once removed. Upon Edward's death, however, the kingdom passed to his brother-in-law Harold, triggering as a consequence the Norman invasion and the end of the Anglo-Saxon era.

The advent of William was not good news for Benedictine monasteries: generally speaking, they were historically well aligned with the Anglo-Saxon monarchy (Knowles 1966, Barlow 1979) and "the initial effect of the Norman conquest was greatly to disturb the monasteries" (Barlow 1979, p. 190). This is suggestive evidence against the idea of them being systematically favoured by the new king.

After seizing power, William replaced the Anglo-Saxon elite with his own people in order to secure his position (Finn 1963). Anglo-Saxon noblemen that fought with Harold were considered as traitors of the legitimate king. Therefore, they were killed and their land was redistributed mainly to Norman noblemen loyal to William and, in some cases, to monasteries. In this enormous redistributive endeavour, William was not in favour of increasing the power of the Church. On the contrary, Finn (1963) notes that William "did not materially increase the church possession under his reign", and it seems that the main reason driving re-assignments was military protection. In particular, Fleming (1991) and Thomas (2008) discuss the methods behind the redistribution, pointing out that it is hard to say in what proportion of the assignment it was due to each of those. According to Thomas (2008, ch. 3), the four methods were:

1. Direct succession, i.e. William designing a Norman landlord as successor of a

[16]Hundreds fairly correspond to contemporary local districts. They were centred upon a local court, responsible for tax collection and military service obligations. As McDonald and Snooks (1986, p. 23) note, the social and economic hierarchical structure within these units had a long history, although it was "probably intensified during the Scandinavian invasions".

[17]Interestingly, the Domesday Book collect data for 1086 and for 1066 as well, i.e. the last point in time where King Edward was alive, precisely for this reason.

dispossessed Anglo-Saxon landlord;

2. Military protection: "in militarily sensitive regions, the King gave blocks of properties, sometimes entire counties such as Cheshire, to individual lords";

3. "in some areas, William lumped together all or most of the lands of minor landholders that were 'left over' after the redistribution of major estates and gave them to a single Norman lord";

4. "Norman nobles, particularly sheriffs, also simply grabbed new lands";

Overall, it seems that William was concerned with two things. First, foreign invasions. Hence his closest allies and family members received control of lands located in regions of strategic importance, either to defend the kingdom against Viking invasions or to protect communication lines between England and Normandy (Finn 1963). Second, he wanted to make sure that none of the nobles was sufficiently strong to rebel successfully (Thomas 2008). As a consequence, he made sure that the new Norman nobility was richer than the Anglo-Saxon one, on average, but none of them was individually able to challenge his power. In all of this, we could not find any historical source pointing out a strategic re-assignment of holdings to monasteries.

Twenty years after the Conquest, he commissioned a complete survey of the land ownership in his reign. The result is known as Domesday Book, "because its decisions, like those of the last judgement, are unalterable".[18]

Several scholars, both in humanities and in economics (Darby 1977, McDonald and Snooks 1987, McDonald 1997), looked in detail at the content of the Domesday Book and at the way those data were gathered. The objective of such an intense data collection ordered by William is still debated, but according to McDonald and Snooks (1986) it was a combination of two instances: assessing the ability to pay taxes and clarifying the feudal structure of the kingdom. William sent royal commissioners everywhere in his kingdom, and the data they gathered (holding's "value to its owner", i.e. our measure of productive capacity, in 1066 and 1086, landlord in 1066 when King Edward was alive and in 1086 etc etc) were then verified in open court (McDonald and Snooks 1987). Valuations, corresponding to holding productive capacity (Roffe 2000), where

[18]Richard Fitz Nigel (treasurer of Henry II), Dialogus de Scaccario, cited in Roffe (2000, p.5).

collected in pounds and shelling that, at the time, existed only as a unit of accountancy (Finn 1963). Overall, the Domesday Book seems to provide a reliable measure of land "value" and tax assessment (McDonald and Snooks 1986), and its data are suitable for cliometric analysis (McDonald and Snooks 1987).

4.3 Data

We assembled an original dataset by gathering information from a variety of different sources, as briefly outlined in Table 4.1, creating a dataset accruing to more than 9,000 observations at holding level.[19] The main source of information is the Domesday Book, DDB henceforth.[20]

The most important information included in the DDB is the name of the lord of the holding in 1066, at the time of King Edward (i.e *before* the Norman conquest), and the name of the lord in 1086, at the time the DDB was compiled under the reign of King William. The Domesday Book reports 48 different landlord names referring to a monastery in 1066 (before the Conquest) and 80 in 1086 (after the Conquest).[21]

To identify whether the lord is a Benedictine monastery, we matched lord names, as recorded in the DDB, with the names of Benedictine houses, retrieved from the English Monastic Archive (D'Avray 2015). Through the investigation of lord names we also identified holdings held by a Catholic bishop. Therefore, we classify all the entries in our dataset according to the following three types, at two point in times (before and after the Norman conquest):

- **Benedictine**, when the lord is a Benedictine monastery;
- **Bishop**, when the lord is either a Bishop or the canons of a cathedral church;
- **Secular**, when the lord is none of the above, mostly a lay nobleman or the King himself.

[19] The actual number of observations in the analysis depends on sparse missing data within each individual variable.

[20] The Domesday Book has been digitised by Palmer and colleagues and a downloadable version is freely available at the Hydra repository of the University of Hull. Please see `http://www.domesdaybook.net/` and `https://hydra.hull.ac.uk/resources/hull:domesdayDisplaySet`.

[21] In our analysis we use 31 and 60 monasteries as we have to restrict the sample to observations where the outcome variable is available for both 1066 and 1086.

We excluded from the sample about 0.2 percent of holdings that were attributed to non Benedictine monasteries and about 0.4 percent that were attributed to Benedictine Nuns, in order to obtain a homogeneous "treatment" group solely composed by male Benedictine monasteries.[22] Figures 4.2 and 4.3 provide a graphical outlook of the geographical distribution of holdings by type of landlords before (1066) and after (1086) the Conquest.

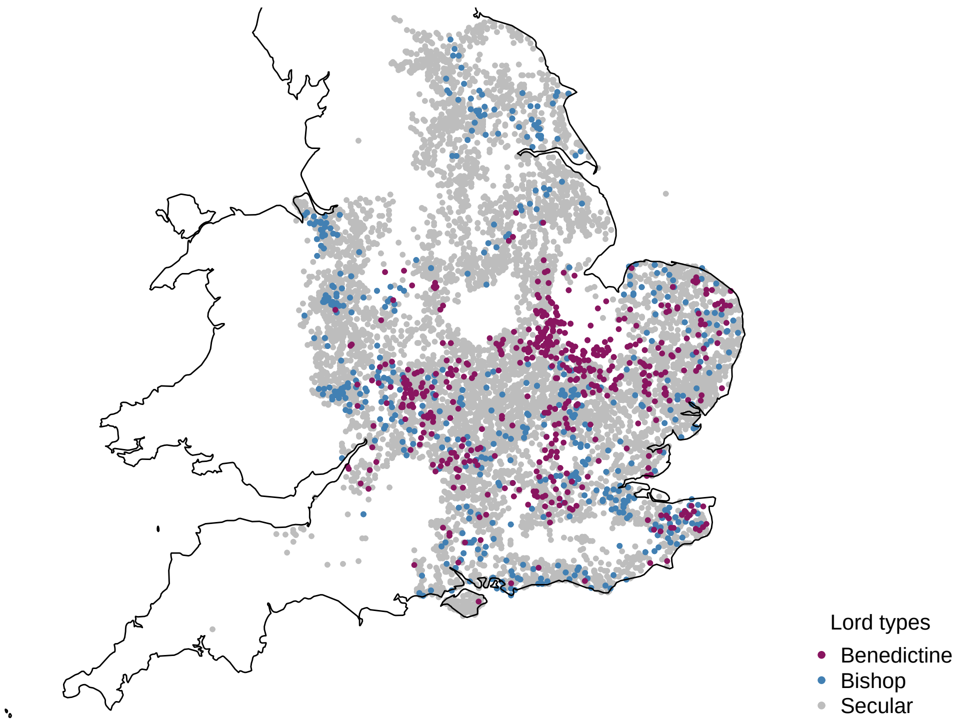


Figure 4.2: Holdings of Benedictine monasteries in England (1066)

4.3.1 Outcome variables

As explained in Section 4.2.3, some of the information in DDB are available at two points in time: the first one refers to the last moment before the Norman conquest, thus we label it 1066; the second information refers to the time DDB was actually written, after the Conquest, in 1086. Exploiting this feature, we gather from the DDB the main outcome variable of our analysis, i.e. the holding's "value" in 1066 and 1086, from which we calculate the overall growth rate.[23] Historians have widely debated the

[22] Alternatively, we included these two small groups of observation in the control group, obtaining substantially unchanged results. We further excluded from our sample 38 observations (about 0.38 percent of total "raw" observations) whose property was shared among lords of different types.

[23] The actual growth rate is computed as $Growthrate = log(Value1086 + 1) - log(Value1066 + 1)$.

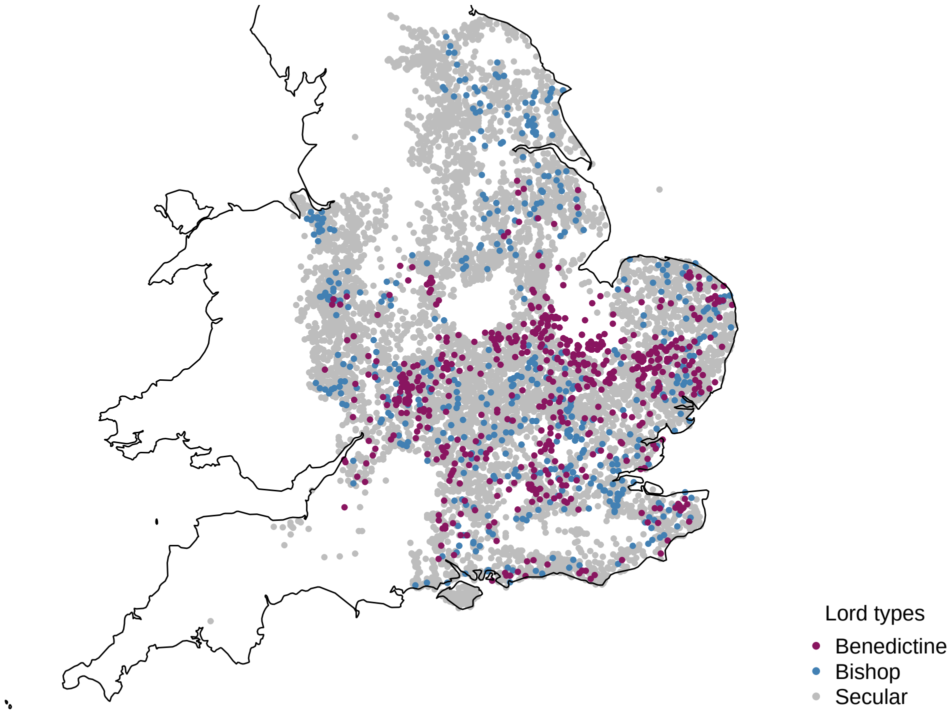


Figure 4.3: Holdings of Benedictine monasteries in England (1086)

actual meaning of the information about "values" recorded in the DDB (McDonald and Snooks 1986, Roffe 2000). After a thorough survey of the most recent developments on the issue, Roffe (2007, p. 241) concludes that the prevailing current opinion is that: "Domesday values are a more or less accurate index of the productive capacity of estates", and we stick to this interpretation throughout the paper. Further, these data have been shown to be highly consistent with all the variables reported in the DDB (McDonald and Snooks 1987, Roffe 2000, Roffe 2007), as well as when compared to subsequent historical surveys of some regions of England (Wareham 2000).

To study additional economic and social mechanisms we also collected information about total population in the holding, as recorded in the Domesday Book.[24] Further, we collected information about the presence of mills and ploughs as recorded in the DDB.

[24]This information is available only at one point in time and only for a limited subsample, therefore it cannot be used to fully replicate our main analysis. Population is calculated as the sum of all categories of people living in the estate, as recorded in the DDB, namely: villagers, smallholders, freemen, cottagers, slaves, burgesses and cases listed as "other population".

4.3.2 Geography

The most important difference across holdings is likely to depend on their geographical location. The digitised version of the DDB includes important geographical information related to each entry:

- Ordinance Survey Grid positions, that we converted into Latitude and Longitude;
- County and Hundred (local district), that we exploited to include geographical fixed effects in our analysis (see "Geographical fixed effect" section in Table 4.1).

Latitude and longitude have been included to account for the uneven distribution of monastic estates throughout England. Further, following Becker et al. (2015), all our models also include an interaction term between Latitude and Longitude. Since all holdings in our dataset have been geolocated we can also match them with geographical information about the features of the terrain, that are likely to affect soil productivity and hence potential productivity. For this reason we include the median altitude of terrain, retrieved from the GAEZ dataset at a very high resolution, which roughly corresponds to grid cells of about 10 km. Higher terrain are expected to be less productive hence median altitude is expected to be negatively correlated with value, as it is the case in our sample. Further, since not all terrain may be equally fertile and suitable for farming, we include a measure of crop suitability, retrieved again from GAEZ. This database provides alternative measures of crop suitability: we decided to include suitability related to cereals for low input level rain-fed terrains, since we expect that such an index better approximate the actual conditions of farming faced around XI c. and the same measure is used in Angelucci et al. (2017). Moreover, we include an index for terrain ruggedness, as proposed by Nunn and Puga (2012). This index targets the potential detrimental effect on soil productivity (and hence on the value of the estate) driven by sharp elevation differences within small areas. Finally, we include agriculture suitability, pastoralism suitability and sedentary animal husbandry suitability, retrieved from Beck and Sieber (2010). These three indices are constructed on a high spatial resolution using raster cells of about 5km and provide indicators of the suitability of land for alternative use in primary economic activities. With the inclusion of these controls we can account for *ex-ante* advantages in the holdings' productivity.

All the geographic controls illustrated in this paragraph have been included in our dataset by superimposing the original raster files on our map of geolocalised holdings.

4.3.3 History

To account for potential heterogeneity in the historical roots of holdings we include two measure to control for proximity to ancient Roman settlements and road network. To obtain the first measure we calculated the geodesic distance from each holding to Roman rural settlements in Britain by relying on the geolocated information provided by Allen et al. (2018). This database provides geographical coordinates of rural placements of Roman Britain, identified through a vast range of archaeological sources. Areas of more ancient settlement may benefit from higher levels of development. The second measure is calculated as the geodesic distance to the nearest Roman road. We retrieved information and maps of the Roman road network in Britain in McCormick et al. (2010). All distances are expressed in kilometres.

4.3.4 Market access

Our measure of productive capacity could be affected by better access to markets. To control for this feature we included a set of (geodesic) distance measures. Firstly, we retrieved historical maps of Britain in the Barrington Atlas of the Greek and Roman World (Talbert 2000) to calculate distance to rivers and coasts. Both distances approximate market access since waterways were an extremely important transportation network in the Early Middle Ages. Further, distance to coast also approximate the strategic position of the holding and its possible vulnerability due to Vikings raids. Secondly, we include measures of distance to actual markets approximated by distance to the nearest borough (as identified in the DDB), to the Tower of London and to the closest fair or market, as recorded in Letters (2005). The latter dataset provide valuable information about the location of markets and fairs in early medieval England. Since the date of the first appearance of chartered markets might be uncertain, we consider all markets or fairs that were recorded before 1066, or that were recorded in a range of time including 1066 or whose mint date was recorded in an analogous way.

4.3.5 Holding-specific

To account for further potentially omitted variables that are specific to individual holdings we include three further controls, retrieved from the DDB. The first one is holding's value at the time of King Edward the Confessor (i.e. 1066). This variable constitutes a proper baseline measure, something extremely rare to be found in historical databases. The second one is a measure of the size of the holding, to account for the high heterogeneity in holdings' features and sizes.[25] Moreover, we calculated the number of holdings held by each individual landlord, as identified in the DDB, to account for the potential effects on governance for landlords owning a large number of properties. Finally, we also acknowledge that more spatially concentrated holdings may be easier to govern and might give rise to economies of scale for the landlord. Therefore, if X_i and Y_i are the spatial coordinates of holding i, we compute the geographic dispersion of the n holdings belonging to the same landlord by calculating the standard distance, H_D as follows: $H_D = \sqrt{\frac{\Sigma\left(X_i - \overline{X}_c\right)^2 + \Sigma\left(Y_i - \overline{Y}_c\right)^2}{n}}$ where $\overline{X}_c$ and $\overline{Y}_c$ are the coordinates of the centre with respect to the n holdings. Standard distance is the spatial equivalent of standard deviation of a distribution and provide an approximation of the degrees of concentration (or dispersion) of a group of spatial units.

4.4 Empirical Strategy

Objective of this paper is the estimation of the effect of being governed by a monastery *vis-à-vis* a secular landlord on economic outcomes. The main explanatory variable is a dummy equal to 1 if an holding was controlled by a Benedictine monastery; the outcome variable is the growth rate of holding's value between 1066 and 1086, as a measure for its change in productive capacity.

Clearly, the relationship between these two variables is likely to be spurious for many reasons. Monasteries (and hence their estates) tend to be located more often in the southern part of England, and they tend to be located in the most ancient settled areas (Postan 1973). Moreover, monasteries may have been granted systematically better (or

[25] To ensure comparability across holdings we only keep observations whose size has been measured in "geld", as recorded in Palmer (2008). According to Palmer, this is the best proxy for the size of estates.

Table 4.1: Variables description

	Description	Source
Benedictine (1066) or (1086)	Dummy variable equal to 1 if the holding's lord is a Benedictine monastery at the time of King Edward, i.e. *before* the Conquest (1066), or at the time of King William (1086),	DDB[1]and EMA[2]
Bishop (1066) or (1086)	Dummy variable equal to 1 if the holding's lord is a Bishop or the Canons of a cathedral church at the time of King Edward, i.e. *before* the Conquest (1066), or at the time of King William (1086),	DDB[1]
Outcome variables		
Growth rate (log)	Natural log of growth of holding's value between 1086 and 1066	DDB[1]
Population (1086)	Number of households settled in the holding	DDB[1]
Mills (1086)	Number of mills recorded in the holding	DDB[1]
Ploughs (1086)	Number of ploughs in the holding	DDB[1]
Geography		
Latitude	Latitude of the holding, decimal degrees	DDB[1]
Longitude	Longitude of the holding, decimal degrees	DDB[1]
Median altitude	Median altitude of terrain, resolution 5 arc-minutes	GAEZ[3]
Crop suitability	Index of cereals suitability for low input level rain-fed cereals, baseline period 1961-1990, resolution 5 arc-minutes	GAEZ[3]
Ruggedness	Index for rugged terrain, measured in hundreds of meters of elevation difference for grid points 30 arc-seconds apart.	Nunn and Pulga (2012)
Agriculture suitability	Probability of occurrence of agriculture estimated through maximum entropy modelling, baseline period 1961-1991, resolution 2.5 arc minutes	Beck and Sieber (2010)
Pasture suitability	Probability of occurrence of pasture estimated through maximum entropy modelling, baseline period 1961-1991, resolution 2.5 arc minutes	Beck and Sieber (2010)
Animal suitability	Probability of occurrence of animal husbandry estimated through maximum entropy modelling, baseline period 1961-1991, resolution 2.5 arc minutes	Beck and Sieber (2010)
History		
Dist. Roman settl. (log)	Geodesic distance to nearest Roman settlement, as identified in the Rural settlements of Roman Britain database	Authors' calculations based on ADS[4]
Dist. Roman roads (log)	Geodesic distance to nearest Roman road	Authors' calculations based on McCormick et al. (2010)
Market access		
Dist. rivers (log)	Geodesic distance to nearest river	Authors' calculations based on BA[5]
Dist. coast (log)	Geodesic distance to coastline	Authors' calculations based on BA[5]
Dist. London (log)	Geodesic distance to the Tower of London (latitude 51.5048, longitude -0.0723)	Authors' calculations
Dist. borough (log)	Geodesic distance to nearest borough, as defined in DDB	Authors calculations based on DDB
Dist. to markets	Geodesic distance to closest market or fair	Authors calculations based on Letters (2005)
Holding-specific		
Holding's value, 1066 (log)	Natural log of value to the lord of the value of the holding in 1066, as recorded in DDB	DDB[1]
Size (log)	Natural log of holding's size as recorded in DDB	DDB[1]
Holdings number (log)	Natural log of the number of holdings held by same lord of the observed holding in 1066 and 1086 (as indicated in parentheses), as recorded in DDB	Authors' calculations based on DDB[1]
Holding's dispersion	Dispersion of holdings of the same Lord calculated both for 1066 and 1086 (as indicated in parentheses) as follows: $H_D = \sqrt{\frac{\Sigma\left(X_i - \overline{X}_c\right)^2 + \Sigma\left(Y_i - \overline{Y}_c\right)^2}{n}}$	Authors' calculations based on DDB[1]
Geographic fixed effects		
County FE	Dummies that uniquely identifies holdings w.r.t. their county	DDB[1]
Hundred FE	Dummies that uniquely identifies holdings w.r.t. their local district, named Hundred, as reported in DDB	DDB[1]
IV for diff-in-diff		
Dist. Benedictine lord	Geodesic distance to the nearest Benedictine Monastery that is enlisted among the land holders in 1066.	DDB[1]and EMA[2]
Other variables		
Dist. to Benedictine Monastery	Geodesic distance to any nearest Benedictine Monastery, regardless of the fact that it is a landlord or not, founded before 1066.	EMA[2]

Notes:
[1] Domesday Book Statistics, Palmer (2008);
[2] English Monastic Archive, D'Avray (2015);
[3] Global Agro-Ecological Zones (GAEZ v3.0), IIASA/FAO (2012);
[4] Archaeology Data Service, The Rural Settlement of Roman Britain: an online resource, Allen et al. (2018);
[5] The Barrington Atlas of the Greek and Roman World, Talbert (2000). Digital maps retrieved at `http://awmc.unc.edu/wordpress/map-files/`;

Table 4.2: Summary statistics

	Mean	SD	Min	Max	Obs
Benedictine (1066)	0.054	0.227	0.000	1.000	9643
Outcome variables					
Growth rate (log)	−0.033	0.398	−3.892	2.818	9643
Population (1086)	15.834	26.232	0.000	1145.000	9638
Mills (1086)	1.419	1.113	0.000	13.500	2494
Ploughs (1086)	4.871	6.987	0.000	173.000	9592
Geography					
Latitude	52.319	0.846	50.502	54.611	9643
Longitude	−0.638	1.145	−4.074	1.742	9643
Latitude*Longitude	−33.529	59.999	−205.757	91.468	9643
Median altitude	70.197	49.895	−1.000	404.000	9642
Crop suitability	4.047	1.384	0.000	7.615	9642
Ruggedness	14.855	19.857	0.000	208.646	9643
Agricultural suitability	0.518	0.064	0.351	0.789	9641
Pasture suitability	0.151	0.076	0.049	0.532	9641
Animal husbandry suitability	0.463	0.165	0.189	0.816	9641
History					
Dist. Roman settl. (log)	1.439	0.556	0.000	3.989	9643
Dist. Roman roads (log)	1.672	0.754	0.002	3.982	9643
Market access					
Dist. rivers (log)	2.993	0.996	0.006	5.209	9643
Dist. coast (log)	3.420	1.061	0.065	4.921	9643
Dist. London (log)	4.766	0.633	0.865	5.919	9643
Dist. borough (log)	3.307	0.726	0.000	4.800	9643
Dist. markets (log)	2.675	0.628	0.044	4.286	9643
Holding-specific					
Holding value, 1066 (log)	1.255	0.831	0.000	5.707	9643
Size (log)	1.253	0.740	0.000	4.646	9643
Holdings of the same lord in 1066 (log)	3.255	1.632	0.693	6.880	9643
Holdings' Dispersion (1066)	0.852	0.664	0.000	3.933	9643

worse) holdings. Indeed, historical evidence has shown that among the most strenuous opponents William faced after he landed on England were some Benedictine monasteries, since most of the monks were part of the Anglo-Saxon aristocracy (Knowles 1966).

To isolate the effect of the institutional structure from everything else, we employ two different strategies. The first one exploits the geographical information contained in our dataset, that allow us to compare holdings of different ownership located in the same small administrative area, controlling for a wide range of geographic and economic factors. The second one uses the post-Hastings re-assignment of holdings formerly owned by the Anglo-Saxon nobility as an historic "natural experiment".

4.4.1 Geographic fixed effects

Since we know the geographic location of every estate, we are able to collect a number of controls, described in Section 4.3, related with geographic conditions (latitude, longitude, altitude, ruggedness, terrain suitability to cereal, agriculture and pasture), history (distance to Roman roads and Roman settlements), market access (distance to marketplaces, boroughs, coasts, rivers and London) and holding-specific (size, number of holdings controlled by the same landlord, their dispersion, value in 1066). Moreover, we exploit information on the administrative structure of the Norman kingdom to include dummies for each county or each hundred. In practice, we estimate by OLS the following specification:

$$Y_{i,g} = \gamma_g + \delta Benedict_{i,g} + \beta^{'} \mathbf{x}_{i,g} + \epsilon_{i,g} \tag{4.1}$$

where $Y_{i,g}$ is the growth rate of value over 20 years for holding i located in the geographic unit g, γ_g are geographic (county or hundred) dummies, $Benedict_{i,g}$ is a dummy equal to one if holding i was controlled by a Benedictine monastery in 1066 and $\mathbf{x}_{i,g}$ is the vector of holding-varying controls mentioned above.

The coefficient of interest is δ and, in the most demanding specification, we are comparing monastic and secular estates located within small administrative units. The combination of this feature with the holding specific controls we collected should address most of the endogeneity concerns coming from pure geographic considerations. It remains possible, however, that monasteries managed to systematically select the best land (in terms of unobservable characteristics we are not accounting for) within each hundred. Moreover, the Norman conquest affected the land controlled by non-monastic lords in a disproportionate way, because of the complete annihilation of the Anglo-Saxon elite. Section 4.4.2 tries to address those concerns as well.

4.4.2 Ownership change and diff-in-diff estimation

With the Norman conquest, holdings belonging to Anglo-Saxon landlords were re-assigned by King William after the Battle of Hastings. Some of them were assigned to new secular landlords, other were assigned to monasteries. In this specification, we focus on the sub-sample of holdings owned by Anglo-Saxon landlords that changed ownership with the Norman conquest, using those that ended up to be controlled by Benedictine monasteries as "treatment group" and the rest as "control group".

The fact that every holding, in this sample, has been reassigned, allows us to disentangle the effect of ownership change from the effect of being controlled by a Benedictine monastery.

Using the panel structure of the dataset, we are substantially estimating a diff-in-diff (DID) specification, where we compare pre- and post-Conquest (log of) value for holdings moving from an Anglo-Saxon secular landlord to a Benedictine monastery and holdings moving from an Anglo-Saxon secular landlord to a Norman landlord. In practice, we estimate

$$Y_{i,t} = \tau_t + \mu_i + \phi Benedict_{i,t} + \lambda^{'} \mathbf{x}_{i,t} + \epsilon_{i,t} \tag{4.2}$$

where $Y_{i,t}$ is the (log of) annual value of holding i at time t, τ_t time fixed effect, μ_i is holding fixed effect, $Benedict_{i,t}$ is a dummy for holdings controlled by Benedictine monasteries and $x_{i,t}$ is a vector of time varying and time invariant controls interacted with time fixed effect. Under the assumptions of parallel trends and that nothing else happens at the same time affecting treatment and control group in a different way, ϕ captures the causal effect of being controlled by a Benedictine monastery on

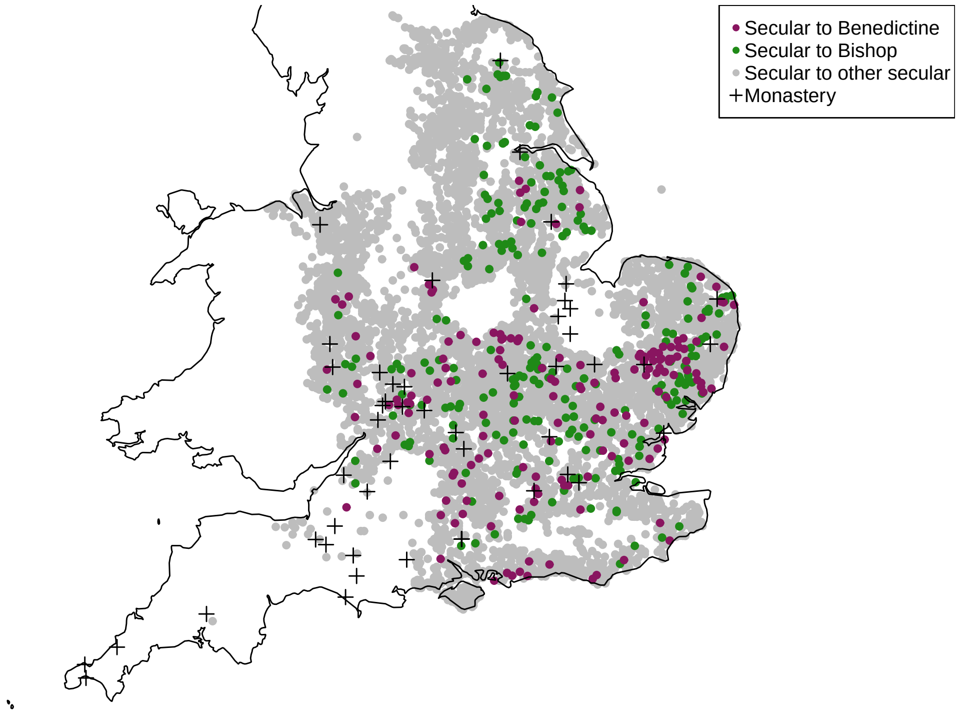


Figure 4.4: Ownership change

Note: Holdings whose landlord changed between 1066 and 1086. Grey dots are those switching from a secular landlord to another secular landlord (presumably, from an Anglo-Saxon to a Norman nobleman), green dots are those switching from a secular landlord to a bishop and purple dots are those switching from a secular landlord to a monastery. Black crosses represent monasteries.

our measure of land productive capacity. Note that we are taking care of every time invariant holding specific characteristic.

Given the fact that we are using historical data and we have only two time periods, unfortunately we cannot test the parallel trend assumption directly. However, as explained in Section 4.2, there is no historical evidence we are aware of reporting King William systematically favouring monasteries in the re-assignment. Consistently with this, the balance table we show in section 4.5.2 looks fairly good, as the main significant difference in observables between treatment and control group comes from the tendency of the former to be located in Southern England. This is clearly due to pre-existing historical reasons, that are time invariant and hence unlikely to induce a systematic bias in the coefficient. New monasteries spread all over England starting from Canterbury, keeping cultural and religious ties with Rome and the continental Europe application of Benedict's Rule.[26] Not only treatment and control estates are fairly balanced in terms of observables: the value level pre-Conquest is also not statistically different so, in terms of outcome, they start from the same point.

Overall, the data are consistent with the idea that William decided the re-assignment without discriminating in favour of monasteries, as the land they received was very similar, on average, to the land assigned to secular landlords. Moreover, our results are robust to the introduction of all the aforementioned controls interacted with time. Finally, we are not aware of other events systematically affecting one of the two groups of holdings[27]. In particular, it is worth pointing out that the rebellions that followed Hastings involved monasteries as well, and William had to re-conquer the monastery of Ely during the most challenging one (Thomas 2008, ch. 2).

4.5 Results

4.5.1 Geographic fixed effects

At first we compare the 20-years value growth rate of holdings that were controlled by Benedictine monasteries in 1066 and holdings controlled by secular landlords at the same time. The idea is that those estates were probably in the same (monastic or secular) hands way before the last day of King Edward, hence we want to test whether there is some persistence of a "monastic effect" in our data. Obviously, the vast majority of secular holdings also changed owner during this period, a concern that will be addressed in Section 4.5.2.

Table 4.3 summarises the main findings. Table A4.2 performs the same exercise with a less demanding specification in terms of geographic controls (i.e. county fixed effects) and again everything remains stable and consistent.

Column (1) of Table 4.3 shows the correlation between the Benedictine dummy and the productive capacity growth rate, with only Hundred FE. Hence, we are comparing

[26] During the reign of Louis the Pious (814-840) the application of Benedict' Rule became mandatory in every monastery throughout his Empire, a provision ultimately affecting also monasteries lying outside the Carolingian Empire, due to cultural and affiliation ties among Benedictine monasteries (Dell'Omo 2011).

[27] A number of Viking raids are known to hit the North-Western part of England around 1070, but the potential detrimental effects of these events are already captured by the geographical control variables we include in our models.

Table 4.3: Growth rate of holding value: Benedictine and secular landlords compared

	(1)	(2)	(3)	(4)	(5)	(6)
Benedictine (1066)	0.057 * **	0.054 * *	0.057 * **	0.058 * **	0.039*	0.037*
	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)
Included controls:						
Geography[1]	*No*	*Yes*	*No*	*No*	*No*	*Yes*
Market access[2]	*No*	*No*	*Yes*	*No*	*No*	*Yes*
History[3]	*No*	*No*	*No*	*Yes*	*No*	*Yes*
Holding specific[4]	*No*	*No*	*No*	*No*	*Yes*	*Yes*
Hundred FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Obs	9287	9285	9287	9287	9287	9285

Notes. Dependent variable: value growth rate. Standard errors clustered at lord level (1066). Shared holdings and outliers (i.e. 7 estates recording growth rate larger than 30) are excluded from the sample.

[1] Geography controls include geographic/agriculture related features, namely: latitude, longitude, latitude×longitude, median altitude, crop suitability, ruggedness, pasture suitability, agricultural suitability, animal husbandry suitability;

[2] Market access controls include proxies for access to markets: (log of) distance (in km) to: rivers, coast, London, nearest borough;

[3] History controls include proxies for the ancientness of the settlement, namely: (log of) distance (in km) to Roman settlements and to Roman roads;

[4] Holding-specific controls include holding's value in 1066, size, the number of holdings held by the same lord and holding dispersion.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Benedictine and non-Benedictine holdings located within the same small geographic unit. On top of that, columns (2) to (6) add different sets of control variables, as outlined above. The most restrictive specification, in column (6), includes all the controls as well as hundred dummies. Even in that case, Benedictine estates exhibit a 3.7 percentage points higher growth rate in the 20 years after the Norman conquest.

Together with overall productive capacity, there are few other outcomes that we can test with this cross sectional approach: population, presence of mills and number of ploughs.[28] Table 4.4 reports the results in the most demanding specification, the same as in column (6) of Table 4.3. Interestingly, Benedictine holdings are not different (or worse) from the others in terms of technology (ploughs and mills) or population. Of course, since we are dealing with levels we cannot test whether there was a change with respect to the pre-Conquest period.

4.5.2 Ownership change and Diff-in-diff estimation

Table A4.4 describes the subsample used for this section and group-specific summary statistics are in Table 4.5. As mentioned, we keep only estates that were held by secular landlords in 1066 and that changed owner between 1066 and 1086. In this case, we are exploiting the panel structure of our dataset and the outcome variable is the log of

[28] Unfortunately, data for these outcomes are available for 1086 only, hence we cannot use them in the panel approach.

Table 4.4: Other outcomes: Benedictine and other landlords compared

	(1) Population	(2) Ploughs	(3) Mills
Benedictine (1066)	0.134	-0.019	$-2.210**$
	(0.971)	(0.336)	(0.898)
Included controls:			
Geography[1]	Yes	Yes	Yes
Market access[2]	Yes	Yes	Yes
History[3]	Yes	Yes	Yes
Holding specific[4]	Yes	Yes	Yes
Hundred FE	Yes	Yes	Yes
Obs	9280	9236	227

Notes. Dependent variable: Population (column 1), Ploughs (col. 2), Presence of Mills (col. 3). Standard errors clustered at lord level (1066). Shared holdings and outliers (i.e. 7 estates recording growth rate larger than 30) are excluded from the sample.

[†] Logit model, estimated margins reported in table.

[1] Geography controls include geographic/agriculture related features, namely: latitude, longitude, latitude×longitude, median altitude, crop suitability, ruggedness, pasture suitability, agricultural suitability, animal husbandry suitability;

[2] Market access controls include proxies for access to markets: (log of) distance (in km) to: rivers, coast, London, nearest borough;

[3] History controls include proxies for the ancientness of the settlement, namely: (log of) distance (in km) to Roman settlements and to Roman roads;

[4] Holding-specific controls include size, the number of holdings held by the same lord and holding's dispersion.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

value in 1066 and 1086. Therefore, we can interpret the result as the differential effect on the productive capacity growth rate calculated over 20 years induced by Benedictine control. As most of the control variables are time-invariant (with the exception of the number of holdings controlled by the same lord and their dispersion), we add them to the specification after multiplying them by time fixed effect. From Table 4.5 it is clear that "treated" holdings tend to have a positive 5% growth rate over 20 years and "control" holdings have a negative 5% growth rate over the same period.

Table 4.5: Summary statistics (subsample) by groups

	Control					Treatment				
	Mean	SD	Min	Max	Obs	Mean	SD	Min	Max	Obs
Outcome variables										
Holding's value, 1066 (log)	1.224	0.802	0.000	4.727	7612	1.455	0.897	0.049	4.466	175
Holding's value, 1086 (log)	1.173	0.856	0.000	4.987	7612	1.501	0.874	0.000	4.615	175
Geography										
Latitude	52.363	0.864	50.502	54.611	7612	52.013	0.581	50.777	53.608	175
Longitude	-0.616	1.157	-4.074	1.742	7612	-0.405	1.154	-2.861	1.709	175
Latitude*Longitude	-32.411	60.659	-205.757	91.468	7612	-21.015	60.156	-151.033	89.981	175
Median altitude	69.871	50.187	-1.000	404.000	7611	70.486	48.708	1.000	259.000	175
Crop suitability	4.063	1.383	0.000	7.615	7611	4.136	1.289	0.000	7.565	175
Ruggedness	14.998	20.065	0.000	208.646	7612	11.776	16.097	0.000	91.389	175
Pasture suitability	0.148	0.073	0.049	0.529	7610	0.153	0.074	0.061	0.416	175
Agricultural suitability	0.516	0.063	0.351	0.789	7610	0.520	0.072	0.383	0.776	175
Animal husbandry suitability	0.459	0.162	0.189	0.816	7610	0.456	0.164	0.222	0.775	175
History										
Dist. Roman settl. (log)	1.458	0.557	0.000	3.989	7612	1.408	0.479	0.361	2.570	175
Dist. Roman roads (log)	1.690	0.754	0.002	3.982	7612	1.588	0.746	0.015	3.181	175
Market access										
Dist. rivers (log)	3.040	0.980	0.006	5.209	7612	3.117	0.944	0.395	4.510	175
Dist. coast (log)	3.385	1.068	0.065	4.908	7612	3.517	0.944	0.116	4.772	175
Dist. London (log)	4.798	0.618	0.865	5.919	7612	4.562	0.578	2.396	5.484	175
Dist. borough (log)	3.339	0.712	0.000	4.800	7612	3.323	0.687	0.882	4.636	175
Dist. markets (log)	2.700	0.623	0.132	4.286	7612	2.532	0.534	0.589	3.571	175
Manor-specific										
Size (log)	1.211	0.714	0.012	4.646	7612	1.383	0.767	0.095	4s.407	175
Holdings of the same lord in 1066 (log)	3.096	1.550	0.693	6.332	7612	3.404	1.527	0.693	6.332	175
Holdings of the same lord in 1086 (log)	3.192	1.616	0.693	7.141	7612	3.080	1.422	0.693	4.990	175
Holdings' dispersion (1066)	0.890	0.692	0.000	3.933	7612	0.930	0.632	0.000	2.380	175
Holdings' dispersion (1086)	0.850	0.645	0.000	3.348	7612	0.602	0.418	0.000	2.326	175

The results of (4.2) are reported in Table 4.6. We report only the estimated values for ϕ. The estimates are pretty consistent across all the columns: holdings that, after the re-assignment, end up being controlled by Benedictine monasteries experience a 20-year productive capacity growth rate that is between 4 and 10 percentage points higher, depending on the specification. Moreover, as shown in Table A4.5, excluding ecclesiastical holdings from the control group does not change the result, and the same is true when we exclude holdings that end up under the direct control of the King, as shown in Table A4.6.

As we have only two time periods, unfortunately we cannot perform the usual test

Table 4.6: Holding value: Benedictine and other lords compared

	(1)	(2)	(3)	(4)	(5)	(6)
Benedictine	0.098 * **	0.048 * *	0.055 * *	0.096 * **	0.096 * **	0.039*
	(0.030)	(0.022)	(0.023)	(0.030)	(0.032)	(0.022)
Manor FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Time FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Geography*time ctrls[1]	*No*	*Yes*	*No*	*No*	*No*	*Yes*
Market*time ctrls[2]	*No*	*No*	*Yes*	*No*	*No*	*Yes*
History*time ctrls[3]	*No*	*No*	*No*	*Yes*	*No*	*Yes*
Holding*time ctrls[4]	*No*	*No*	*No*	*No*	*Yes*	*Yes*
Obs	15574	15570	15574	15574	15574	15570

Notes. Dependent variable: log of value. Standard errors clustered at lord level (1086). Shared holdings and outliers are excluded from the sample;

[1] Geography controls include geographic/agriculture related features, namely: latitude, longitude, latitude×longitude, median altitude, crop suitability, ruggedness, pasture suitability, agricultural suitability, animal husbandry suitability;

[2] Market access controls include proxies for access to markets: (log of) distance (in km) to: rivers, coast, London, nearest borough;

[3] History controls include proxies for the ancientness of the settlement, namely: (log of) distance (in km) to Roman settlements and to Roman roads;

[4] Holding-specific controls include size, the number of holdings held by the same lord and holding dispersion.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

for the parallel trend assumption. The best proxy we could find is a set of balance tests between treatment and control groups on the sample we are using. Results are reported in Table 4.7. Treatment and control groups are statistically different only with respect to few dimensions. Two of them (distance to London and latitude) have a clear interpretation based on the development of monasteries in England. As there were more monasteries in the south, it seems reasonable that holdings in the south are more likely to end up under the control of a monastery, after the re-assignment. The effect of this feature on the growth perspective is likely to be time invariant, hence it is taken care of by the holding fixed effect.

The third dimension is the number of holdings controlled by the same lord in 1066. As the measure is time varying, we can take care of it directly in the regression. Moreover, this seems consistent with the idea that holdings re-assigned to monasteries were probably a “residual” of big pre-existing holdings divided in different ways for reasons related with internal or external security. In fact, the two groups are no longer significantly different on this dimension in 1086. Fourth, holdings assigned to Benedictine monasteries are closer to marketplaces. Finally, note that the outcome variable “pre-treatment” is not statistically different between the two groups, hence there is no evidence of selection in terms of observable *ex ante* productive capacity.

As the re-assignment procedure was clearly not random, however, it is possible that the King could have chosen not to re-assign to his allies holdings with a higher growth prospect. Obviously we cannot enter in King William’s mind, but historians seem to agree that the main reasons were internal and external security. On top of that, the King could have assigned those high-prospect estates to himself directly. However, this seem not to be the case, as a comparison between King's and secular landlords' holdings, summarised in Table 4.8, shows a consistently negative (or close to zero) sign.

Table 4.7: Balance test (subsample)

	(1) Value (1066)	(2) Size	(3) Holdings same lord (1066)	(4) Holdings' dispersion (1066)
Benedictine (1086)	0.231	0.172	0.308***	0.041
	(0.185)	(0.145)	(0.116)	(0.097)
Constant	1.224***	1.211***	3.096***	0.890***
	(0.017)	(0.018)	(0.049)	(0.015)
Obs	7787	7787	7787	7787
	(5) Latitude	(6) Longitude	(7) Lat × Lon	(8) Median altitude
Benedictine (1086)	−0.350***	0.211	11.396	0.615
	(0.103)	(0.319)	(16.633)	(6.230)
Constant	52.363***	−0.616***	−32.411***	69.871***
	(0.041)	(0.039)	(2.073)	(1.216)
Obs	7787	7787	7787	7786
	(9) Crop suitability	(10) Agr. suitability	(11) Past. suitability	(12) Anim. suitability
Benedictine (1086)	0.073	0.005	0.005	−0.002
	(0.193)	(0.009)	(0.011)	(0.038)
Constant	4.063***	0.516***	0.148***	0.459***
	(0.030)	(0.001)	(0.002)	(0.005)
Obs	7786	7785	7785	7785
	(13) Ruggedness	(14) Dist. to Roman roads	(15) Dist. to Roman settl.	(16) Dist. to markets
Benedictine (1086)	−3.223	−0.102	−0.050	−0.168***
	(2.295)	(0.068)	(0.036)	(0.055)
Constant	14.998***	1.690***	1.458***	2.700***
	(0.428)	(0.014)	(0.011)	(0.020)
Obs	7787	7787	7787	7787
	(17) Dist. to rivers	(18) Dist. to coast	(19) Dist. to London	(20) Dist. to borough
Benedictine (1086)	0.077	0.131	−0.236***	−0.016
	(0.164)	(0.102)	(0.078)	(0.130)
Constant	3.040***	3.385***	4.798***	3.339***
	(0.026)	(0.034)	(0.029)	(0.023)
Obs	7787	7787	7787	7787

Notes. Regression of the dependent variable stated in each column on Benedictine (1086) dummy. Sample limited to non-Benedictine holdings in 1066 that switched owner, excluding shared and outliers. Distances, value and holdings of the same lord are all in logs. Standard errors clustered at lord level (1086).
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.6 Mechanisms

As discussed in Section 4.2, there are few possible reasons economic historians have highlighted and that may explain our results. However, we are able to rule out those

Table 4.8: Holding value: King vs secular landlords

	(1)	(2)	(3)	(4)	(5)	(6)
King	$-0.081***$	-0.020	$-0.043***$	$-0.078***$	-0.019	0.013
	(0.010)	(0.018)	(0.015)	(0.011)	(0.025)	(0.025)
Manor FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Time FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Geography*time ctrls[1]	*No*	*Yes*	*No*	*No*	*No*	*Yes*
Market*time ctrls[2]	*No*	*No*	*Yes*	*No*	*No*	*Yes*
History*time ctrls[3]	*No*	*No*	*No*	*Yes*	*No*	*Yes*
Holding*time ctrls[4]	*No*	*No*	*No*	*No*	*Yes*	*Yes*
Obs	14666	14662	14666	14666	14666	14662

Notes. Dependent variable: log of value. Standard errors clustered at lord level (1086). Shared holdings and outliers are excluded from the sample, as well as holdings controlled by monasteries or bishops in 1086;

[1] Geography controls include geographic/agriculture related features, namely: latitude, longitude, latitude×longitude, median altitude, crop suitability, ruggedness, pasture suitability, agricultural suitability, animal husbandry suitability;

[2] Market access controls include proxies for access to markets: (log of) distance (in km) to: rivers, coast, London, nearest borough;

[3] History controls include proxies for the ancientness of the settlement, namely: (log of) distance (in km) to Roman settlements and to Roman roads;

[4] Holding-specific controls include size, the number of holdings held by the same lord and holding dispersion.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

that are not related with the institutional elements of the Rule. In particular, we control for the historical presence of Roman settlements in the area and for a wide range of other geographic characteristics, hence the effect we find is not due to those elements (whose importance has been stressed, for example by Postan 1973). Moreover, the effect is still present when we control for the number of holdings owned by the same landlord and for their dispersion, hence it is not due to learning or economies of scale, as proposed by Campbell (2006). The availability of cheap labour as a channel for prosperity, stressed by Ekelund et al. (1996), plays no role here, as *conversi* were not there yet. On top of this, monastic holdings seem not to be different from the others in term of population or technology, as highlighted by Table 4.4.

Another mechanism could be cultural proximity, meaning that it is not being governed by a monastery that matters, but just being close, so that people can acquire the related "hard work" mentality. This is the mechanism broadly suggested by Andersen et al. (2017), for example. In order to test this idea, we geo-coded the location of every English Benedictine monastery, irrespective of whether it was a landlord or not. Then, for all our holdings, we can calculate the distance between them and the closest monastery, and see whether it can explain productive capacity growth. Table 4.9 shows that it is not the case: the distance to the nearest monastery has no statistically significant effect on the growth rate, irrespective of whether we use hundred (column 2) or county (column 3) dummies.

Table 4.9: Growth rate of holding value: Benedict's rule and proximity to monastery compared

	(1) Benchmark	(2) Excl. Bened. (HFE)	(3) Excl. Bened. (CFE)
Dist. nearest monastery (log)		−0.012	−0.003
		(0.013)	(0.008)
Benedictine (1066)	0.037*		
	(0.022)		
Hundred FE	*Yes*	*Yes*	*No*
County FE	*No*	*No*	*Yes*
Included controls:			
Geography[1]	*Yes*	*Yes*	*Yes*
Market access[2]	*Yes*	*Yes*	*Yes*
History[3]	*Yes*	*Yes*	*Yes*
Holding specific[4]	*Yes*	*Yes*	*Yes*
Obs	9285	8761	8762

Notes. Dependent variable: value growth rate. Standard errors clustered at lord level (1066). Shared holdings and outliers (i.e. 7 estates recording growth rate larger than 30) are excluded from the sample, as well as holdings held by Female and Non-Benedictine monastic orders.

[1] Geography controls include geographic/agriculture related features, namely: latitude, longitude, latitude×longitude, median altitude, crop suitability, ruggedness, pasture suitability, agricultural suitability, animal husbandry suitability;

[2] Market access controls include proxies for access to markets: (log of) distance (in km) to: rivers, coast, London, nearest borough;

[3] History controls include proxies for the ancientness of the settlement, namely: (log of) distance (in km) to Roman settlements and to Roman roads;

[4] Holding-specific controls include size, the number of holdings held by the same lord and holding dispersion.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.10 helps ruling out two further channels, i.e. the possibility that the effect we find comes from better human capital or that it is due to the different incentives that may arise when the estate is controlled by a religious institution. We estimate the same model as described in Section 4.4.2, but now the "treatment" dummy becomes one when the land owned by an Anglo-Saxon nobleman in 1066 is controlled by a bishop in 1086. We test the same specifications as in Section 4.5.2 and we find no statistically significant difference, meaning that being re-assigned and controlled by a bishop does not lead to a statistically significant change in productive capacity growth rate. The dimension of the coefficients is halved with respect to those associated with monasteries. Note that bishops where usually educated figures and their authority was based on a religious element, as much as in the case of abbots.

Table 4.10: Holding value, bishops vs secular landlords

	(1)	(2)	(3)	(4)	(5)	(6)
Bishop	0.052	0.026	0.027	0.051	0.051	0.022
	(0.047)	(0.036)	(0.036)	(0.048)	(0.049)	(0.036)
Manor FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Time FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Geography*time ctrls[1]	*No*	*Yes*	*No*	*No*	*No*	*Yes*
Market*time ctrls[2]	*No*	*No*	*Yes*	*No*	*No*	*Yes*
History*time ctrls[3]	*No*	*No*	*No*	*Yes*	*No*	*Yes*
Holding*time ctrls[4]	*No*	*No*	*No*	*No*	*Yes*	*Yes*
Obs	15056	15052	15056	15056	15056	15052

Notes. Dependent variable: log of value. Standard errors clustered at lord level (1086). Shared holdings and outliers are excluded from the sample, as well as estates with a Benedictine owner in 1086;

[1] Geography controls include geographic/agriculture related features, namely: latitude, longitude, latitude×longitude, median altitude, crop suitability, ruggedness, pasture suitability, agricultural suitability, animal husbandry suitability;

[2] Market access controls include proxies for access to markets: (log of) distance (in km) to: rivers, coast, London, nearest borough;

[3] History controls include proxies for the ancientness of the settlement, namely: (log of) distance (in km) to Roman settlements and to Roman roads;

[4] Holding-specific controls include size, the number of holdings held by the same lord and holding dispersion.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Another way to assess the importance of human capital is through the comparison of monasteries of old and more recent foundation. Older monasteries may have had more time to develop specific skills, learn about specialisation, acquire expertise, invest in training etc. If those factors are important for the result we find, we should expect holdings assigned to old monasteries to perform better than holdings assigned to more recently-founded monasteries. We test this hypothesis adding to equation (4.2) an interaction term between the Benedictine dummy and a dummy equal to 1 if the landlord-monastery was founded before year 900 AD.[29] The results of Table 4.11,

[29] The number of observations drops because we do not have the foundation date for every monastery

however, point towards the opposite direction: the coefficient of the interaction term is never statistically different from zero and its sign is always negative. We interpret this result as further suggestive evidence against the "human capital" channel.

Table 4.11: Holding value, old vs new monasteries

	(1)	(2)	(3)	(4)	(5)	(6)
Benedictine	0.110***	0.053*	0.069***	0.111***	0.116***	0.039
	(0.043)	(0.032)	(0.027)	(0.043)	(0.044)	(0.029)
Benedictine*Pre900	−0.062	−0.032	−0.051	−0.066	−0.075	−0.025
	(0.066)	(0.048)	(0.055)	(0.067)	(0.070)	(0.054)
Manor FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Time FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Geography*time ctrls[1]	*No*	*Yes*	*No*	*No*	*No*	*Yes*
Market*time ctrls[2]	*No*	*No*	*Yes*	*No*	*No*	*Yes*
History*time ctrls[3]	*No*	*No*	*No*	*Yes*	*No*	*Yes*
Holding*time ctrls[4]	*No*	*No*	*No*	*No*	*Yes*	*Yes*
Obs	15448	15444	15448	15448	15448	15444

Notes. Dependent variable: log of value. Standard errors clustered at lord level (1086). Shared holdings and outliers are excluded from the sample;
[1] Geography controls include geographic/agriculture related features, namely: latitude, longitude, latitude×longitude, median altitude, crop suitability, ruggedness, pasture suitability, agricultural suitability, animal husbandry suitability;
[2] Market access controls include proxies for access to markets: (log of) distance (in km) to: rivers, coast, London, nearest borough;
[3] History controls include proxies for the ancientness of the settlement, namely: (log of) distance (in km) to Roman settlements and to Roman roads;
[4] Holding-specific controls include size, the number of holdings held by the same lord and holding dispersion.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.12 compares values of holdings controlled by Benedictines in both 1066 and 1086 and holdings that were secular in 1066 and did not change landlord, hence that probably remained Anglo-Saxon throughout the period. Again, we see a positive and significant effect of Benedictine control, and this suggests that - in the main specification - we are not capturing just an "Anglo-Saxon effect", or the consequence of a better knowledge of the territory by Benedictine decision makers when compared with the newly arrived Norman conquerors. They were better than Anglo-Saxon landlords as well.[30]

Table 4.13 ensures that our effect is not due to the monks working the land directly. They may have different incentives, motivations and a different "hard work ethic" with respect to ordinary peasants so, since we want to isolate the effect of a better decision making structure, it makes sense to exclude places where they could work directly. In particular, Table 4.13 excludes all the holdings that are closer than 5km to a monastery, and the coefficients are basically unaffected.[31]

in our sample.

[30]Importantly, none of those holdings has been re-assigned, hence we do not need to worry about the effect coming from a replacement of the landlord.

[31]Results are very similar using 4km and 6km as thresholds.

Table 4.12: Holding value, Benedictines vs Anglo-Saxon landlords

	(1)	(2)	(3)	(4)	(5)	(6)
Benedictine	0.262∗∗	0.171	0.156∗	0.258∗∗∗	0.207∗	0.073
	(0.107)	(0.112)	(0.086)	(0.096)	(0.114)	(0.096)
Manor FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Time FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Geography*time ctrls[1]	*No*	*Yes*	*No*	*No*	*No*	*Yes*
Market*time ctrls[2]	*No*	*No*	*Yes*	*No*	*No*	*Yes*
History*time ctrls[3]	*No*	*No*	*No*	*Yes*	*No*	*Yes*
Holding*time ctrls[4]	*No*	*No*	*No*	*No*	*Yes*	*Yes*
Obs	2812	2812	2812	2812	2812	2812

Notes. Dependent variable: log of value. Standard errors clustered at lord level (1086). Shared holdings and outliers are excluded from the sample. We keep only holdings that where either always Benedictine or that remained in Anglo-Saxon hands after the Conquest.

[1] Geography controls include geographic/agriculture related features, namely: latitude, longitude, latitude×longitude, median altitude, crop suitability, ruggedness, pasture suitability, agricultural suitability, animal husbandry suitability;

[2] Market access controls include proxies for access to markets: (log of) distance (in km) to: rivers, coast, London, nearest borough;

[3] History controls include proxies for the ancientness of the settlement, namely: (log of) distance (in km) to Roman settlements and to Roman roads;

[4] Holding-specific controls include size, the number of holdings held by the same lord and holding dispersion.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.13: Holding value, Benedictines vs secular landlords (excluding holdings close to a monastery)

	(1)	(2)	(3)	(4)	(5)	(6)
Benedictine	0.103∗∗∗	0.050∗∗	0.057∗∗	0.102∗∗∗	0.101∗∗∗	0.040∗
	(0.029)	(0.023)	(0.022)	(0.029)	(0.030)	(0.022)
Manor FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Time FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Geography*time ctrls[1]	*No*	*Yes*	*No*	*No*	*No*	*Yes*
Market*time ctrls[2]	*No*	*No*	*Yes*	*No*	*No*	*Yes*
History*time ctrls[3]	*No*	*No*	*No*	*Yes*	*No*	*Yes*
Holding*time ctrls[4]	*No*	*No*	*No*	*No*	*Yes*	*Yes*
Obs	15182	15178	15182	15182	15182	15178

Notes. Dependent variable: log of value. Standard errors clustered at lord level (1086). Shared holdings and outliers are excluded from the sample. We keep only observations whose distance from a monastery is bigger than 5km.

[1] Geography controls include geographic/agriculture related features, namely: latitude, longitude, latitude×longitude, median altitude, crop suitability, ruggedness, pasture suitability, agricultural suitability, animal husbandry suitability;

[2] Market access controls include proxies for access to markets: (log of) distance (in km) to: rivers, coast, London, nearest borough;

[3] History controls include proxies for the ancientness of the settlement, namely: (log of) distance (in km) to Roman settlements and to Roman roads;

[4] Holding-specific controls include size, the number of holdings held by the same lord and holding dispersion.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.7 Conclusion

This paper suggests that a certain degree of democratic selection, even in absence of full accountability, positively impacts economic outcomes. We exploit the historical case of Benedictine monasteries in England at the turn of Norman conquest to show that they were more efficient landlords than their secular counterparts: Benedictine monasteries were abiding Benedict's Rule which prescribes, among other provisions, the election of the abbot. We obtain this result at a very disaggregated level, comparing a contemporaneous measure of productive capacity of medieval English holdings controlled by Benedictine monasteries and by secular landlords.

In order to address the selection concerns, we exploit the massive re-assignment of land possession that followed the Norman Conquest, focusing on estates that switched ownership from Anglo-Saxon secular lords to Norman secular lords or to monasteries. Benedictine-controlled holdings experienced, in the most conservative estimation, a 3.9 percentage points higher growth in productive capacity over the 20 years period 1066-1086.

In terms of mechanisms, we are able to rule out many alternative stories, showing - among others - that our effect is not due to geographic proximity to a monastery, that it does not depend on holdings where the monks could have worked directly and that religious institutions unaffected by Benedict's Rule (i.e. bishops) have an economic performance that is indistinguishable from that of secular landlords. In this way, we prove (indirect) evidence of the importance of the Rule and of its institutional features.

These results suggest ways for further research. First of all, having information about abbots' tenure and the number of monks in each monastery, one could measure freedom and competitiveness of the elections. Further, it could be interesting to look at long term outcomes, particularly before the suppression of monasteries ordered by Henry VIII (as the long term effect of this policy are already considered by Heldring et al. 2017).

While all these questions certainly deserve further research, our paper sets a first step in shedding new light on the economic effect of democratic selection, even when accountability cannot be fully guaranteed.

Appendices

A4 Additional tables

Cross-sectional regression

Table A4.1: Growth rate of holding value: Benedictine and secular landlords compared, full table displaying all coefficients

	(1)	(2)	(3)	(4)	(5)	(6)
Benedictine (1066)	0.057***	0.054**	0.057***	0.058***	0.039*	0.037*
	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)
Latitude		−0.095				−0.152*
		(0.075)				(0.087)
Longitude		−4.253*				−2.768
		(2.496)				(2.286)
Latitude*Longitude		0.082*				0.055
		(0.048)				(0.044)
Median altitude		−0.000				−0.000
		(0.000)				(0.000)
Crop suitability		−0.005				−0.003
		(0.004)				(0.004)
Ruggedness		−0.000				−0.000
		(0.000)				(0.000)
Agricultural suitability		−0.285*				−0.369**
		(0.159)				(0.164)
Pasture suitability		0.646***				0.674***
		(0.187)				(0.190)
Animal husbandry suitability		−0.468***				−0.410***
		(0.134)				(0.122)
Dist. rivers (log)			−0.010			−0.014
			(0.010)			(0.012)
Dist. coast (log)			−0.010			0.007
			(0.013)			(0.016)
Dist. London (log)			−0.023			0.134***
			(0.063)			(0.048)
Dist. borough (log)			0.017			0.012
			(0.018)			(0.017)
Dist. markets (log)			−0.049***			−0.046***
			(0.016)			(0.014)
Dist. Roman settl. (log)				−0.021***		−0.017**
				(0.008)		(0.008)
Dist. Roman roads (log)				0.007		0.009
				(0.007)		(0.007)
Holding's value, 1066 (log)					−0.232***	−0.234***
					(0.019)	(0.019)
Size (log)					0.233***	0.234***
					(0.018)	(0.018)
Holdings of the same lord in 1066 (log)					0.003	0.004
					(0.007)	(0.007)
Dispersion of holdings (1066)					0.001	−0.000
					(0.012)	(0.012)
Hundred FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Obs	9287	9285	9287	9287	9287	9285

Dependent variable: value growth rate. Standard errors clustered at lord level (1066). Shared holdings and outliers (i.e. 7 estates recording growth rate larger than 30) are excluded from the sample.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A4.2: Growth rate of holding value: Benedictine and secular landlords compared, county FE, full table

	(1)	(2)	(3)	(4)	(5)	(6)
Benedictine (1066)	0.091**	0.081**	0.090**	0.090***	0.065**	0.055*
	(0.036)	(0.034)	(0.037)	(0.035)	(0.030)	(0.028)
Included controls:						
Geography[1]	*No*	*Yes*	*No*	*No*	*No*	*Yes*
Market access[2]	*No*	*No*	*Yes*	*No*	*No*	*Yes*
History[3]	*No*	*No*	*No*	*Yes*	*No*	*Yes*
Holding specific[4]	*No*	*No*	*No*	*No*	*Yes*	*Yes*
County FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Adj. R-sq.	0.273	0.277	0.277	0.274	0.340	0.346
Obs	9288	9286	9288	9288	9288	9286

Dependent variable: value growth rate. Standard errors clustered at lord level (1066). Shared holdings and outliers (i.e. 7 estates recording growth rate larger than 30) are excluded from the sample.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A4.3: Growth rate of holding value: Benedict's rule and proximity to monastery compared

	(1) Benchmark	(2) Excl. Bened. (HFE)	(3) Excl. Bened. (CFE)
Dist. nearest monastery (log)		−0.012	−0.003
		(0.013)	(0.008)
Benedictine (1066)	0.037∗		
	(0.022)		
Latitude	−0.152∗	−0.149∗	−0.002
	(0.087)	(0.088)	(0.034)
Longitude	−2.768	−3.020	0.682
	(2.286)	(2.290)	(0.856)
Latitude*Longitude	0.055	0.059	−0.013
	(0.044)	(0.044)	(0.016)
Median altitude	−0.000	−0.000	−0.000 ∗∗
	(0.000)	(0.000)	(0.000)
Crop suitability	−0.003	−0.004	−0.000
	(0.004)	(0.004)	(0.004)
Ruggedness	−0.000	−0.000	−0.000
	(0.000)	(0.000)	(0.000)
Agricultural suitability	−0.369 ∗∗	−0.403 ∗∗	−0.343 ∗∗∗
	(0.164)	(0.174)	(0.107)
Pasture suitability	0.674 ∗∗∗	0.720 ∗∗∗	0.353 ∗∗
	(0.190)	(0.225)	(0.145)
Animal husbandry suitability	−0.410 ∗∗∗	−0.410 ∗∗∗	−0.166 ∗∗
	(0.122)	(0.129)	(0.072)
Dist. rivers (log)	−0.014	−0.015	−0.011
	(0.012)	(0.013)	(0.008)
Dist. coast (log)	0.007	0.009	0.016∗
	(0.016)	(0.016)	(0.009)
Dist. London (log)	0.134 ∗∗∗	0.121 ∗∗	−0.045 ∗∗
	(0.048)	(0.049)	(0.021)
Dist. borough (log)	0.012	0.026	0.019∗
	(0.017)	(0.019)	(0.010)
Dist. markets (log)	−0.046 ∗∗∗	−0.049 ∗∗∗	−0.033 ∗∗∗
	(0.014)	(0.015)	(0.009)
Dist. Roman settl. (log)	−0.017 ∗∗	−0.017 ∗∗	−0.021 ∗∗∗
	(0.008)	(0.008)	(0.008)
Dist. Roman roads (log)	0.009	0.014 ∗∗	0.008
	(0.007)	(0.007)	(0.005)
Holding's value, 1066 (log)	−0.234 ∗∗∗	−0.234 ∗∗∗	−0.217 ∗∗∗
	(0.019)	(0.020)	(0.018)
Size (log)	0.234 ∗∗∗	0.232 ∗∗∗	0.218 ∗∗∗
	(0.018)	(0.018)	(0.017)
Holdings of the same lord in 1066 (log)	0.004	0.005	0.006
	(0.007)	(0.008)	(0.007)
Dispersion of holdings (1066)	−0.000	−0.001	−0.003
	(0.012)	(0.012)	(0.013)
Obs	9285	8761	8762

Dependent variable: value growth rate. Standard errors clustered at lord level (1066). Shared holdings and outliers (i.e. 7 estates recording growth rate larger than 30) are excluded from the sample, as well as holdings held by Female and Non-Benedictine monastic orders.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Diff-in-diff: robustness checks

Table A4.4: Summary statistics (subsample)

	Mean	SD	Min	Max	Obs
Benedictine (1086)	0.022	0.148	0.000	1.000	7787
Outcome variables					
Holding's value, 1066 (log)	1.229	0.805	0.000	4.727	7787
Holding's value, 1086 (log)	1.180	0.858	0.000	4.987	7787
Geography					
Latitude	52.356	0.860	50.502	54.611	7787
Longitude	−0.611	1.157	−4.074	1.742	7787
Latitude*Longitude	−32.155	60.668	−205.757	91.468	7787
Median altitude	69.885	50.151	−1.000	404.000	7786
Crop suitability	4.065	1.381	0.000	7.615	7786
Ruggedness	14.926	19.989	0.000	208.646	7787
Pasture suitability	0.148	0.073	0.049	0.529	7785
Agricultural suitability	0.516	0.063	0.351	0.789	7785
Animal husbandry suitability	0.459	0.162	0.189	0.816	7785
Market access					
Dist. rivers (log)	3.042	0.979	0.006	5.209	7787
Dist. coast (log)	3.388	1.065	0.065	4.908	7787
Dist. London (log)	4.793	0.618	0.865	5.919	7787
Dist. borough (log)	3.339	0.712	0.000	4.800	7787
Dist. markets (log)	2.696	0.622	0.132	4.286	7787
History					
Dist. Roman settl. (log)	1.456	0.556	0.000	3.989	7787
Dist. Roman roads (log)	1.688	0.754	0.002	3.982	7787
Holding-specific					
Size (log)	1.215	0.715	0.012	4.646	7787
Holdings of the same lord in 1066 (log)	3.103	1.550	0.693	6.332	7787
Holdings of the same lord in 1086 (log)	3.190	1.612	0.693	7.141	7787
Holdings' dispersion (1066)	0.891	0.690	0.000	3.933	7787
Holdings' dispersion (1086)	0.844	0.642	0.000	3.348	7787

Table A4.5: Holding value: Benedictine and secular landlord (bishops excluded)

	(1)	(2)	(3)	(4)	(5)	(6)
Benedictine	0.099 * **	0.050 * *	0.055 * *	0.098 * **	0.098 * **	0.042*
	(0.030)	(0.022)	(0.023)	(0.030)	(0.032)	(0.022)
Manor FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Time FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Geography*time ctrls[1]	*No*	*Yes*	*No*	*No*	*No*	*Yes*
Market*time ctrls[2]	*No*	*No*	*Yes*	*No*	*No*	*Yes*
History*time ctrls[3]	*No*	*No*	*No*	*Yes*	*No*	*Yes*
Holding*time ctrls[4]	*No*	*No*	*No*	*No*	*Yes*	*Yes*
Obs	15092	15088	15092	15092	15092	15088

Dependent variable: log of value. Standard errors clustered at lord level (1086). Shared holdings and outliers are excluded from the sample, as well as holdings with an ecclesiastical non monastic landlord;

[1] Geography controls include geographic/agriculture related features, namely: latitude, longitude, latitude×longitude, median altitude, crop suitability, ruggedness, pasture suitability, agricultural suitability, animal husbandry suitability;

[2] Market access controls include proxies for access to markets: (log of) distance (in km) to: rivers, coast, London, nearest borough;

[3] History controls include proxies for the ancientness of the settlement, namely: (log of) distance (in km) to Roman settlements and to Roman roads;

[4] Holding-specific controls include size, the number of holdings held by the same lord and holding dispersion.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A4.6: Holding value: Benedictine vs other landlords (excluding the King)

	(1)	(2)	(3)	(4)	(5)	(6)
Benedictine	0.095 ∗ ∗∗	0.055 ∗ ∗	0.061 ∗ ∗∗	0.093 ∗ ∗∗	0.098 ∗ ∗∗	0.051 ∗ ∗
	(0.030)	(0.022)	(0.023)	(0.030)	(0.030)	(0.021)
Manor FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Time FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Geography*time ctrls[1]	*No*	*Yes*	*No*	*No*	*No*	*Yes*
Market*time ctrls[2]	*No*	*No*	*Yes*	*No*	*No*	*Yes*
History*time ctrls[3]	*No*	*No*	*No*	*Yes*	*No*	*Yes*
Holding*time ctrls[4]	*No*	*No*	*No*	*No*	*Yes*	*Yes*
Obs	14572	14568	14572	14572	14572	14568

Dependent variable: log of value. Standard errors clustered at lord level (1086). Shared holdings and outliers are excluded from the sample, as well as holdings with the King as a landlord in 1086;

[1] Geography controls include geographic/agriculture related features, namely: latitude, longitude, latitude×longitude, median altitude, crop suitability, ruggedness, pasture suitability, agricultural suitability, animal husbandry suitability;

[2] Market access controls include proxies for access to markets: (log of) distance (in km) to: rivers, coast, London, nearest borough;

[3] History controls include proxies for the ancientness of the settlement, namely: (log of) distance (in km) to Roman settlements and to Roman roads;

[4] Holding-specific controls include size, the number of holdings held by the same lord and holding dispersion.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A4.7: Diff-in-Diff: same sample

	(1)	(2)	(3)	(4)	(5)	(6)
Benedictine	0.098 ∗ ∗∗	0.048 ∗ ∗	0.055 ∗ ∗	0.096 ∗ ∗∗	0.096 ∗ ∗∗	0.039∗
	(0.030)	(0.022)	(0.023)	(0.030)	(0.032)	(0.022)
Manor FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Time FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Geography*time ctrls[1]	*No*	*Yes*	*No*	*No*	*No*	*Yes*
Market*time ctrls[2]	*No*	*No*	*Yes*	*No*	*No*	*Yes*
History*time ctrls[3]	*No*	*No*	*No*	*Yes*	*No*	*Yes*
Holding*time ctrls[4]	*No*	*No*	*No*	*No*	*Yes*	*Yes*
Obs	15570	15570	15570	15570	15570	15570

Dependent variable: log of value. Standard errors clustered at Lord86 level. The sample of Column 6 is usef for all the regressions. Shared holdings and outliers are excluded from the sample.

[1] Geography controls include geographic/agriculture related features, namely: latitude, longitude, latitude×longitude, median altitude, crop suitability, ruggedness, pasture suitability, agricultural suitability, animal husbandry suitability;

[2] Market access controls include proxies for access to markets: (log of) distance (in km) to: rivers, coast, London, nearest borough;

[3] History controls include proxies for the ancientness of the settlement, namely: (log of) distance (in km) to Roman settlements and to Roman roads;

[4] Holding-specific controls include size, the number of holdings held by the same lord and holding dispersion.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Diff-in-diff: different clustering strategies

Table A4.8: Diff-in-Diff: Conley standard errors

	(1)	(2)	(3)	(4)	(5)	(6)
Benedictine	0.098***	0.048**	0.055***	0.096***	0.096***	0.039*
	(0.030)	(0.021)	(0.019)	(0.029)	(0.029)	(0.021)
Manor FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Time FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Geography*time ctrls[1]	*No*	*Yes*	*No*	*No*	*No*	*Yes*
Market*time ctrls[2]	*No*	*No*	*Yes*	*No*	*No*	*Yes*
History*time ctrls[3]	*No*	*No*	*No*	*Yes*	*No*	*Yes*
Holding*time ctrls[4]	*No*	*No*	*No*	*No*	*Yes*	*Yes*
Obs	15574	15570	15574	15574	15574	15570

Dependent variable: log of value. Standard errors calculated using Spatial HAC method. Shared holdings and outliers are excluded from the sample.

[1] Geography controls include geographic/agriculture related features, namely: latitude, longitude, latitude×longitude, median altitude, crop suitability, ruggedness, pasture suitability, agricultural suitability, animal husbandry suitability;

[2] Market access controls include proxies for access to markets: (log of) distance (in km) to: rivers, coast, London, nearest borough;

[3] History controls include proxies for the ancientness of the settlement, namely: (log of) distance (in km) to Roman settlements and to Roman roads;

[4] Holding-specific controls include size, the number of holdings held by the same lord and holding dispersion.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A4.9: Diff-in-Diff: clustering at Hundred level

	(1)	(2)	(3)	(4)	(5)	(6)
Benedictine	0.098***	0.048*	0.055**	0.096***	0.096***	0.039
	(0.029)	(0.026)	(0.025)	(0.028)	(0.029)	(0.025)
Manor FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Time FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Geography*time ctrls[1]	*No*	*Yes*	*No*	*No*	*No*	*Yes*
Market*time ctrls[2]	*No*	*No*	*Yes*	*No*	*No*	*Yes*
History*time ctrls[3]	*No*	*No*	*No*	*Yes*	*No*	*Yes*
Holding*time ctrls[4]	*No*	*No*	*No*	*No*	*Yes*	*Yes*
Obs	15574	15570	15574	15574	15574	15570

Dependent variable: log of value. Standard errors clustered at hundred level. Shared holdings and outliers are excluded from the sample.

[1] Geography controls include geographic/agriculture related features, namely: latitude, longitude, latitude×longitude, median altitude, crop suitability, ruggedness, pasture suitability, agricultural suitability, animal husbandry suitability;

[2] Market access controls include proxies for access to markets: (log of) distance (in km) to: rivers, coast, London, nearest borough;

[3] History controls include proxies for the ancientness of the settlement, namely: (log of) distance (in km) to Roman settlements and to Roman roads;

[4] Holding-specific controls include size, the number of holdings held by the same lord and holding dispersion.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A4.10: Diff-in-Diff: clustering at County level

	(1)	(2)	(3)	(4)	(5)	(6)
Benedictine	0.098*	0.048*	0.055 **	0.096*	0.096*	0.039*
	(0.055)	(0.027)	(0.025)	(0.053)	(0.055)	(0.023)
Manor FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Time FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Geography*time ctrls[1]	*No*	*Yes*	*No*	*No*	*No*	*Yes*
Market*time ctrls[2]	*No*	*No*	*Yes*	*No*	*No*	*Yes*
History*time ctrls[3]	*No*	*No*	*No*	*Yes*	*No*	*Yes*
Holding*time ctrls[4]	*No*	*No*	*No*	*No*	*Yes*	*Yes*
Obs	15574	15570	15574	15574	15574	15570

Dependent variable: log of value. Standard errors clustered at county level. Shared holdings and outliers are excluded from the sample.

[1] Geography controls include geographic/agriculture related features, namely: latitude, longitude, latitude×longitude, median altitude, crop suitability, ruggedness, pasture suitability, agricultural suitability, animal husbandry suitability;

[2] Market access controls include proxies for access to markets: (log of) distance (in km) to: rivers, coast, London, nearest borough;

[3] History controls include proxies for the ancientness of the settlement, namely: (log of) distance (in km) to Roman settlements and to Roman roads;

[4] Holding-specific controls include size, the number of holdings held by the same lord and holding dispersion.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Bibliography

Acemoglu, D., Johnson, S. and Robinson, J. (2005) The rise of europe: Atlantic trade, institutional change, and economic growth. *American economic review*, 95(3): 546–579.

Acemoglu, D., Naidu, S., Restrepo, P. and Robinson, J. A. (2019) Democracy does cause growth. *Journal of Political Economy*, 127(1): 47–100.

Acemoglu, D., Reed, T. and Robinson, J. A. (2014) Chiefs: Economic development and elite control of civil society in Sierra Leone. *Journal of Political Economy*, 122(2): 319–368.

Aghamolla, C. (2016) *Essay on Analyst Herding.* Ph. D. thesis, Columbia University.

Alesina, A. and Passarelli, F. (2019) Loss Aversion in Politics. *American Journal of Political Science*, forthcoming.

Alesina, A. and Tabellini, G. (2007) Bureaucrats or politicians? Part I: a single policy task. *American Economic Review*, 97(1): 169–179.

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

Allcott, H., Braghieri, L., Eichmeyer, S. and Gentzkow, M. (2019) The Welfare Effects of Social Media. *Working paper.*

Allen, M., Blick, N., Brindle, T., Evans, T., Fulford, M., Holbrook, N., Lodwick, L., Richards, J. D. and Smith, A. (2018) *The rural settlement of roman britain: an online resource* [data-set].

Andersen, T. B., Bentzen, J., Dalgaard, C.-J. and Sharp, P. (2017) Pre-reformation roots of the protestant ethic. *The Economic Journal*, 127(604): 1756–1793.

Anderson, S. P. and McLaren, J. (2012) Media Merges and Media Bias with Rational Consumers. *Journal of the European Economic Association*, 10(4): 831-859.

Andina-Diaz, A. and Garcia-Martinez, J. E. (2016) Reputation and news suppression in the media industry. *working paper*.

Andreottola, G. (2018) Flip-flopping and Electoral Concerns. *Working paper*.

Andreottola, G. and De Moragas, A. I. (2017) Fake news, media competition and political accountability. *Working paper*.

Angelucci, C. and Cagé, J. (2019) Newspapers in times of low advertising revenues. *American Economic Journal: Microeconomics*, 11(3): 319-364.

Angelucci, C., Meraglia, S. and Voigtländer, N. (2017) The medieval roots of inclusive institutions: From the norman conquest of england to the great reform act. *National Bureau of Economic Research*.

Argentesi, E. and Ivaldi, M. (2005) Market definition in printed media industry: theory, practice and lessons for broadcasting. *CEPR Working Paper*.

Armstrong, M. (2005) Public service broadcasting. *Fiscal Studies*, 26(3): 281-299.

Ashworth, S. and Bueno de Mesquita, E. (2014) Is Voter Competence Good for Voters? Information, Rationality, and Democratic Performance. *The American Political Science Review*, 108(3):565-587.

Ashworth, S. and Shotts, K. (2010) Does informative media commentary reduce politicians' incentives to pander? *Journal of Public Economics*, 94(11-12): 838-847.

Ashworth, S., Bueno de Mesquita, E. and Friedenberg, A. (2017) Accountability and information in elections. *American Economic Journal: Microeconomics*, 9(2): 95–138.

Aston, T. H. (1958) The origins of the manor in England 1. *Transactions of the Royal Historical Society*, 8: 59–83.

Ayton, A. and Davis, V. (1987) Ecclesiastical wealth in England in 1086. *Studies in Church History*, 24: 47–60.

Barberá, P. (2018) The Consequences of Exposure to Disinformation and Propaganda in Online Settings. In Tucker, J. A. et al. (2018) *Social Media, Political Polarization and Political Disinformation*, Hewlett Foundation.

Barlow, F. (1979) *The English Church 1066-1154*. Longman.

Baron, D. (2006) Persistent Media Bias. *Journal of Public Economics*, 90: 1-36.

Barone, G., D'Acunto, F. and Narciso, G. (2015) Telecracy: Testing for Channels of Persuasion. *American Economic Journal: Economic Policy*, 7(2): 30-60.

Barrera, O., Guriev, S., Henry, E. and Zhuravskaya, E. (2017) *Facts, alternative facts and fact checking in times of post-truth politics*. Centre for Economic Policy Research.

Barro, R. J. and Lee, J. W. (2013) A new data set of educational attainment in the world, 1950-2010, *Journal of Development Economics*, 104, 184-198.

Beattie, G., Durante, R., Knight, B. and Sen, A. (2017) Advertising Spending and Media Bias: Evidence from News Coverage of Car Safety Recalls. *NBER working paper 23940*.

Beck, J. and Sieber, A. (2010) Is the spatial distribution of mankind's most basic economic traits determined by climate and soil alone? *PloS one*, 5(5): e10416.

Becker, S. O., Boeckh, K., Hainz, C. and Woessmann, L. (2015) The empire is dead, long live the empire! Long-run persistence of trust and corruption in the bureaucracy. *The Economic Journal*, 126(590): 40–74.

Bernhardt, D., Krasa, S. and Polborn, M. (2008) Political Polarization and the Electoral Effects of Media Bias. *Journal of Public Economics*, 98(5-6): 1092-1104.

Besley, T. (2006) *Principled Agents? The Political Economy of Good Government*. New York: Oxford University Press.

Besley, T. and Burgess, R. (2002) The Political Economy of Government Responsiveness. Theory and Evidence from India. *Quarterly Journal of Economics*, 117(4): 1415-1451.

Besley, T. and Prat, A. (2006) Handcuffs for the Grabbing Hand? The Role of the Media in Political Accountability. *American Economic Review*, 96(3): 720-736.

Blasco, A., Pin, P. and Sobbrio, F. (2015) Paying Positive to Go Negative: Advertisers' Competition and Media Reports. *Working paper.*

Bobtcheff, C., Bolte, J. and Mariotti, T. (2016) Researchers dilemma. *The Review of Economic Studies*, 84(3): 969-1014.

Borusyak, K. and Jaravel, X. (2016) Revisiting Event Study Designs. *Working paper.*

Burgess, R., Jedwab, R., Miguel, E., Morjaria, A., and Padró i Miquel, G. (2015) The value of democracy: evidence from road building in Kenya. *American Economic Review*, 105(6): 1817–1851.

Cagé, J. (2014) Media Competition, Information Provision and Political Participation. *Working paper.*

Cairncross, F. (2019) *The Cairncross review. A sustainable future for journalism.* Department for Digital, Culture, Media & Sport, HM Government of UK.

Campbell, B. M. S. (1983) Arable Productivity in Medieval England: Some Evidence from Norfolk. *Journal of Economic History*, 43(2): 379–404.

Campbell, B. M. S. (2006) *English Seignioral Agriculture, 1250-1450.* Cambridge University Press.

Canes-Wrone, B., Herron, M. C. and Shotts, K. W. (2001) Leadership and Pandering: A Theory of Executive Policymaking. *American Journal of Political Science*, 45(3): 532-550.

Card, D. (1992) Using Regional Variation in Wages to Measure the Effects of the Federal Minimum Wage. *Industrial and Labor Relations Review*, 46(1): 22-37.

Carson, A. (2019) *Investigative Journalism, Democracy and the Digital Age.* Routledge.

Chan, Y. (2014) Journalism and digital times: Between wider reach and sloppy reporting. In Dragomir, M. and Thompson, M. (Eds.), *Digital Journalism. Making News, Breaking News.* Open Society Foundation.

Chan, J. and Stone, D. F. (2013) Media proliferation and partisan selective exposure. *Public Choice*, 156: 467-490.

Chen, H. and Suen, W. (2016) Competition for attention in the news media market. *Working paper.*

Corneo, G. (2006) Media Capture in a Democracy: The Role of Wealth Concentration. *Journal of Public Economics*, 90(1-2): 37-58.

Darby, H. C. (1977) *Domesday England.* Cambridge University Press.

D'Avray, D. (2015) *English monastic archive.*

Dell'Omo, M. (2011) *Storia del monachesimo occidentale dal medioevo all'età contemporanea: il carisma di San Benedetto tra VI e XX secolo.* Milano: Jaca Book.

Di Tella, R. and Franceschelli, I. (2011) Government Advertising and Media Coverage of Corruption Scandals. *American Economic Journal: Applied Economics*, 3(4): 119-51.

Dobkin, C., Nicosia, N. and Weinberg, M. (2014) Are supply-side drug control efforts effective? Evaluating OTC regulations targeting methamphetamine precursors. *Journal of Public Economics*, 120: 48-61.

Drago, F., Nannicini, T. and Sobbrio, F. (2014) Meet the Press: How Voters and Politicians Respond to Newspapers Entry and Exit. *American Economic Journal: Applied Economics*, 6(3): 159–188.

Driver, A. (2018) Trump's Mexico wall would be a gift to the drug cartels. *Cnn.com.* [Accessed 01/08/2018].

Drufuca, S. M. (2014) Information, Media and Elections: Incentives for Media Capture. *working paper.*

Duggan, J. and Martinelli, C. (2011) A Spatial Theory of Media Slant and Voter Choice. *Review of Economic Studies*, 78(2): 640-666.

Dukes, A. (2006) Media Concentration and Consumer Product Prices. *Economic Inquiry*, 44(1): 128-141.

Durante, R. and Knight, B. (2012) Partisan Control, Media Bias and Viewer Responses: Evidence from Berlusconi's Italy. *Journal of the European Economic Association*, 10(3): 451-481.

Ehrmann, T., Rost, K. and Inauen, E. (2013) Location of decision rights in catholic religious orders. *Managerial and Decision Economics*, 34: 29–39.

Ekelund, R. B., Ebert, R. F. H., Tollison, R. D., Anderson, G. M. and Davidson, A. B. (1996) *Sacred Trust: The Medieval Church as an Economic Firm.* Oxford University Press.

Ellman, M. and Germano, F. (2009) What do the Papers Sell? A Model of Advertising and Media Bias. *The Economic Journal*, 119: 680-704.

Enikolopov, R. and Petrova, M. (2016) Media Capture: Empirical Evidence. In *Handbook of Media Economics*, S. Andersen, D. Strömberg, J. Waldfogel (eds.), Elsevier.

Enikolopov, R., Petrova, M. and Zhuravskaya, E. (2011) Media and Political Persuasion: Evidence from Russia. *American Economic Review*, 101(7): 3253-85.

European Audiovisual Observatory (1998-2015) *Yearbook. Television, cinema, video and on-demand audiovisual services in Europe*, 1998-2015 Editions, Strasbourg.

Ferraz, C. and Finan, F. (2008) Exposing corrupt politicians: the effects of Brazil's publicly released audits on electoral outcomes. *Quarterly Journal of Economics*, 123(2): 703-745.

Fetzer, T. and Kyburz S. (2018) Cohesive institutions and political violence. *Empirical Studies of Conflict Project.*

Finn, R. W. (1963). *An Introduction to Domesday Book.* Barnes and Noble.

Fleming, R. (1991). *Kings and Lords in Conquest England.* Cambridge University Press.

Fox, J. (2007) Government transparency and policymaking. *Public Choice*, 131(1/2): 23-44.

Freedom House (n.d.) Freedom of the Press database, `https://freedomhouse.org/report-types/freedom-press`.

Freedom House (n.d.) Freedom in the World database, `https://freedomhouse.org/report-types/freedom-world`.

Frisell, R. (2009) A theory of self-fulfilling political expectations. *Journal of Public Economics*, 93: 715-720.

Fudenberg, D., Gilbert, R., Stiglitz, J. and Tirole, J. (1983) Preemption, leapfrogging and competition in patent races. *European Economic Review*, 22(1): 3-31.

Fudenberg, D. and Tirole, J. (1985) Preemption and rent equalization in the adoption of new technology. *The Review of Economic Studies*, 52(3): 383-401.

Galvis, A. F., Snyder Jr., J. M. and Song, B. K. (2016) Newspaper Market Structure and Behavior: Partisan Coverage of Political Scandals in the United States from 1870 to 1910. *The Journal of Politics*, 78(2): 368-381.

Gehlbach, S. and Sonin, K. (2014) Government control of the media. *Journal of Public Economics*, 118, 163-171.

Gentzkow, M. (2014) Trading dollars for dollars: The price of attention online and offline. *American Economic Review: Papers & Proceedings*, 104(5): 481-488.

Gentzkow, M. and Shapiro, J. M. (2006) Media Bias and Reputation. *Journal of Political Economy*, 114(2): 280-316.

Gentzkow, M., and Shapiro, J. M. (2008) Competition and Truth in the Market for News. *Journal of Economic Perspectives*, 22(2): 133-154.

Gentzkow, M., Shapiro, J.M. and Stone, D.F. (2016) Media Bias in the Marketplace: Theory. In *Handbook of Media Economics*, S. Andersen, D. Strömberg, J. Waldfogel (eds.), Elsevier.

Germano, F. and Meier, M. (2013) Concentration and self-censorship in commercial media. *Journal of Public Economics*, 97: 117-130.

Gerring, J., P. Bond, W. T. Barndt, and C. Moreno (2005). Democracy and economic growth: A historical perspective. *World Politics* 57(3), 323–364.

Glaeser, E. L. and Ponzetto, G. A. M., (2017) Fundamental Errors in the Voting Booth. *working paper.*

Godes, D., Ofek, E. and Sarvary, M. (2009) Content vs. Advertising: The Impact of Competition on Media Firm Strategy. *Marketing Strategy*, 28(1): 20-35.

Gratton, G., Holden, R. and Kolotilin, A. (2017) When to drop a bombshell. *The Review of Economic Studies*, 85(4): 2139-2172.

Guttman, I. (2010) The timing of analysts' earnings forecasts. *The Accounting Review*, 85(2): 513-545.

Guttman, I., Kremer, I. and Skrzypacz, A. (2014) Not only what but also when: A theory of dynamic voluntary disclosure. *American Economic Review*, 104(8): 2400-2420.

Hafer, C. Landa, D. and Le Bihan, P. (2017) Who owns the story? News-gathering competition and political Accountability. *working paper.*

Hafer, C., Landa, D. and Le Bihan, P. (2018) Who gets the credit? News-gathering competition and political accountability. *Working paper.*

Hafer, C., Landa, D. and Le Bihan, P. (2019) Breaking news: Media competition and political polarization. *Working paper.*

Hamilton, J. T. (2004) *All the News That's Fit To Sell.* Princeton, Princeton University Press.

Heldring, L., Robinson, J. A. and Vollmer, S. (2017) The long-run impact of the dissolution of the English monasteries. *Working paper.*

Hollifield, C. A. (2006) News Media Performance in Hypercompetitive Markets: An Extended Model of Effects. *International Journal on Media Management*, 8: 60-70.

Hopenhayn, H. A. and Squintani, F. (2011) Preemption games with private information. *The Review of Economic Studies*, 78(2): 667-692.

Hopenhayn, H. A. and Squintani, F. (2015) Patent rights and innovation disclosure. *The Review of Economic Studies*, 83(1): 199-230.

IIASA/FAO (2012) *Global agro-ecological zones* (gaez v3.0).

Iloska, A. (2014) *FYR Macedonia: A large decrease in press freedom since 2008 – prevailing economic interests and increasing politicisation of the media scene.* Available from: `http://www.fairpress.eu/blog/2014/07/07/macedonia-a-large-decrease-in-press-freedom-since-2008-prevailing-economic-interests-and-increasing-politicisation-of-the-media-scene/`. [Accessed: 29th May 2015].

Institute for Democracy and Electoral Assistance (n.d.) database, `http://www.idea.int/vt/viewdata.cfm`.

International Foundation for Electoral System (n.d.), database, `electionguide.org`.

IREX (2015) *Media Sustainability Index (MSI) study for Europe and Eurasia. Executive Summary.* Availabe form: `https://www.irex.org/projects/media-sustainability-index-msi` [Accessed: 31st July 2015].

Johnson, J. W. and Wallack, J. S. (2010) *ELECTORAL SYSTEMS AND THE PERSONAL VOTE. Update of database from "Particularism Around the World"*, `http://dss.ucsd.edu/~jwjohnso/espv.html`.

Jones, A. (1972) The rise and fall of the manorial system: a critical comment. *The Journal of Economic History*, 32(4): 938–944.

Judson, H. P. (1898) The Political Effects of the Teaching of Jesus. *The Biblical World*, 11(4): 229-238.

Kenny, R., Foster, R. and Suter, T. (2014) *The value of Digital terrestrial Television in an era of increasing demand for spectrum.* Digital UK report.

Kibris, O. and Kocak, K. (2017) The Effect of Social Media on Media Capture and Government Accountability. *working paper.*

Knobel, B. (2018) *The watchdog still barks: How accountability reporting evolved for the digital age.* Fordham Univ Press.

Knowles, D. (1966). *The Monastic Order in England. A History of Its Development from the Times of St. Dunstan to the Fourth Lateran Council, 940-1216.* Cambridge University Press.

Kosminsky, E. A. (1961). *Studies in the Agrarian History of England in the Thirteenth Century.* Oxford: Basil Blackwell.

Lapidge, M. (1995) *Archbishop Theodore: Commemorative Studies on his Life and Influence.* Cambridge University Press.

Laussel, D. and Van Long, N. (2020) Tying the politicians' hands: The optimal limits to representative democracy. *Jorunal of Public Economic Theory*, 22(1): 25-48.

Lee, A. M.-C. (2014) *How fast is too fast?: examining the impact of speed-driven journalism on news production and audience reception.* Ph. D. thesis, University of Texas at Austin.

Letters, S. (2005) *Gazzetteer of Markets and Fairs in England and Wales to 1516*, `http://www.british-history.ac.uk/list-index-soc/markets-fairs-gazetteer-to-1516`.

Levy, G. and Razin, R. (2015) Correlation Neglect, Voting Behavior, and Information Aggregation. *The American Economic Review*, 105(4): 1634-1645.

Levy, G., Morendo de Barrida, I. and Razin, R. (2017) Persuasion and Correlation Neglect. *working paper.*

Lewis, J., Cushion, S. and Thomas, J. (2005) Immediacy, Convenience or Engagement? An analysis of 24-hour news channels in the UK. *Journalism Studies*, 6(4), 461-477.

Lockwood, B. (2016) Confirmation Bias and Electoral Accountability. *Quarterly Journal of Political Science*, 11(4): 471-501.

Lockwood, B. and Rockey, J. (2018) Negative Voters? Electoral Competition with Loss-Aversion. *Working paper.*

Martinez-Bravo, M., i Miquel, G. P., Qian, N. and Yao, Y. (2014) Political reform in China: the effect of local elections. *NBER working paper* 18101.

Martinez-Bravo, M., Mukherjee, P. and Stegmann, A. (2017) The non-democratic roots of elite capture: Evidence from soeharto mayors in indonesia. *Econometrica*, 85(6): 1991–2010.

Maskin, E. and Tirole, J. (2004) The Politician and the Judge: Accountability in Government. *The American Economic Review*, 94(4): 1034-1054.

Matejka, F. and Tabellini, G. (2017) Electoral Competition with Rationally Inattentive Voters. *Working paper.*

McCannon, B. C. and Pruitt, J. (2018) Taking on the boss: Informative contests in prosecutor elections. *Jorunal of Public Economic Theory*, 20(5): 657-671.

McCormick, M., Huang, G., Gibson, K. (2010) *The digital atlas of roman and medieval civilizations* (darmc).

McDonald, J. (1997) Manorial Efficiency in Domesday England. *Journal of Productivity Analysis*, 8: 199–213.

McDonald, J. and Snooks, G. (1987) The suitability of domesday book for cliometric analysis. *The Economic History Review*, 40(2): 252–261.

McDonald, J. and Snooks, G. (1986) *Domesday economy: a new approach to Anglo-Norman history.* Oxford University Press.

McMillan, J. and Zoido, P. (2004) How to Subvert Democracy: Montesinos in Perù. *Journal of Economic Perspectives*, 18(4): 69-92.

Morelli, M. and Van Weelden, R. (2013) Ideology and Information in Policymaking. *Journal of Theoretical Politics*, 25(3): 412-439.

Mosquera, R., Odunowo, M., McNamara, T., Guo, X. and Petrie, R. (2019) The Economic Effects of Facebook. *Working paper.*

Moulin, L. (2016) *La vita quotidiana secondo san Benedetto.* Jaca Book.

Mullainathan, S. and Shleifer, A. (2005) The Market for News. *American Economic Review*, 95(4): 1031-1053.

North, D. C. and Thomas, R. P. (1971) The rise and fall of the manorial system: A theoretical model. *The Journal of Economic History*, 31(4): 777–803.

North, D. C. and Thomas, R. P. (1973) *The rise of the western world: A new economic history.* Cambridge University Press.

Nunn, N. and Puga, D. (2012) Ruggedness: The blessing of bad geography in africa. *Review of Economics and Statistics*, 94(1): 20–36.

Ortoleva, P. and Snowberg, E. (2015) Overconfidence in Political Behavior. *The American Economic Review*, 105(2): 504-535.

Palmer, J. (1987) The Domesday Manor. In Holt, J. C. (Eds.), *Domesday Studies.* Woodbridge, Suffolk: The Boydell Press.

Palmer, J. (2008) *Domesday statistics database.*

Perego, Y. and Yuksel, S. (2018) Media competition and social disagreement. *Working paper.*

Persico, N. (2004) Committee design with endogenous information. *The Review of Economic Studies*, 71(1): 165–191.

Petrova, M. (2008) Inequality and Media Capture. *Journal of Public Economics*, 92(1-2): 183-212.

Petrova, M. (2011) Newspapers and Parties: How Advertising Revenues Created an Independent Press. *American Political Science Review*, 105(4): 790-808.

Pew Research Center (2016) 2016 *Campaign: Strong Interest, Widespread Dissatisfaction.*

Polo, M. (2005) Regulation for Pluralism in the Media Markets. *Working paper.*

Postan, M. M. (1973) *The medieval economy and society: an economic history of Britain, 1100-1500, Volume 1.* University of California Press.

Prat, A. (2005) The Wrong Kind of Transparency. *The American Economic Review*, 95(3): 862-877.

Prat, A. (2014) Media Power. *CEPR Discussion Paper No.* 10094.

Prat, A. (2016) Media Capture and Media Power. In *Handbook of Media Economics*, S. Andersen, D. Strömberg, J. Waldfogel (eds.), Elsevier.

Prat, A. and Strömberg, D. (2013) The Political Economy of Mass Media. *Advances in economics and econometrics*, 2.

Prato, C. and Wolton, S. (2016) The Voters' Curses: Why We Need Goldilocks Voters. *American Journal of Political Science*, 60(3): 726-737.

Prato, C. and Wolton, S. (2018) Rational Ignorance, Populism, and Reform. *European Journal of Political Economy*, 55; 119-135.

Puglisi, R. and Snyder, J.M. (2016) Empirical Studies of Media Bias. In *Handbook of Media Economics*, S. Andersen, D. Strömberg, J. Waldfogel (eds.), Elsevier.

Rochet, J.-C. and Tirole, J. (2003) Platform competition in two-sided markets. *Journal of the European Economic Association*, 1(4): 990-1029.

Rochet, J.-C. and Tirole, J. (2006) Two-sided markets: a progress report. *The RAND Journal of Economics*, 37(3): 645-667.

Roffe, D. (2000) *Domesday: the Inquest and the Book.* New York, USA: Oxford University Press.

Roffe, D. (2007) *Decoding Domesday.* Woodbridge, USA: Boydell & Brewer Ltd.

Rosenberg, H. and Feldman, C. S. (2008) *No time to think: The menace of media speed and the 24-hour news cycle.* A&C Black.

Rost, K. (2017) Introduction to the corporate governance of religion. In Scott, R. A. and Buchmann, M. (Eds.), *Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource.* Hoboken.

Rost, K. and Graetzer, G. (2013) Multinational organizations as rule-following bureaucracies — the example of catholic orders. *Journal of International Management*, 20: 290–311.

Seabright, P. and von Hagen, J. (2007) *The Economic Regulation of Broadcasting Markets: Evolving Technology and Challenges for Policy.* Cambridge, Cambridge University Press.

Shapiro, J. M. (2016) Special interests and the media: Theory and an application to climate change. *Journal of public economics*, 144: 91-108.

Sims, C. A. (2003) Implications of rational inattention. *Journal of Monetary Economics*, 50(3): 665-690.

Snyder, J. and Strömberg, D. (2010) Press Coverage and Political Accountability. *Journal of Political Economy*, 118(2): 355-408.

Sobbrio, F. (2014) Citizen-editors' endogenous information acquisition and news accuracy. *Journal of Public Economics*, 113: 43-53.

Starbird, K., Dailey, D., Mohamed, O., Lee, G. and Spiro, E. S. (2018) Engage early, correct more: How journalists participate in false rumors online during crisis events. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.*

Strömberg, D. (2004) Mass Media Competition, Political Competition, and Public Policy. *Review of Economic Studies*, 71(1): 265-284.

Strömberg, D. (2015) Media and Politics. *Annual Review of Economics.* 7: 173-205.

Swanson, R. N. (1979) *Church & Society in Late Medieval England.* Blackwell.

Szeidl, A. and and Szucs, F. (2017) Media Capture through Favor Exchange. *Working paper.*

Talbert, R. J. (2000) *Barrington Atlas of the Greek and Roman World: Map-by-map Directory, Volume 1.* Princeton University Press.

Teorell, J., Dahlberg, S., Holmberg, S., Rothstein, B., Hartmann, F. and Svensson, R. (2015) *The Quality of Government Standard Dataset, version Jan15*, University of Gothenburg: The Quality of Government Institute, `http://www.qog.pol.gu.se`.

Thomas, H. M. (2008) *The Norman Conquest: England after William the Conqueror*. Rowman Littlefield.

Tirole, J. (1986) Hierarchies and Bureaucracies: On the Role of Collusion in Organizations. *Journal of Law, Economics and Organization*, 2(2), 181-214.

Torrens, G. (2014) Optimal Regulation of the Media Industry. *Working paper.*

Trombetta, F. (2020) When the light sines too much: Rational inattention and pandering. *Jorunal of Public Economic Theory*, 22(1): 98-145.

UNESCO Institute for Statistics, database, `http://data.uis.unesco.org/`.

Vaccari, F. (2018) Direct competition in the news market. *Working paper.*

Vaidya, S. and Gupta, R. (2016) Corruption via Media Capture: the Effect of Competition. *Southern Economic Journal*, 82(4): 1327-1348.

Wareham, A. (2000) The feudal revolution in eleventh-century east anglia. *Anglo-Norman Studies*, 22: 293–321.

Wihbey, J. (2015) How does social media use influence political participation and civic engagement? A meta-analysis. *journalistresource.org.* [Accessed 28/08/2017].

Wolfers, J. (2006) Did Unilateral Divorce Laws Raise Divorce Rates? A Reconciliation and New Results. *American Economic Review*, 96(5): 1802-1820.

Wolton, S. (2019) Are Biased Media Bad for Democracy? *American Journal of Political Science*, forthcoming.

worldaudit.org (2015) *Press freedom survey on 186 countries.*
Available from: `http://www.worldaudit.org/presstable.html`. [Accessed: 29th May 2015].

Zibak, F. I. (2015) Diversity not viewed as a measure to assess media freedom in Turkey. *Today's Zaman*, 20th April. Available from: `http://www.todayszaman.com`. [Accessed: 29th May 2015].