European University Institute

DEPARTMENT OF ECONOMICS

Essays in Political Economy

Giovanni Andreottola

Thesis submitted for assessment with a view to obtaining the degree of Doctor of Economics of the European University Institute

Florence, 25 October, 2017

European University Institute
Department of Economics

Essays in Political Economy

Thesis submitted for assessment with a view to obtaining the degree of Doctor of Economics of the European University Institute

Examining Board

Prof. Andrea Mattozzi, EUI, Supervisor
Prof. David K. Levine, EUI
Prof. Ronny Razin, London School of Economics
Prof. Alessandro Riboni, Ecole Polytechnique

European University Institute

DEPARTMENT OF ECONOMICS

Researcher declaration to accompany the submission of written work

I, **Giovanni Andreottola** certify that I am the author of the work **Essays in Political Economy** I have presented for examination for the PhD thesis at the European University Institute. I also certify that this is solely my own original work, other than where I have clearly indicated, in this declaration and in the thesis, that it is the work of others.

I warrant that I have obtained all the permissions required for using any material from other copyrighted publications.

I certify that this work complies with the *Code of Ethics in Academic Research* issued by the European University Institute (IUE 332/2/10 (CA 297).

Statement of inclusion of previous work (if applicable):

I confirm that **Chapter 3** was jointly co-authored with **Mr. Antoni-Italo de Moragas** and I contributed **50%** of the work.

Signature and Date:

Napoli, October 11, 2017

Giovanni Andreottola

Acknowledgements

The five years I have had the privilege to spend at the EUI have been one of the most incredible periods of my life: therefore, my first acknowledgement goes to the institute for being such a unique place, which allows a doctorate to be a highly enriching experience in so many dimensions.

The key figure for any PhD student is of course the supervisor, and in my years at the EUI I have had the luck to have Andrea Mattozzi as my first advisor: I am hugely indebted to Andrea not only for teaching me how to be a researcher and following the development of this thesis step by step, contributing with countless brilliant suggestions and having the patience to correct my mistakes, but also for all the understanding and the support, always with a smile and a touch of irony, that made me feel I had someone on my side and a role model to follow.

I would also like to extend my gratitude to David Levine for his invaluable advice and for giving me the privilege to confront myself and draw from a kind of economist one would think only exists on the cover page of books.

My gratitude also goes to Alessandro Riboni, for supporting me during the academic job market and being part of my defense committee, as well as to Ronny Razin who also kindly agreed to join the committee: thanks to your comments and discussion I hope to greatly improve the papers composing this thesis.

During my research at the EUI I also benefitted from very useful discussions with Piero Gottardi and Andrea Galeotti, and I gained some very valuable experience by assisting Andrea Ichino in his editorial work. I would also like to thank Evi Pappa for being my provisional advisor during my first year, as well as Arpad Abraham, Juan Dolado and all the other members and visiting members of the faculty.

The Economics department at the EUI is also made special by its great staff: I would therefore like to give special thanks to Jessica, Sarah, Lucia, Anne, Julia and Lori, Antonella and Sonia from the bar.

In the last 5 years, I had the pleasure to share countless invaluable moments with my fellow researchers at the EUI: I would like to thank all of you for making the department feel like home. In particular, a big thanks to Gazza, Omar, Pitone, Andresa, Nastia, Toni, Gatto, Franci, Matic, Philipp, Sebastian and many others for making me look forward to going to work every day!

Far away from Florence, but close in so many other respects, I also owe to Tizi and to my late friend Hausby.

This thesis is dedicated to my family, which despite me living out of home for many years, has always been and will always be a pillar of all I do. I would have never been able to succeed academically without the support, the love and the teaching of my parents and grandparents: I have had the luck to grow up in a family with a formidable mix of talents and experiences, starting from my grandparents Franco and Marisa, to my late grandfather Giovanni, to my incredible parents Gianni and Susanna. Thank you all for always believing in me, for giving me the passion for knowledge and hard work and for always being role models to follow. Acknowledging my family could also not be complete without mentioning my unique crew of siblings Ce, Matti, Mac, Ste, Ciara, Pino and Yann: thank you for always being there for me.

Finally, this thesis is also dedicated to the most important person in my life, my amazing girlfriend Johanna. Along with having played a fundamental role in everything I did in the last year, with inspiring discussions, rehearsals, proofreading, and a crucial contribution to my handing in of this thesis, she sweetly stormed into my life during the last year of my doctorate, making it the fullest and happiest I could think of.

Contents

Abstract

My thesis revolves around the question of how information asymmetries affect elections. In particular, I am interested in how electoral concerns shape policy choices and in the effects of institutional arrangements aimed to providing voters with information on politicians.

In the first chapter I model a primary election, i.e. an election to choose a candidate. I show that if party members do not know the quality of candidates, high quality candidates distinguish themselves by proposing more extreme policies. As a result, introducing primary elections increases the quality of candidates but it might lead to policy polarization.

The second chapter, which is my job market paper, develops a model in which a politician takes a repeated action over an issue and is evaluated by a voter through an election. I show that politicians who flip-flop, i.e. change their decision on the issue, are penalized by voters, because flip-flopping signals incompetence. As a result, politicians have an incentive to protect their reputation by inefficiently sticking to their initial policy choice. This decreases the quality of both policy and electoral choices. The paper also discusses how changes in transparency and term limits can discipline the behaviour of politicians.

My third and final chapter, which is joint work with Antoni-Italo de Moragas, describes a media market in which a set of news outlets compete to break a news concerning a politician in office; after receiving a signal of whether the politician is corrupt, media outlets can either fact-check and learn the truth, or publish the news immediately. We show that increasing the number of outlets competing in the market results in less fact-checking and more fake corruption scandals being published. By making the re-election of honest incumbents more difficult, the increase in competition might therefore be detrimental to social welfare.

1 Signalling Valence in Primary Elections

I build a model of two-stage elections in which candidates differ in terms of a privately-observed quality dimension (valence) and they commit to a policy platform before the primary election. I show that primaries select better candidates at the cost of increased polarization. I also endogenize the choice of holding primaries and find that it can lead to asymmetric equilibria, in which only one political party holds primaries. This outcome is consistent for example with the history of several social-democratic parties in Europe, which were often holding primaries while their opponent right-wing parties were not.

1.1 Introduction

"*Mr. Trump's outrageous statements signal that he has some other political virtue some voters value.*"

Justin Wolfers

In 2010, Charlie Bass decided to enter the Republican primary for the House of Representatives in New Hampshire's second congressional district. An experienced career politician with a moderate resume, and a member of the centrist group Republican Main Street Partnership, Bass performed a considerable shift during his primary campaign, frequently appearing at Tea party rallies and taking conservative positions on salient issues such as taxes and abortion [1]. His move proved successful, propelling him to the primary victory and subsequently to a narrow victory against Democrat Anne McLane Kuster [2].

Also the recent election of Donald Trump as 45^{th} president of the United States of America is an interesting example of primary election extremism. A seemingly moderate if not even Democratic-leaning New Yorker, Mr Trump emerged as a dark-horse candidate

[1] During the campaign, for example, he signed an anti-tax pledge by the Tea-Party related group Americans for Prosperity.

[2] As a matter of fact, there are several other Republican politicians in North-Eastern states who in the same period have followed tactics similar to Bass' in a successful manner: two examples are the primary race between Joe Malone and Jeff Perry in the Congress primary in Massachusetts and that between Linda McMahon and Rob Simmons in the Connecticut Senate primary.

and shocked the world with a primary election campaign in which he took radical positions on many crucial issues, such as immigration and international trade [3].

Was the move to the right performed by these candidates just a way to cater to partisan primary voters, or is there more to that? Why would non-ideological candidates risk jeopardizing their general election prospects with extreme proposal that do not resonate well with the moderate electorate? The point is that the policy platform is not all a politician has to offer: voters also value other qualities such as competence, integrity, or the fact of being an outsider to the establishment. However, given the more intangible character of these qualities, it is fundamental for a politician to find a way to credibly convey that he is competent, honest or truly different from establishment politicians.

In this paper I propose a mechanism by which politicians use the policy platforms proposed in a primary election to send voters a signal of their quality (valence). As Justin Wolfers wrote in the New York Times concerning Trump's proposal of a complete shut down of Muslims from the United States, the economic theory of signalling "suggests that his statement is less a calculated attempt at feeding a demand for bigotry and more an effort to fuel the hunger for authenticity". This is precisely what happens in my model: valent politicians run on platforms which are more extreme than those party voters would ideally want them to propose, but the additional extremism serves the purpose of convincing voters about the candidate's valence.

The intuition for this result is the following: first of all, primary elections make sure that politicians need to develop a platform well in advance of the general election. At that stage, politicians are likely to still not be very well known by voters. Between the primary and the general election, however, there is plenty of time in order for voters to learn about the quality of a politician, thanks to the discussion on the media and the increased attention politicians receive. For example, a scandal might reveal the candidate's incompetence or dishonesty. In this environment, a non-valent politician running on an extreme platform runs the risk of getting caught: not being able to change his policy position, the candidate would see his chances of winning the general election sharply reduced. This allows valent politicians to separate from non-valent ones by choosing an extreme enough platform that discourages non-valent politicians from mimicking them.

If this mechanism holds, then the existence of primary elections creates an important trade-off: primary elections make it possible to select more valent politicians, but also lead to increased polarization of platforms.

Do these stylized facts square with the evidence we have on the effects of primary elec-

[3] Some particularly famous examples are the statement concerning the wall on the border with Mexico, the ban on Muslim migrants and the idea of fully renegotiating NAFTA and exiting from the TPP.

tions? As far as candidate quality is concerned, Snyder and Hirano 2014 show that open-seat primary elections of the advantaged party in a safe district are effective at selecting high quality candidates [4]. A question that naturally arises is therefore what mechanism lies behind the improved selection of candidates achieved by primary elections, and the aim of the paper is to describe one possible such mechanism.

Along with improving the quality of candidates, primary elections are often associated with polarization: an explanation that is often given is that voters participating in primaries (especially closed) tend to be a selected sample of politically engaged partisans. From an empirical point of view, however, this question is still not settled: Hirano, Snyder, Ansolabehere, and Hansen (2010), for example, find no conclusive evidence of a link between primaries and polarization. At the same time, however, there are also studies suggesting the existence of an effect: Brady, Han, and Pope (2007), for example, find that candidates that are too close to the general electorate preferences suffer a penalty in primary elections [5]. In light of this conflicting evidence, a reasonable hypothesis is to think that primaries lead to polarization only under some conditions: had the anti-Washington mood not been so strong in the 2010 congressional primary, Charlie Bass would have probably run on a more moderate platform. In other words, this model provides a possible explanation of what the factors needed for primaries to result in higher polarization are.

Finally, another important stylized fact about primaries concerns the decision of holding them: although now fully part of the US electoral system, the subset of countries which regularly use primaries is rather limited [6]. Moreover, it is sometimes the case that only one major political party [7] uses them, as it happened several times in Italy, France and in many South American countries. How can this phenomenon be explained? Would parties benefit from primaries being compulsory? As I explain later, this research is also capable

[4] Although they do not directly measure quality, Carey and Polga-Hecimovic 2006 find that candidates elected through primaries perform better than candidates selected without a primary elections using a sample of Latin American elections.

[5] Whereas Ansolabehere et al. use the introduction of mandatory primaries to answer their question, the majority of studies looking at the connection between primary elections and polarization use the variation (either cross section of in the time-series) in the degree of openness of primary elections: in this context, McGhee, Masket, Shor, and McCarty (2014) find little variation in candidate polarization, whereas Kaufmann, Gimpel, and Hoffman (2003) and Elisabeth R. Gerber (1998) suggest that closed primaries select more polarized candidates than open primaries. Bullock and Clinton (2011), on the other hand, find a polarizing effect of primaries only in competitive districts.

[6] See Carey and Polga-Hecimovic (2006) and Kemahlioglu and Hirano (2009) for a discussion of the use of primaries in Latin America; in Europe, primaries have been used in several countries including Italy, France, Greece and Spain.

[7] In particular, especially in Europe, it seems that when primaries are held in only one party, centre-left parties are more likely to hold them: examples include the Partito Democratico in Italy, the PS in France (although the UMP is set to hold primaries for the coming election, too), the PASOK in Greece, the Labour Party in the UK.

of providing answers to these questions.

The model I develop is a simple Downsian model of a two-stage election. Two potential candidates are drawn to compete in the primary, the winner becoming the candidate for the general election against an incumbent. Candidates are office-oriented but they are able to commit to the policy they choose when entering the primary election race. The key feature of the model is that valence is private information of candidates at the time of the primary election, but there is a positive chance that it gets revealed before the general election: as anticipated above, this can be thought of as the probability that a scandal emerges in the time interval between the primaries and the general election, or as the fact that general election valence can sometimes only be observed once the general election campaign starts and the primary nominee is pitched against the incumbent. These two ingredients - commitment [8] and the possibility of valence being revealed [9] - make platform choice a credible signal of valence.

The main result of the paper is that in such a model, under some conditions there exists a unique separating equilibrium in which valent candidates differentiate themselves by choosing more partisan platforms. In other words, primaries manage to select better candidates even if valence is unobservable at the time of primaries, but they do it at the cost of increased polarization. This result goes in the opposite direction compared to what happens in models where valence is publicly known, such as Hummel (2013). In such models, valent candidates pick more moderate platforms than their competitors, since the observability of valence gives them a bargaining power which they do not have in my model. Moreover, the outcome of primaries in my model is that voters will not only choose valent candidates, but they will also like their platforms more than those of their non-valent opponents in terms of liberal-conservative location.

My result can also help shed some light on the relationship between electorate and platform polarization. Even in districts in which polarizing forces such as the Tea Party or confessional organizations such as the Evangelical Church play a secondary role, as for example in the New Hampshire district where Charlie Bass won his congressional race, politicians willing to showcase qualities such as integrity and willingness to act independently from the establishment might take polarizing partisan positions. With such a mechanism in place, the polarization of candidates' platforms does not require polarization in the underlying electorate [10].

[8] If flip-flops were possible, but still costly to the politician, then my result would not change. As examples of primary elections models with flip-flopping see Hummel (2010) and Agranov (2016).

[9] Snyder and Ting (2011) build their model on a similar assumption, according to which the type of politician can be revealed before or after the primaries.

[10] There is an open debate on whether the American electorate is becoming more polarized. Abramowitz

Last but not least, concerning the question of whether parties will choose to hold primaries, I show that there are several possibilities: apart from the two polar cases in which primaries are either held or not held by both parties, there is an intermediate region in which the only pure-strategy equilibria are asymmetric and involve only one party holding primaries, which reflects the examples of countries in which only one major party holds primaries. Notice that given the linearity of preferences I use in my model, parties would always prefer an equilibrium where both candidates are selected through primaries, but that is not always sustainable: the model therefore suggests that political parties should be in favour of making primary elections compulsory[11].

1.2 Related Literature

My paper mainly relates to the literature on primary elections, adding the novel feature of signalling valence through the choice of a more radical political platform. More generally, however, this paper develops a theory on the interaction between candidate valence and policy platform choice: in this respect, an important reference is Kartik and McAfee (2007): in both models, valent candidates (or candidates with character, as they call them) are more extreme than non-valent ones, but whereas in their model it is up to non-valent candidates to locate in such a way to maximize their probability of winning elections, in my model it is valent candidates who strategically take more extreme positions in order to separate from non-valent candidates [12]. Another similarity between the models is that the more frequent valent candidates are, the more extreme a policy needs to be in order for voters to know that the candidate is valent: in their model, this happens because the support of policies chosen by strategic candidates broadens, whereas in mine the reason is that valent candidates need to pick more extreme policies to signal their valence. What is more, Kartik and McAfee also show how their mechanism translates into a simple two-stage election setup: when some candidates have character, what they call "General Election Indifference" can exist. In this respect, my paper can be seen as an extension of their work, focusing on two-stage elections and fully strategic candidates.

Whereas the aforementioned paper by Kartik and McAfee is the one sharing the most

(2010) argues that the American public is getting more polarized, whereas studies such as Seth J. Hill (2015) find that the polarization of the electorate has not increased.

[11] A caveat to this conclusion, however, lies in the fact that my model does not take into account the potential advantage that party elites might have from being able to choose candidates directly rather than through a primary election process, which might favour anti-establishment outsiders.

[12] Moreover, in my model both types are strategic, whereas in Kartik and McAfee (2007) valent candidates are non-strategic.

common ground in terms of relation between valence and candidate location, the most similar study in terms of model setup is Hummel (2013). What is interesting in the comparison between my paper and Hummel's is the fact that the assumption of privately observed valence I make in my model significantly changes the result: in Hummel's paper, valent candidates choose more moderate policy platforms than non-valent ones: the reason is that when valence is observable, it is sufficient for valent politicians to make voters indifferent between them and more partisan non-valent candidates and be in a better position to win the general election. My contribution might therefore also be seen as a way to qualify Hummel's result and underline how the observability of valence is crucial in obtaining his result. Moreover, my model can also be though of as a way to bring some of the ideas of Kartik and McAfee inside a primary election model a la Hummel [13].

Another paper that finds a similar connection between candidate location and valence is Casas (2014): in his model, voters cannot observe policies before elections, but they can observe party affiliation and valence. An implication of his model which is similar to my findings is that valent candidates are more ideologically volatile, hence more likely to be extreme. In other words, policy moderation implies a lower valence, just as in my model. The interesting parallelism between the two papers is that the complementarity between valence and moderation turns into substitutability when either policy preferences (as in Casas) or valence (as in my paper) become unobservable. Another similar implication of the two papers is that the smaller the differentiation, the more profitable the "investment" in valence (i.e. primaries in my model) is. However, the unobservability of policy as opposed to valence in Casas' paper also brings about some differences: in his model, the more moderate the "nominator" is (i.e. the more open primaries are), the lower the valence of candidates. This is due to the fact that a moderate nominator dislikes policy variance. In my model, on the other hand, primaries increase the probability of selecting valent candidates. In other words, whereas in both models valence and moderation are substitutes given an institutional setup, in Casas' model they are complements *across* institutional setups: this means that, unlike what I conclude in my model, his model predicts that (closed) primaries select more partisan platforms but also less valent ones compared to handpicking by the party leader.

As we have seen, Casas (2014) analyzes the effects of candidate nomination rules on party platforms and candidate valence. Whereas I don't study the difference between closed and

[13] It is also useful to point out that if Hummel's is the primary election paper that mostly resembles mine, there are other similar models in the literature, all differing along some features: Grofman, Troumpounis, and Xefteris (2016) consider both closed and open primaries with candidates of potentially different (observable) valence and sincere voting; Takayama (2014) Adams and Merrill (2008)

open primaries, I do address the question of whether a party would like to hold (closed) primaries. In this respect, my paper is similar to Serra (2011): in both models, the trade-off for party leaders is that handpicking the candidate gives them control over the platform but results in a lower chance of selecting a valent candidate. In his paper, however, the increased probability of getting a valent candidate through the primaries is an assumption of the model, whereas the improved selection of candidates occurs endogenously in my model. In this respect, this paper provides a microfoundation of Serra's guiding assumption.

Finally, the problem of deciding to hold primaries is addressed also by Snyder and Ting (2011): in their model, the cost of primaries is that they can reveal the candidate as being bad before the general election. For this reason, they conclude that primaries are optimal in sufficiently safe districts. The implication of my model, on the other hand, is that primaries are more likely to be optimal in when parties are not very differentiated (which is to say in competitive districts): this result is particularly interesting because at first sight it appears counterintuitive. In a competitive district, as a matter of fact, policy polarization is more costly for the general election, but it is also simpler for valent politicians to separate. Moreover, the policy cost of losing the election is smaller in competitive districts than in safe districts.

The paper is structured as follows: Section 3 will present the model in which only one party holds a primary election: I call this the single-primary model. The main results from this model are described in Section 4, along with a discussion and a welfare analysis. After that, Section 5 presents a more general version of the model, in which both parties hold primaries. The case of endogenous primaries is also treated there. Section 6 concludes.

1.3 The Baseline Model

I build a very simple two-stage election game based on a standard Downsian spatial voting model. A unit mass of citizens have policy preferences distributed over the $x \in [0, 1]$ interval. All that matters about the distribution of voters is that there is a median voter m who is randomly located in the $[\frac{1}{2} - \epsilon, \frac{1}{2} + \epsilon]$ interval following a uniform distribution. There are two political parties, whose policy bliss points are for simplicity fixed at the extremes of the interval: I call the party located at 0 Democratic party and the party located at 1 Republican party, or D and R respectively. The policies proposed by each party in the general election are denoted by r and d respectively. The voters of each party are identical and their condition of party voters is predetermined and fixed throughout the whole model [14]. There are two stages of elections: in the first stage, the primaries, platforms and candidates are selected. In this stage, parties choose one candidate between two randomly drawn politicians. As soon as their are drawn, politicians propose a platform and commit to their choice for the whole game. In the second stage, the general election, the two winners of the primaries face each other and every citizen votes for one of these two candidates. In other words, this is a model of closed primary elections, because only party members vote in the primaries. Politicians are purely office motivated, receiving a utility of 1 if elected in the general election and 0 otherwise. There are two types of politicians, valent and non-valent. The share of valent politicians in the population of potential candidates of both parties is α. Valence is a set of non-policy characteristics of a politician that provide utility to all voters across the ideological spectrum: these characteristics can be thought of as honesty, work ethics, international reputation but also as campaigning ability. I assume that valence is binary, so that a valent politician has valence $v_j = v$, whereas a non-valent politician has valence $v_j = 0$. Voters' utility is additively linear in the elected politician's valence and the distance between his policy platform p_j and the voter's bliss point x_i:

$$U(x_i, v_j, p_j) = v_j - |p_j - x_i|$$

In the first stage, when the primaries take place, valence is private information of politicians. In the baseline version of the model, I assume that after the primaries take place, valence is exogenously revealed with probability 1: this model is therefore particularly apt at considering valence as general election campaigning ability or popularity with the general electorate. The fact that valence can be exogenously revealed after primaries is similar to what Snyder and Ting (2011) assume in their model, and it particularly fits the case of any attribute that gives a politician an electoral advantage and that is acknowledged by voters following the primaries. Moreover, the more general case where valence is only revealed with some probability $z \in (0, 1)$ is very similar and it is treated in the extensions. When voting in the primary stage, party members are fully rational: they maximize their expected utility by taking into account the policy platform but also the probability of winning that each candidate faces in the general election given his policy platform and expected valence. To further simplify the problem, for now I assume that a primary election only takes place within the Democratic party, whereas the Republican party candidate is already set at location r and I normalize his valence to zero. Moreover, I assume that $r \geq \frac{1}{2}$. A useful notation that will be used throughout the model is $\tilde{r} \equiv r - (\frac{1}{2} - \epsilon)$, which denotes the

[14] The party bliss point could also be thought of as the bliss point of the decisive median voter inside the party.

dstance between the Republican policy and the most leftwing media voter. The analogue expression for the Democratic policy d is therefore $\tilde{d} \equiv \frac{1}{2} - \epsilon - d$. I further assume that valence is not too large, and in particular that $v \leq \tilde{r}$. This assumption makes sure that a valent democratic candidate located at $\frac{1}{2} - \epsilon$ would not win the election for sure. To sum up, the timing of the model is the following:

1. Two Democratic primary candidates are drawn.
2. Primary candidates simultaneously choose a policy platform.
3. Party members observe the platforms and choose one candidate.
4. All citizens observe the valence of the chosen democratic candidate and vote in the general election.
5. The winning policy is implemented and payoffs are distributed.

At the heart of the model, in other words, is the fact that parties (for now only the Democratic party) have to choose a candidate before valence is revealed: the fact that politicians cannot change their platforms after primary elections gives politicians the possibility to credibly signal their valence, as we will see in the next sections.

1.4 Results

1.4.1 General Election

In order to understand the solution of the game I start from the second stage, i.e. the general election. Once the Democratic party primary has taken place and the Democratic platform is set to d , the game is a standard Downsian election with the possibility of a valence advantage v for the Democratic politician. The winning party is determined by the location of the median voter m compared to that of the indifferent voter $z \equiv \frac{d+r+v}{2}$: D wins if $m \leq z$ and R wins if $m > z$. The probability of winning for the Democratic candidate is:

$$\pi(d, r, v) = \frac{1}{2} + \frac{1}{4\epsilon}(d + r + v - 1)$$

The value of π is important since it defines the politicians' payoff conditional on winning the primary.

In order to understand primary election behaviour, the first step is to study party preferences. As mentioned above, parties maximize the expected utility of a candidate's policy, thus trading off location and electability. This turns out to be particularly simple thanks to the assumption of a linear utility function: neither valence nor the opponent's location affect the utility maximizing location of the Democratic candidate, which uniquely depends on the median uncertainty parameter ϵ:

Lemma 1. *In the general election, the location maximizing the Democratic party's utility is* $d^* = \frac{1}{2} - \epsilon$.

Proof. See Appendix 1.B. □

1.4.2 The Primary Election

The heart of the model is the primary election. Since I do not model disagreement within the party [15] party members vote unanimously, as if a single party entity was taking the decision. Moreover, given that two candidates are randomly drawn from the pool [16] there are three possibilities: both are valent, none is valent or only one is valent. Without private information over valence, primaries would have no effect other than that of getting politicians to compete and locate at d^*, which is an outcome equivalent to the party having direct control over the candidate. With private information over valence, candidates can use the primaries to send signals about their valence. The way this works is simple: the policy platform announced before the primary election remains the same for the general election and therefore choosing a platform further away from the expected median decreases a politician's probability of winning the general election. Valent politicians, however, are able to perform better thanks to their valence. It follows that if valent politicians choose an extreme enough policy, they prevent non-valent politicians from mimicking them, making their platform choice a credible signal of valence. This, however, is only feasible if, despite their lower chances of success in the general election, party voters are willing to elect signalling politicians in the primaries, which requires the following condition to hold:

Condition 1. $\tilde{\rho} \equiv \frac{v}{\tilde{r}} \geq \frac{\sqrt{1+(2-\alpha)^2}}{2-\alpha} - 1.$

Derivation. See Appendix 1.B. □

The main result of the paper can be summarized in the following theorem:

[15] Two things have to be noted: first, even with disagreement within the party, results would not change as long as the median voter theorem applies. What is more, even if assuming that party voters are distributed in some interval, given these preferences their first-best policy choice d^* would be the same for all of them if $x < \frac{1}{2} - \epsilon$.

[16] The fact that two and not more candidates are drawn is arbitrary and chosen for simplicity; in fact, an interesting question, which I will partially address in the extensions, concerns the optimal number of candidates drawn for the primaries.

Theorem 1. *Assume that Condition 1 holds. Then, the Democratic primary election game has a Perfect Bayesian Equilibrium (PBE) in which low-valence candidates locate at* $d_L = d^* = \frac{1}{2} - \epsilon$ *and high-valence candidates locate at* $d_H = \frac{1}{2} - \epsilon - \frac{1}{2-\alpha}\tilde{r}$. *This equilibrium is the only one which survives the Intuitive Criterion by Cho and Kreps (1987).*

Proof. In a separating equilibrium, valent candidates know that they will win the primary with probability $1 - \frac{\alpha}{2}$, since they only lose if also the other candidate is valent and they are not selected in the random tie-breaking. Non-valent politicians, on the other hand, can still enjoy the probability of winning the primaries if none of the two candidates is valent, which from their perspective translates in a winning probability of $\frac{(1-\alpha)}{2}$. It follows that politicians can credibly signal their valence by moving their platform to the left. The size of this movement is given by the incentive compatibility constraint for low types, which reads:

$$\frac{1-\alpha}{2}\left[\frac{1}{2} + \frac{1}{4\epsilon}(d_L + r - 1)\right] \geq \left(1 - \frac{\alpha}{2}\right)\left[\frac{1}{2} + \frac{1}{4\epsilon}(d_H + r - 1)\right]$$

and yields

$$d_H \leq \frac{1-\alpha}{2-\alpha}d_L + \frac{1}{2-\alpha}(1 - 2\epsilon - r)$$

Thanks to the intuitive criterion, all platform choices satisfying the above condition will be attributed to the valent type; ceteris paribus, valent politicians prefer to locate closer to the center, so that in the unique separating equilibrium, the above constraint will hold with equality.

In order to pin down the location of d_L, notice that politicians are office-motivated, but in the primary election they are pitched against each other and this leads them to compete. Hence, if $d_L \neq d^* = \frac{1}{2} - \epsilon$ in a separating equilibrium, any deviation to some $\hat{d}$ closer to d^* will lead the non valent politician to win for sure at least whenever the other candidate is another non-valent politician locating at d_L. The reason is simply that d^* is the bliss point for party voters, and in a separating equilibrium beliefs are already the worst possible for a politician at d_L, i.e. $\mu = 0$. For the same reason, any non-valent politician deviating out of d^* in a separating equilibrium will never win the primary (unless she were to deviate to d_H, of course, but that is already ruled out by the incentive compatibility constraint).

In order to see that a pooling equilibrium is also not sustainable, consider without loss of generality a pooling at d^*: a movement to the left of $\tilde{d} \geq \frac{\tilde{r}}{2}$ will be interpreted as coming from the valent type, given the intuitive criterion [17]. Then, however, any valent politician would strictly profit by carrying out such a deviation, as it can be immediately verified.

[17] This simply comes from the incentive compatibility constraint when low types win with probability $\frac{1}{2}$ in the pooling equilibrium at d^* and deviating results in winning with probability 1

□

In terms of comparative statics, the movement to the left that valent democratic politicians have perform is $\tilde{d}_H = \frac{1}{2-\alpha}\tilde{r}$. This means that the higher α is, the more radical the signalling policy becomes: when α moves closer to 1, the separating policy moves towards a policy with zero winning probability in the general election for a hypothetical non-valent candidate mimicking the valent type's policy: this is intuitive since as α goes to 1, the probability to win the primary election for a non-valent politician goes to zero.

Moreover, the further to the right is the location of the Republican opponent, the more movement to the left is required from the Democratic candidate. In other words, primaries contribute to political polarizarion. Notice that $\tilde{r}$ might also be interpreted as a parameter representing district safety, or the advantage of the Democratic party: the interpretation hence would be that primary elections create more polarization in safer districts [18]. Notice that the valence of the Democratic politician does not affect the movement necessary to signal, which depends on the probability of winning of the non-valent candidate. If the Republican candidate were valent himself, what's more, the signalling location would become more moderate, because the low-valence candidate would win with a lower probability [19].

In terms of probability of winning the general election, valent politicians, notwithstanding their more partisan platform, win the general election more often than non-valent ones as long as $\frac{v}{\tilde{r}} \geq \frac{1}{2-\alpha}$, so that below that threshold primaries actually decrease the expected probability of seeing a valent (Democratic) politician in office. Overall, the probability (compared to a hypothetical pooling equilibrium) of electing a valent politician in the general election increases as long as $\frac{v}{\tilde{r}} \geq \frac{\alpha}{1-\alpha}$. The interpretation of this threshold in terms of district safety is that primary elections increase the presence of valent politicians in competitive districts, while they decrease it in safe districts (the level of safety depending on the likelihood of drawing a valent politician α). This seems at first sight to be at odds with the evidence presented by Snyder and Hirano (2014), but allowing for a larger share of valent candidates in safe districts could bring my results in line with theirs. The comparative statics of the model can be collected in the following proposition:

Corollary 1. *Primary elections generate increasingly partisan platforms whenever α is larger (more valent politicians in the challenger pool) or $\tilde{r}$ is larger (i.e. more partisan*

[18] The fact that the more advantaged the party is, the more valent candidates have to be radical in order to win the primary partially offsets the initial advantage of the party. In other words, the signalling happening in my model acts to make what Grofman, Troumpounis, and Xefteris (2016) call *matching effect of primaries* smaller, making competitiveness across districts more homogeneous.

[19] Adjusting the model to deal with a valent opponent is straightforward and yields a separating location of $\tilde{d}_H \geq \frac{1}{2-\alpha}(\tilde{r} - v)$.

Republican candidate or safer district). Conditional on winning the primary, valent politicians still win general elections more often than non-valent politicians as long as $\frac{v}{\bar{r}} \geq \frac{1}{2-\alpha}$*. Moreover, the probability of electing a valent politician in the general election, compared to a hypothetical pooling equilibrium, increases as long as* $\frac{v}{\bar{r}} \geq \frac{\alpha}{1-\alpha}$*.*

As a final note, in the model I assume that valence is a quality benefitting all voters across the ideology line, but the results would not change, except for some welfare comparisons, if I assumed that valence is only benefitting the (in this case Democratic) party voter and the median voter.

1.4.3 Discussion

Having presented the main result of the paper, it is now useful to discuss some of the features of the model. The main insight of the theorem presented above is that in a primary election where candidates can commit to a policy platform, choosing relatively extreme positions can be an instrument to signal a politician's valence. Two things seem particularly worthy of a discussion: the first one is the main prediction of the model, i.e. the fact that, in a primary election, valent (i.e. *good*) candidates will tend to propose more partisan policies than non valent ones.

On one hand, there is some evidence point to the fact that valent primary candidates choose more moderate policies (See for example Ansolabehere, Snyder, and Stewart III (2001), Stone and Simas (2010) and Brady, Han, and Pope (2007)). However, there are also elements pointing in the opposite direction: for example, Stone and Simas (2010) shows that challengers choosing more extreme platforms have a higher probability of replacing the incumbent in House elections; similarly, Brady, Han, and Pope (2007) and Hirano, Snyder, Ansolabehere, and Hansen (2010) show that ideologically moderate incumbents are more vulnerable to primary challenges. All these findings are compatible with the results of my model. Notice that even if it were the case that valent candidates take more moderate positions on average, this could just mean that valence is more likely to be observable (or, if we have both observable and unobservable valence, that the former effect dominates): the reason is that with observable valence, valent candidates would play more moderate policies, such as in Hummel (2013). Moreover, the situation my model described most accurately is one in which a primary needs to decide the candidate who will challenge an incumbent from the opposing party: in other words, the Democratic primary in a Republican district or viceversa.

The second issue that I would like to discuss is the possibility that this kind of signalling could also happen in a context different from that of a primary election [20]. However, the two fundamental ingredients, which is to say the possibility to commit to a policy platform and the sequentiality whereby valence is private information at the moment of choosing the platform but can be exogenously revealed before the general election, suggest that the context of primaries might be the most likely to see this type of signalling happen. Apart from the sequentiality issue, which is straightforward, it seems plausible to assume that participating in a primary offers a much better commitment-to-policy device than writing a political manifesto within the walls of a party's *smoke-filled-room.* As a matter of fact, in a later section I will briefly describe an alternative model, in which it is the party who selects candidates by offering a menu of platforms: valent politicians will take the more partisan platform compared to non-valent ones. However, the commitment assumption appears less realistic in that model: without the exposure given by a primary, it is less credible for politicians taking the more partisan platform to not flip-flop when running in an actual election. Of course it has to be taken into consideration that in this model the absence of flip-flopping is not derived endogenously, but rather ruled out by assumption.

The discussion about the screening alternative, which as you will see later on in the paper is, as intuition suggests, superior to primaries, leads me to the other point that I would like to consider: the comparison between an environment with primary elections and an alternative environment. To this moment, as a matter of fact, I have not specified how the selection of candidates would work if primary elections where not held. Apart from the theoretically optimal candidate screening, there are several other possibilities, of which I will just mention some benchmarks: the first one is to think that without primaries, candidates would be completely unrestrained by the party. In that case, we could imagine that they would run on a moderate platform [21] such as $\frac{1}{2}$. On the other hand, we might think that without primaries, the party will choose an insider to run as candidate, but that this candidate can never have the extra valence. Lastly, we might think that without primaries, the party can get candidates to run at d^*, but that candidates are just randomly drawn from the pool of politicians: in this case, no primaries would be equivalent to a pooling equilibrium. When talking about the welfare effects of primary elections, therefore, one alternative has to be picked as no-primary benchmark. In the following paragraph I will use the pooling equilibrium at d^* as benchmark.

[20] Since I do not model the possibility of ideological differences between the party elite and the party median voter, in my model holding primaries means letting candidates propose a platform.

[21] Since r is exogenously fixed in this version of the model, a politician free to locate would choose a position just to the left of r, unless this is not exactly $\frac{1}{2}$

1.4.4 Welfare Analysis

The fundamental trade-off driving the welfare comparisons is between more extreme policies and valence of politicians. Let's start by looking at politicians' welfare. As in all signalling games, a separating equilibrium potentially favours the types carrying out the signalling, which in this case is valent politicians. Non-valent politicians therefore lose compared to a hypothetical pooling equilibrium where both primary candidates locate at d^* and one is picked at random (which is equivalent to having no primary elections). As far as valent politicians are concerned, the trade off they face is that they are more likely to be chosen as candidates for the general election, but at the same time they need to pick a more extreme policy, so that their probability of winning the general election decreases. It turns out that they only gain if the following condition is satisfied:

$$\frac{v}{\tilde{r}} \geq \frac{\alpha}{1-\alpha}. \tag{1.1}$$

This is (intuitively so) the same condition according to which the probability of seeing valent candidates elected to office increases thanks to the primary. Valent politicians gain if either the right-hand side is large, which happens if α is small (it's easier to separate when there are few good politicians), or if the left-hand side is small, which happens when either $\tilde{r}$ is low (the Republican candidate is moderate, or the district is competitive) or v is high (valence is high). Interestingly, politicians are more likely to gain from primaries when the opponent is moderate (or the district is competitive) and that is because separating requires a smaller movement to the extreme (in this case the left). Finally, notice that only if $\alpha < \frac{1}{2}$ valent politicians can gain from the primaries under the assumption which binds the value of valence [22] to $\frac{v}{\tilde{r}} \leq 1$.

Consider now party welfare. As a result of primaries, the party can choose valent politicians when they are available. Although valent politicians choose policies that are closer to the party ideal, they win the general election less often. As a result, the Democratic party gains from primaries when either valence is high enough or the district is competitive enough:

$$\frac{v}{\tilde{r}} \geq \frac{1}{(2-\alpha)\left[(1-\alpha)+\sqrt{\frac{(1-\alpha)}{(2-\alpha)}\left[1+(1-\alpha)(2-\alpha)\right]}\right]}. \tag{1.2}$$

Notice from (1.2) that as α grows closer to one, a point is reached where the party never gains from primaries, since the right-hand side grows unboundedly but we assume that

[22] The assumption means that valence is never so large to make a hypothetical valent politician located at $\frac{1}{2} - \epsilon$ win the election for sure

$\frac{v}{\tilde{r}} < 1$; the intuition is that if the likelihood of having a valent politician is very high, there is little reason to hold a primary.

An interesting question concerns the alignment of interests between valent politicians and the party. By analyzing (1.1) and (1.2), one derives that there is a value of α above which the party threshold is higher than the politician threshold. This means that if α is large enough, then it is possible for parties to gain from primaries when politicians don't and viceversa if α is low enough. In the former case, politicians would prefer a pooling equilibrium, but in this game they have no way to commit to it. In the latter case, the party would gain from the ability to commit to rejecting a politician who signals that he is valent. The possibility of these conflicts occurring is an indication of the importance of electoral institutions regulating the primary election process.

What about the welfare of the Republican candidate and party? As far as the Republican candidate is concerned, it is easy to show that he gains if

$$\frac{v}{\tilde{r}} \leq \frac{1}{1-\alpha}.$$

Notice that this inequality is not just (1.1) with a flipped sign, since the Republican politician doesn't know the type of candidate (valent or not valent) he will be facing.

As far as the Republican party is concerned, it gains from the Democratic party's primaries if $\frac{v}{\tilde{r}} \geq \frac{1}{(2-\alpha)-\sqrt{2-\alpha}}$ or if $\frac{v}{\tilde{r}} \leq \frac{1}{(2-\alpha)+\sqrt{2-\alpha}}$, but notice that only the latter is relevant because of the assumption that $\frac{v}{\tilde{r}} \leq 1$. In other words, the Republican party gains from the primaries when either valence is sufficiently low (which is intuitive) or when the district is leaning to the Democratic side, i.e. high $\tilde{r}$: this is less intuitive but it is due to the fact that a larger $\tilde{r}$ leads to more extreme (and therefore less competitive) valent Democratic candidates.

Last but not least, the welfare of the median voter is characterized by an algebraically harder condition to analyse:

$$\frac{v}{\tilde{r}} \geq \frac{1}{(2-\alpha)\alpha\frac{\sqrt{1+\frac{(1-\alpha)(3-2\alpha)}{\alpha^2(2-\alpha)}}-1}{3-2\alpha}}$$

The intuition, however, is analogous to that previously encountered: it takes either a sufficiently high valence or a sufficient competitiveness of the district for the median voter to gain from the institution of primaries.

1.4.5 The Screening Alternative

In this section I describe the selection of politicians taking place if parties could screen candidates by offering a menu of policy platforms. The intuition is the same as that applying to the signalling model described in the paragraphs above, but with one important difference. Since parties can choose the screening policies before drawing the two potential candidates, they can choose the utility maximizing menu of policies, which can be either a pooling or a separating one. As a consequence, not only pooling becomes feasible, but the pair of separating policies are not the same as those described for the signalling model. The reason is very simple: in the signalling model of primaries, candidates compete, and party voters cannot commit not to vote for a candidate offering them their bliss point once they have learnt about the candidate's valence. This means that, in equilibrium, low valence candidates have to locate at $d^* = \frac{1}{2} - \epsilon$. In the screening model, on the other hand, the party is able to distort the policy offered to low valence politicians, making it more moderate in order to make the separating policy offered to high valence politicians closer to their bliss point. The result can be summarized in the following proposition:

Proposition 1. *The candidate screening game has a unique equilibrium: if* $\tilde{\rho} < \sqrt{1 + \frac{4-\alpha}{4(1-\alpha)}} - 1$*, then parties offer a pooling menu with* $d = d^*$*; if* $\tilde{\rho} \geq \sqrt{1 + \frac{4-\alpha}{4(1-\alpha)}} - 1$*, then parties offer a separating menu of policies* (d_L^{scr}, d_H^{scr})*, with* $d_L^{scr} \geq d^*$ *and* d_H^{scr} *satisfying the same incentive compatibility constraint as in the signalling game.*

Proof. See Appendix 1.B. □

1.5 Primaries in Both Parties

In this section I present a generalized version of the model, in which both parties contemporaneously hold a primary election. The consequence of having a double primary is that, on each side, the separating equilibrium is played against an opponent of uncertain location and valence. The core of the analysis is very similar to the single primary case: under a condition analogous to Condition 1 for the single-primary model, the game has a unique separating equilibrium in which candidates of both parties signal by choosing more partisan platforms. Denote by $\rho \equiv \frac{v}{2\epsilon}$ the double-primary environment analogue of $\tilde{\rho} = \frac{v}{\tilde{r}}$. The double primary non-rejection condition reads:

Condition 2. *Assume that* α *and* ρ *satisfy the following condition:* $A(\alpha)\rho^2 + B(\alpha)\rho - C(\alpha) \geq 0$*, where* $\rho \equiv \frac{v}{2\epsilon}$ *and* A*,* B*,* C *are functions of* α *derived in the appendix.*

Derivation. See Appendix 1.B. □

The following proposition summarizes the main results of the double primary model:

Proposition 2. *Assume that Condition 2 is satisfied. Then, the double primary game has a unique separating equilibrium under the intuitive criterion. On the democratic side, valent politicians separate by proposing platform:*

$$d_H^{dp} = d^* - \frac{1}{1-\alpha}\left[\frac{2\epsilon - \alpha(2-\alpha)v}{(2-\alpha)}\right], \tag{1.3}$$

whereas non-valent politicians choose $d_L^{dp} = d^* = \frac{1}{2} - \epsilon$*. The locations of the* R *candidates are* $r_H^{dp} = 1 - d_H^{dp}$ *and* $r_L^{dp} = 1 - d_L^{dp}$*, hence the equilibrium is symmetric.*

Proof. See Appendix 1.B. □

As one can see from the expression for d_H^{dp}, there are two opposite effects at work, as anticipated in the previous discussion: uncertainty in the opponent's location and uncertainty in the opponent's valence. Uncertainty in the opponent's location has a multiplier effect on the movement required to separate, which is represented by the $\frac{1}{1-\alpha}$ factor, which was not present in the single primary expression: this strategic complementarity reinforces the polarization effect of primaries. The multiplier effect is increasing in the frequency of valent politicians. As far as the valence of the opponent is concerned, its effect is one of strategic substitution which weakens the polarizing effect of primaries. It can be notices that the numerator of the second term of (1.3) is simply the expected value of the opponent's valence (notice that $\alpha(2-\alpha) = 1 - (1-\alpha)^2$, i.e. the probability that at least one valent politician is drawn on one side). Up to the multiplier effect, therefore, this expression is equivalent to the separation condition in a single primary model with $r = \frac{1}{2} + \epsilon$ and valence of v with probability α. Notice that compared to a single-primary case where $\tilde{r} = 2\epsilon$, the movement to the left of the Democratic candidate in the double primary is larger if and only if $\frac{v}{2\epsilon} \geq \frac{1}{2-\alpha}$. As far as the discussion of welfare is concerned, a full analysis is beyond the scope of this section, but it can be noticed that, similarly to the single primary case, where welfare depended on $\frac{v}{\tilde{r}}$, here the driver is $\frac{2\epsilon}{v}$ and in particular valent politicians gain when this ratio is low enough. The intuition is the following: a higher v is an incentive to separate, since clearly it increases the probability of winning for any given location. On the other hand, the smaller the uncertainty, the riskier the gamble of a politician playing a separating equilibrium. As a matter of fact, for any given v moving to the extreme means relying on the possibility of getting a favourable draw of the median: this probability increases with the range 2ϵ along which the median moves.

1.5.1 Endogenous Primary Choice

In the analysis of the previous paragraph I studied the equilibrium arising when both parties hold primaries, but the question of whether parties would endogenously choose to hold primaries remained unanswered. In this section I answer the question of what would happen if both parties were to choose, at the beginning of the game, whether to hold a primary or play a pooling equilibrium with candidates located at d^*. The result is that depending on the parameters, there can be either a no-primary equilibrium (i.e. an equilibrium in which parties do not hold primaries), or a double-primary equilibrium (i.e. one in which both parties endogenously decide to hold a primary). However, it is also possible for an asymmetric equilibrium to arise, in which only one party holds primaries [23]. Whenever asymmetric equilibria exist, there exists also a symmetric equilibrium in mixed strategies, in which each party holds primaries with some interior probability. Intuitively, the no primary equilibrium requires low values of $\frac{v}{2\epsilon}$, whereas the primary equilibrium requires higher values of $\frac{v}{2\epsilon}$. Moreover, for α high enough the primary equilibrium ceases to exist, whereas for very high levels of α the no-primary equilibrium is the unique outcome of the game: the idea is simply that parties benefit from primaries when valence is not too common among the pool of potential candidates.

Proposition 3. *The endogenous double-primary game has three potential outcomes: for some values of α and ρ, in the unique equilibrium parties choose not to hold primaries. At the other end of the spectrum, the unique equilibrium has parties endogenously choosing to hold primaries. Lastly, there is an intermediate region in which either only one party holds primaries, or both hold them with some symmetric interior probability.*

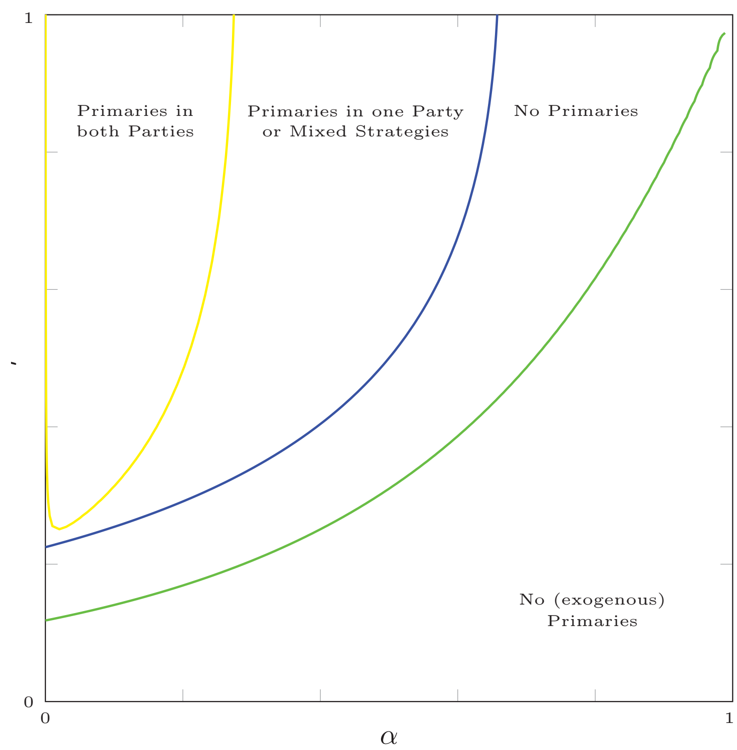
Proof. See Appendix 1.B. □

Figure 1.1 summarizes the results from this section: first of all, the green line represents the non-rejection condition. In other words, separating equilibria exist only in the region north-west of the green line. The orange line, on the other hand, represents the indifference line between not holding primaries and holding primaries when the other party is not holding primaries. This means that to the right of the orange line, a no-primary equilibrium exists. Finally, the purple line is the indifference condition between holding a primary and not holding it whenever the other party is holding a primary. This means that a double primary equilibrium is only sustainable to the left of this line. This means that between the

[23] It is in fact the case that in some countries, primaries have been held by only one of the major parties, such as the the Partito Democratico in Italy.

orange and the purple line, the only equilibria are either asymmetric primary-no primary equilibria, or a symmetric mixed strategy equilibrium.

Figure 1.1: Double Primary Equilibrium Characterization



1.6 Conclusion

This paper develops a model of signaling in primary elections, in which high quality candidates use their positioning along the liberal-conservative axis to separate from low valence ones. Therefore, radical campaign statements prior to a primary election might serve the purpose of convincing voters that a candidate is valent. This means that primaries are polarizing not much because primary electors are more radical, but because politicians exploit primaries to send voters signals about their valence. Under this mechanism, the presence of primaries introduces a trade-off between quality of politicians and policy polarization.

This is a novel result which goes in the opposite direction to what has been found in models where valence is observable at the time of the primary vote. In other words, my model predicts that if valence is unobservable, primary winners (at least in a primary selecting the candidate running against an incumbent, or a fully open seat primary) will generally choose partisan platforms. Moreover, I also show that if parties can choose whether to hold primaries, an equilibrium with primaries is not always sustainable, even if parties are always better off when both hold primaries than when they both do not hold them. In this context it is also possible that only one political side decides to hold primaries in equilibrium, something of which there are several instances in Western Europe.

Future extensions research on this topic could investigate how the mechanism I describe interacts with elements of primary elections that I currently exclude from the model, such as the openness of primary elections, the potential conflict between party leadership and party base and the possibility for politicians to flip-flop between the primary and the general election.

Appendix 1.A Observability of Valence Off-Equilibrium

To build the model presented in the paper I assumed that valence always gets revealed before the general election. In this section I relax this assumption. Notice first of all that the arising of a separating equilibrium makes the observability of valence superfluous along the equilibrium path, whereas it plays an important role in the characterization of the equilibrium. Relaxing the observability assumption (but keeping in place the commitment to platforms), a separating equilibrium is still sustainable. However, it becomes more difficult for valent politicians to prevent non-valent politicians from mimicking their actions. The incentive compatibility constraint now is the following:

$$\left(\frac{2-\alpha}{2}\right)\left\{z\left[\frac{1}{2}+\frac{1}{4\epsilon}(d_H+r-1)\right]+(1-z)\left[\frac{1}{2}+\frac{1}{4\epsilon}\left(d+H+r+v-1\right)\right]\right\}\leq$$
$$\leq\frac{1-\alpha}{2}\left\{\frac{1}{2}+\frac{1}{4\epsilon}\left(d_L+r-1\right)\right\}$$

This can be rewritten as:

$$d_H\leq\frac{1}{2-\alpha}(1-2\epsilon-r)+\frac{1-\alpha}{2-\alpha}d_L-(1-z)v$$

As a result, separation requires valent politicians to move to the left by at least:

$$\tilde{d}_H^z=\frac{1}{2-\alpha}\tilde{r}+(1-z)v$$

which is larger than the amount required in the observable valence case and is increasing in v: this makes primaries less attractive compared to the $z=1$ benchmark for both party members and valent politicians. However, the intuition for the model remains unchanged.

Appendix 1.B Proofs

Proof of Lemma 1. Let μ_j denote the probability that the Democratic candidate j is valent given the information conveyed by the platform location chosen to participate in the primary elections. If the primary conveyed no information (i.e. types pool), then $\mu_j=\alpha$ and the outcome would be the same as drawing a random candidate from the pool. At the other extreme, $\mu_j=1$ indicates a case in which primaries perfectly reveal j's valence. The expected utility the party derives from a candidate locating at d_j and who is believed to be valent with probability μ_j is:

$$\mathbb{E}U_D(d_j,\mu_j) = -r + \mu_j \left[\frac{\tilde{r} - \tilde{d}_j + v}{4\epsilon}(\tilde{r} + \tilde{d}_j + v)\right] + (1-\mu_j)\left[\frac{\tilde{r} - \tilde{d}_j}{4\epsilon}(\tilde{r} + \tilde{d}_j)\right]$$

In order to get the optimal location the party would choose ceteris paribus, all which is left to do is differentiate with respect to d_j, which rather straightforwardly yields:

$$\tilde{d}_j = 0 \Leftrightarrow d_j = \frac{1}{2} - \epsilon \equiv d^*$$

□

Derivation of Condition 1. Separation is feasible only if the party is willing to accept the new platform, which despite being closer to its bliss point, wins the general election with a smaller probability. Using the expression for the probability of winning the general election of a Democratic candidate, $\frac{1}{2} + \frac{1}{4\epsilon}(d + r + v - 1)$, the probability of winning the general election for a non valent candidate located at $d^* = \frac{1}{2} - \epsilon$ can be expressed as $\frac{\tilde{r}}{4\epsilon}$, whereas that for a valent candidate located at $d_H = d^* - \tilde{d}_H$ the probability is $\frac{\tilde{r} - \tilde{d} + v}{4\epsilon}$. Comparing expected utilities for a party voter, the condition for a party voter to prefer the valent politician who is signalling reads:

$$\frac{1}{4\epsilon}(\tilde{r} - \tilde{d}_H + v)(\tilde{r} + \tilde{d}_H + v) - r \geq \frac{1}{4\epsilon}\tilde{r}^2$$

which can be rearranged to:

$$\bar{d}_H \leq \sqrt{(v + \tilde{r})^2 - \tilde{r}^2}.$$

Now, substituting for $\tilde{d}_H = \frac{1}{2-\alpha}\tilde{r}$, dividing by $\tilde{r}^2$ and solving the resulting quadratic equation for $\tilde{\rho}$ we obtain the condition in the statement. Notice that if this condition does not hold, there are two possibilities. If $\frac{\tilde{r}}{2} > \sqrt{(v + \tilde{r})^2 - \tilde{r}^2}$, then the game has a pooling equilibrium at d^*. Otherwise, pooling is not sustainable and no equilibrium (at least in pure strategies) exists. □

Proof of Proposition 2. Assume that the screening party wants to implement a separating equilibrium. Then it has to maximize the following expression:

$$(1-\alpha)^2 \left[\frac{1}{4\epsilon}(2\epsilon - 1 + d_L + r)(r - d_L) - r\right] + [1-(1-\alpha)^2]\left\{\frac{1}{4\epsilon}(2\epsilon - 1 + d_H + r + v)(r - d_H + v)\right\}$$

subject to

$$d_H = \frac{1-\alpha}{2-\alpha} d_L + \frac{1}{2-\alpha}(1-r-2\epsilon)$$

which becomes

$$(1-\alpha)^2 \left[\frac{1}{4\epsilon}(2\epsilon - 1 + d_L + r)(r - d_L) - r\right] + [1-(1-\alpha)^2]\times$$
$$\times \left\{\frac{1}{4\epsilon}(2\epsilon - 1 + \frac{1-\alpha}{2-\alpha}d_L + \frac{1}{2-\alpha}(1-r-2\epsilon) + r + v)(r - \frac{1-\alpha}{2-\alpha}d_L - \frac{1}{2-\alpha}(1-r-2\epsilon) + v)\right\}.$$

Differentiating with respect to d_L and multiplying by the constant 4ϵ we obtain the following:

$$(1-\alpha)^2[r - d_L - 2\epsilon + 1 - r - d_L] + [1-(1-\alpha)^2]\frac{1-\alpha}{2-\alpha}\times$$
$$\times \left[r + v - \frac{1-\alpha}{2-\alpha}d_L - \frac{1}{2-\alpha}(1-r-2\epsilon) - \left(2\epsilon - 1 + \frac{1-\alpha}{2-\alpha}d_L + \frac{1}{2-\alpha}(1-r-2\epsilon) + r + v\right)\right]$$

In order to find the maximum (notice that the second order condition is negative), equate the derivative to zero and rearrange to obtain:

$$d_L^{scr} = \frac{1}{2} - \epsilon + \frac{\alpha}{2(1-\alpha)}\left(r - \left(\frac{1}{2} - \epsilon\right)\right) > d^*$$

As one can see, the optimal screening policy for the low type is more moderate than that resulting from the signalling model, which is $d^* = \frac{1}{2} - \epsilon$. Moreover, in order to conclude whether it is optimal to separate types or not, notice that the utility the party derives from playing a separating equilibrium is:

$$-r + \frac{1}{4\epsilon}\left[(2\epsilon - 1 + r + d_L)(r - d_L)\right](1-\alpha)^2 + \frac{1}{4\epsilon}\alpha(2-\alpha)\left[(2\epsilon - 1 + r + d_H + v)(r - d_H + v)\right]$$

whereas the utility arising from a pooling equilibrium is

$$-r + \alpha\left[\frac{1}{4\epsilon}(r - \frac{1}{2} + \epsilon + v)(r - \frac{1}{2} + \epsilon + v)\right] + \frac{1-\alpha}{4\epsilon}\left[(r - \frac{1}{2} + \epsilon)(r - \frac{1}{2} + \epsilon)\right]$$

This reduces to the inequality:

$$\alpha(2-\alpha)v^2 + 2\alpha(2-\alpha)\tilde{r}v - \frac{\alpha(4-\alpha)}{4}\tilde{r}^2 \geq \alpha v^2 + 2\alpha v\tilde{r}$$

which finally leads to the condition

$$\frac{v}{\tilde{r}} \equiv \tilde{\rho} \geq \sqrt{1 + \frac{4-\alpha}{4(4-\alpha)}} - 1$$

under which screening is preferred to pooling. □

Derivation of Condition 2. This condition determines the parameter values for which a separating equilibrium is feasible, in the sense that the party voter prefers the separating high valence candidate at d_H^{DP} (or r_H^{DP}) to a low valence candidate at d^* (or r^*), when having the choice. I consider an interim condition, meaning that when choosing the candidate, the party voter does not know the valence and location of the opponent yet. The expected utility the party voter gets from a valent candidate at d_H^{DP} is the following:

$$U(d_H^{DP}, v) = \alpha(2-\alpha)\left[\frac{1}{2}(2\epsilon + \tilde{d}_H^{DP}) - \frac{1}{2}\tilde{d}_H^{DP} + v\right] + (1-\alpha)^2\left[\left(\frac{1}{2} - \frac{\tilde{d}_H^{DP} - v}{4\epsilon}\right)\left(2\epsilon + \tilde{d}_H^{DP} + v\right)\right] - r^*$$

On the other hand, if the party voter chooses the non-valent candidate, her expected utility is:

$$U(d^*, 0) = \alpha(2-\alpha)\left[\left(\frac{1}{2} + \frac{\tilde{d}_H^{DP} - v}{4\epsilon}\right)2\epsilon + \left(\frac{1}{2} - \frac{\tilde{d}_H^{DP} - v}{4\epsilon}\right)(v - \tilde{d}_H^{DP})\right] + (1-\alpha)^2\left[\frac{1}{2}2\epsilon\right] - r^*$$

The former can be simplified to (I drop the exponent DP from the notation for the d location):

$$U(d_H, v) = \frac{1}{2} + v - (1-\alpha)^2\frac{\tilde{d}_H^2}{4\epsilon} + (1-\alpha)^2\frac{v^2}{4\epsilon}$$

whereas the utility from the low valence candidate can be written as:

$$U(d^*, 0) = -\frac{1}{2} + \alpha(2-\alpha)\frac{\tilde{d}_H^2}{4\epsilon} + \alpha(2-\alpha)\frac{v^2}{4\epsilon} - 2\alpha(2-\alpha)\frac{v\tilde{d}_H}{4\epsilon}$$

In order for separation to be feasible, it has to hold that $U(d_H, v) \geq U(d^*, 0)$, which requires:

$$v + [2(1-\alpha)^2 - 1]\frac{v^2}{4\epsilon} - \frac{\tilde{d}_H^2}{4\epsilon} + 2\alpha(2-\alpha)\frac{v\tilde{d}_H}{4\epsilon} \geq 0$$

Using the fact that $\tilde{d}_H = \frac{2\epsilon}{(1-\alpha)(2-\alpha)} - \frac{\alpha}{1-\alpha}v$ one obtains the following expression in terms of $\rho \equiv \frac{v}{2\epsilon}$:

$$\underbrace{\left[2(1-\alpha)^2 - 1 - \frac{\alpha^2}{(1-\alpha)^2} - \frac{2\alpha^2(2-\alpha)}{1-\alpha}\right]}_{A(\alpha)}\rho^2 + 2\underbrace{\left[1 + \frac{\alpha}{(2-\alpha)(1-\alpha)^2} + \frac{\alpha}{1-\alpha}\right]}_{B(\alpha)}\rho +$$

$$-\underbrace{\frac{1}{(1-\alpha)^2(2-\alpha)^2}}_{C(\alpha)} \geq 0$$

□

Proof of Proposition 3. The proof is in many aspects analogous to the one for the single-primary game. In particular, the argument according to which non-valent politicians locate at d^* and $1-d^*$ respectively is exactly the same as that in the single-primary game. I will therefore focus here on the incentive compatibility constraint, i.e. the derivation of the separating policy for the valent politician. The constraint for the D side of the game, cancelling out $\frac{1}{4\epsilon}$, writes:

$$\left(1-\frac{\alpha}{2}\right)\left[(1-\alpha)^2(2\epsilon - 1 + d_H + r_L) + \alpha(2-\alpha)(2\epsilon - 1 + d_H + r_H - v)\right] \leq$$
$$\leq \frac{1-\alpha}{2}\left[(1-\alpha)^2(2\epsilon - 1 + d_L + r_L) + \alpha(2-\alpha)(2\epsilon - 1 + d_L + r_H - v)\right].$$

One can also write a constraint for R:

$$\left(1-\frac{\alpha}{2}\right)\left[(1-\alpha)^2(2\epsilon + 1 - d_L - r_H) + \alpha(2-\alpha)(2\epsilon + 1 - d_H - r_H - v)\right] \leq$$
$$\leq \frac{1-\alpha}{2}\left[(1-\alpha)^2(2\epsilon + 1 - d_L - r_L) + \alpha(2-\alpha)(2\epsilon + 1 - d_H - r_L - v)\right].$$

Letting the constraints hold with equality and substituting the second into the first yields the following expression:

$$d_H = \frac{1+\alpha^3-2\alpha^2}{(2-\alpha)(1-\alpha)(1+\alpha)}d_L - \frac{1}{(2-\alpha)(1+\alpha)}r_L + \frac{1}{(2-\alpha)(1+\alpha)} - \frac{2\epsilon}{(2-\alpha)(1-\alpha)} + \frac{\alpha}{1-\alpha}v$$

However, notice that in any separating equilibrium, it has to be that $d_L = \frac{1}{2} - \epsilon$ and $r_L = \frac{1}{2} + \epsilon$. Substituting for these values in the expression above yields:

$$\frac{1}{2} - \epsilon - d_H = \frac{2\epsilon}{(2-\alpha)(1-\alpha)} - \frac{\alpha}{1-\alpha}v$$

and noticing that by definition $\frac{1}{2} - \epsilon - d_H^{DP} = \tilde{d}_H^{DP}$, one obtains the separating condition stated in the theorem.

□

Proof of Proposition 4. Given that the characterization of equilibria involves rather complicated expressions, I will derive the expected utility expressions and then plot them to show the three different regions of equilibria in the (α, ρ) space. I begin by finding the expected utility for each party (since the equilibrium is symmetric, expected utility is the same for both parties) in the no-primary equilibrium and in the double primary equilibrium. $U(P, P)$ denotes the utility from the double-primary equilibrium and $U(NP, NP)$ the utility from the no-primary equilibrium. The utilities are written in terms of the D party, but given symmetry analogous expressions hold for the R party. Denote by $r^* \equiv \frac{1}{2} + \epsilon$:

$$U(NP, NP) = -r^* + \alpha^2 \left[\frac{1}{2}2\epsilon + v\right] + (1-\alpha)^2 \left[\frac{1}{2}2\epsilon + v\right] + \\ + \alpha(1-\alpha) \left[\left(\frac{1}{2} + \frac{v}{4\epsilon}\right)(2\epsilon + v)\right] + \alpha(1-\alpha) \left[\left(\frac{1}{2} + \frac{v}{4\epsilon}\right) v + \left(\frac{1}{2} - \frac{v}{4\epsilon} 2\epsilon\right)\right]$$

Taking out ϵ and noting that $-r^* + \epsilon = \frac{1}{2}$ this can simplified to yield: $-\frac{1}{2} + [\alpha^2 + 2\alpha(1-\alpha)(\frac{1}{2} + \frac{v}{4\epsilon})]$ which can be further simplified to yield:

$$\frac{1}{2} + \alpha v - 2\alpha(1-\alpha)\frac{v^2}{4\epsilon}$$

Notice that this can also be seen as the following:

$$-\frac{1}{2} + [1 - (1-\alpha)^2]v - 2\alpha(1-\alpha)\left(\frac{1}{2} - \frac{v}{4\epsilon}\right)v$$

which can be interpreted as the utility from the choice of a social planner located at $\frac{1}{2}$ minus the loss in valence due to the random location of the median voter. It can be shown that any symmetric equilibrium provides an expected utility of this form: however, for the sake of completeness, I will nonetheless derive the expected utility from the primary equilibrium from scratch:

$$
\begin{aligned}
U(P,P) = & - r^* + \alpha^2(2-\alpha)^2 \left[\frac{1}{2}(-\tilde{d}_H^{DP}) + \frac{1}{2}(2\epsilon + \tilde{d}_H^{DP}) + v\right] + (1-\alpha)^4 \left[\frac{1}{2}2\epsilon\right] + \\
& + \alpha(2-\alpha)(1-\alpha)^2 \left[\left(\frac{1}{2} + \frac{\tilde{d}_H^{DP} - v}{4\epsilon}\right) 2\epsilon + \left(\frac{1}{2} - \frac{\tilde{d}_H^{DP} - v}{4\epsilon}\right)(-\tilde{d}_H^{DP} + v)\right] + \\
& + (1-\alpha)^2\alpha(2-\alpha) \left[\left(\frac{1}{2} - \frac{\tilde{d}_H^{DP} - v}{4\epsilon}\right)(-\tilde{d}_H^{DP} + v)\right]
\end{aligned}
$$

This can be simplified to the following expression:

$$
-\frac{1}{2} + [1-(1-\alpha)^4]v - 2\alpha(2-\alpha)(1-\alpha)^2\left(\frac{1}{2} + \frac{\tilde{d}_H^{DP} - v}{4\epsilon}\right)v
$$

and further to:

$$
-\frac{1}{2} + [1-(1-\alpha)^4 - \alpha(1-\alpha)((1-\alpha)(2-\alpha)+1)]v + [\alpha(2-\alpha)(1-\alpha)]\frac{v^2}{4\epsilon}
$$

where I used the fact that $\tilde{d}_H^{DP} = \frac{2\epsilon}{(1-\alpha)(2-\alpha)} - \frac{\alpha}{1-\alpha}v$. I now calculate the payoff from a deviation out of the no-primary and double-primary equilibrium. The first reads:

$$
\begin{aligned}
U(P,NP) = & - r^* + (1-\alpha)^3\left(\frac{1}{2}2\epsilon\right) + \alpha^2(2-\alpha)\left[\left(\frac{1}{2} - \frac{\tilde{d}_H}{4\epsilon}\right)(2\epsilon + \tilde{d}_H)\right] + \\
& + \alpha(1-\alpha)^2\left[\left(\frac{1}{2} + \frac{v}{4\epsilon}\right)v + \left(\frac{1}{2} - \frac{v}{4\epsilon}\right)2\epsilon\right] + \\
& + \alpha(2-\alpha)(1-\alpha)\left[\left(\frac{1}{2} - \frac{\tilde{d}_H - v}{4\epsilon}\right)(2\epsilon + \tilde{d}_H + v)\right]
\end{aligned}
$$

which can be simplified to:

$$
-\frac{1}{2} + \alpha(1-\alpha)^2\frac{v^2}{4\epsilon} + \alpha^2(2-\alpha)\left[v - \frac{\tilde{d}_H^2}{4\epsilon}\right] + \alpha(2-\alpha)(1-\alpha)\left[v + \frac{v^2}{4\epsilon} - \frac{\tilde{d}_H^2}{4\epsilon}\right]
$$

and finally to:

$$
-\frac{1}{2} + \alpha\left[(2-\alpha) + \frac{\alpha}{2}\right]v + \alpha\left[(1-\alpha)(3-2\alpha) - \frac{\alpha}{2-\alpha}\right]\frac{v^2}{4\epsilon} - \frac{\alpha}{2-\alpha}\epsilon
$$

Finally, the utility from a deviation out of the primary equilibrium can be written as:

$$U(NP,P) = -r^* + (1-\alpha)^3\left[\frac{1}{2}2\epsilon\right] + \alpha^2(2-\alpha)\left[\left(\frac{1}{2}-\frac{\tilde{d}_H}{4\epsilon}\right)(-\tilde{d}_H) + \left(\frac{1}{2}+\frac{\tilde{d}_H}{4\epsilon}\right)2\epsilon + v\right] +$$
$$+\alpha(1-\alpha)^2\left[\left(\frac{1}{2}+\frac{v}{4\epsilon}\right)(2\epsilon+v)\right]$$
$$+\alpha(1-\alpha)(2-\alpha)\left[\left(\frac{1}{2}-\frac{\tilde{d}_H - v}{4\epsilon}\right)(\tilde{d}_H+v) + \left(\frac{1}{2}+\frac{\tilde{d}_H - v}{4\epsilon}\right)2\epsilon\right]$$

This can be simplified to get:

$$-\frac{1}{2}+\alpha^2(2-\alpha)\left[\frac{\tilde{d}_H^2}{4\epsilon}+v\right]+\alpha(1-\alpha)^2\left[\frac{v}{2}+\frac{v}{2}+\frac{v^2}{4\epsilon}\right]+\alpha(1-\alpha)(2-\alpha)\left[\frac{\tilde{d}_H^2}{4\epsilon}+\frac{v\tilde{d}_H}{4\epsilon}+\frac{v\tilde{d}_H}{4\epsilon}+\frac{v^2}{4\epsilon}\right]$$

which can further simplified to yield:

$$\left[\frac{\alpha(1-\alpha)(3-2\alpha)(1-\alpha)^2+2\alpha^2(2-\alpha)(1-\alpha)^2+2\alpha^3(2-\alpha)}{(1-\alpha)^2}\right]\frac{v^2}{4\epsilon}-\frac{\alpha^2}{(1-\alpha)^2}v+\frac{\alpha}{(1-\alpha)^2(2-\alpha)}$$

The next step involves the checking for the conditions under which a no-primary and a primary equilibrium exist: a no-primary equilibrium exists if and only if $U(NP,NP) \geq U(P,NP)$; dividing both sides by 2ϵ and defining $\rho \equiv \frac{v}{2\epsilon}$ one obtains:

$$\alpha\rho - \alpha(1-\alpha)\rho^2 \leq \alpha\left[(2-\alpha)+\frac{\alpha}{2}\right]\rho + \frac{\alpha}{2}\left[(1-\alpha)(3-2\alpha)-\frac{\alpha}{2-\alpha}\right]\rho^2 - \frac{\alpha}{2(2-\alpha)}$$

which can be simplified to:

$$[(1-\alpha)(2-\alpha)(1-2\alpha)-\alpha]\rho^2 + (2-\alpha)^2\rho - 1 \leq 0$$

From the solution of this quadratic equation it can be seen that a no-primary equilibrium exists if and only if ρ is sufficiently small, with the threshold depending on α. In particular, for α sufficiently large a no-primary equilibrium always exists. I will now move on to the double-primary equilibrium. Dividing the expressions for $U(P,P)$ and $U(NP,P)$ by 2ϵ, the condition $U(P,P) - U(NP,P) \geq 0$ can be rewritten as:

$$\left[\alpha(2-\alpha)(1-\alpha) - \frac{\alpha(1-\alpha)^3(3-2\alpha)+2\alpha^2(1-\alpha)^2(2-\alpha)+2\alpha^3(2-\alpha)}{2(1-\alpha)^2}\right]\rho^2 +$$
$$+\left[1-(1-\alpha)^4-\alpha(1-\alpha)[(1-\alpha)(2-\alpha)+1]-\frac{\alpha^2}{(1-\alpha)^2}\right]\rho - \frac{\alpha}{2(2-\alpha)(1-\alpha)^2} \geq 0$$

Solving this quadratic equation produces another threshold on ρ, depending on α, above which a double-primary equilibrium is sustainable. A numerical exploration shows that this threshold is, for any $\alpha \in [0, 1]$, higher than the one below which a no-primary equilibrium is sustainable. This means that there exist values of ρ for which a symmetric pure strategy equilibrium does not exist. In this interval, two types of equilibria are possible: a symmetric equilibrium in mixed strategies (which I will not characterize) or asymmetric equilibria in which one party holds the primary and the other party doesn't. $\square$

2 Flip-flopping and Electoral Concerns

Politicians who change their mind on a policy issue are often confronted with the accusation of being flip-floppers. However, a changing environment sometimes makes policy revisions necessary. The present analysis suggests that flip-flopping on a policy issue is detrimental to a politician's reputation because it sends a bad signal on the accuracy of his information. As a result, electorally concerned politicians can have the incentive to stick to a no longer efficient policy choice in order to avoid the stigma of flip-flopping. This distorted behaviour is not only damaging in terms of policy welfare, but also in terms of a worse selection of competent politicians through elections. In this context, a single-term limit rule can improve welfare, achieving undistorted policies, although at the cost of worsened politicians' selection. At the same time, introducing some noise in voters' observation of the policy choice is always optimal, whereas providing voters with a signal about the state of the world can lead to increased distortions. These results lend themselves to an interesting interpretation in terms of ability of the media to discipline politicians.

But with Kerry the charge isn't that he's inconstant. It's that in his inconstancy he flips wrong – the far more serious charge of bad judgment.
Mickey Kaus (Slate)

Internal discussion and advice can only be withheld where disclosure of the information in question would be harmful to the frankness and candour of future discussions.
Campaign for Freedom of Information, 1997

2.1 Introduction

Consistency is one of the qualities that voters value the most in a politician. As the political scientist Fearon (1999) wrote: "*If I think of elections as a problem of choosing a competent, like-minded type not easily bought by special interests, then it makes perfect sense to be highly concerned with principledness and consistency*". As a result, voters tend to dislike politicians who change their mind on a policy issue, which is disparagingly denoted as *flip-flopping* [1]. As a matter of fact, one of the most frequent attacks used in electoral races is the allegation of being a flip-flopper. Two famous cases of presidential candidates that have considerably suffered from being viewed as flip-floppers are John Kerry and Mitt Romney. The fact that voters tend to punish politicians who flip-flop is also to a large extent confirmed by survey evidence collected by political scientists[2].

In a changing world in which politicians are constantly exposed to new information, however, changing one's mind on an issue is natural and actually the optimal thing to do in many situations: as Keynes put it, *"When the facts change, I change my mind"*. In this respect, the reputational stigma of flip-flopping can seem puzzling. In this paper I show how flip-flopping can be detrimental to a politician's reputation even in cases where ideological or private-interest related concerns are absent, i.e. situations in which changing one's mind simply reflects a change in the information available to the politician. In particular, flip-flopping is rationally punished by voters if i) the optimal policy choice is persistent ii) politicians have private information which cannot be credibly revealed to the public and iii) voters are not (fully) capable of judging the validity of a policy choice. The reason for this penalization is that policy shifts are associated with poorly informed politicians; therefore, voters trying to select well-informed (competent) politicians assign a better reputation to politicians who do not flip-flop. There is indeed evidence of such a rationale in political commentary: Jack Shafer of the media outlet Politico, for example, wrote the following [3] referring to Hillary Clinton: "*So if new or better information has been the impetus for her policy shifts, she must concede that she has a fat history of taking the wrong position in the early going and then requiring a re-do*".

If one part of the story has to do with how voters perceive policy shifts, the flip side of it has to do with how the reputational stigma attached to flip-flopping affects politicians' behaviour. Since flip-flopping is bad for a politician's electoral prospects, a strongly office-motivated politician will have the incentive to distort his behaviour in order to avoid flip-flops: my model therefore describes a form of electoral pandering which is endogenously induced by the previous action of the politician. In other words, politicians display an excess of conservatism, or postured consistency, with respect to their previous decisions.

One historical example in which such a logic seems to have played an important role is the

[1] Use of this expression dates back to at least 1890 according to The New York Times archives. Other terms used to shed negative light on the change of course of a politician are *u-turn* and *backflip*.

[2] See Tomz and Houweling (2012) and Doherty, Dowling, and Miller (2015): I will elaborate more about these and other studies in the related literature section.

[3] The full article can be found at this link: http://www.politico.com/magazine/story/2015/10/democratic-debate-hillary-clinton-flip-flop-213247

decision to start the Iraq war: George Bush and Tony Blair believed Saddam Hussein had (or was in the process of developping) weapons of mass destruction and that waging war on Iraq was necessary to stop him. When evidence pointing in the opposite direction was revealed (including intelligence reports and the work of UN inspectors), the two presidents should have realized that the threat posed by Saddam was after all not so large and reconsidered their plans to invade Iraq. However, they never performed this (beneficial) *flip-flop*. The publication of the Chilcot report in the United Kingdom has recently provided further evidence that Tony Blair's decision to maintain his support for the Iraq invasion was not in accordance with the information he had received from his intelligence sources. As The Guardian put it: "*That was the point at which the UK government could and should have said the US must count the UK out. Blair should have admitted that this was a line in the sand. But he didn't call a halt*".

Another example can be found in the Greek bailout referendum held in July 2015. The Greek prime minister, Alexis Tsipras, thought he could get the upper hand in the negotiations with the Troika by holding a referendum on the bailout proposal Greece had received: if creditors had interpreted the referendum as a threat to break away from Europe, they might have agreed to give Greece better conditions in order for example to avoid panic on the financial markets. After the referendum was called, however, it started becoming clear that European institutions were not willing to make concessions: Greece was not going to leave Europe, and Tsipras' bluff was called. Cancelling the referendum would then have been the optimal choice in order to foster new negotiations and a renewed cooperation between Greece and the European authorities. Tsipras, however, decided to proceed with the referendum, urging his supporters to vote against the (by then expired) offer. Despite the victory of the *no* front, Greece had to agree to a new bailout under conditions considered by many as punishing.

After establishing that reputational concerns lead to an insufficient amount of flip-flopping, the second part of the paper discusses some institutional design approaches to tackle the problem: the first-one is a single term limit policy. By forbidding the re-election of incumbents, such a policy eliminates all policy distortions: as a matter of fact, in this model electoral rents are the only source of misalignment between voters and politicians. At the same time, however, the single term limit forces voters to forgo all the potential gains of learning about a politician's type from his track-record. Not least, such policies require commitment, for example through constitutions.

The other institutional feature that I discuss is the media. The importance of the media in informing citizens and evaluating policies goes without explanation, but how does that affect the issue of distorted consistency highlighted in my model? I focus my attention on the two main roles played by the media: the reporting of politicians' policy choices (reporting media) and the evaluation of policy (commentator media).

Two results from the analysis of the media model stand out: first of all, fully accurate reporting of policy choices is never optimal. A noisy media can insulate the politician from the reputational stigma of flip-flopping: lies are thus crowded out by noise. Moreover, the net effect of this substitution is such that the performance of elections in selecting competent types improves. If the noise is too large, however, then the selection eventually worsens (no learning is possible when the media is fully noisy).

The non-monotonic effects of the media reporting accuracy suggest that the significantly simpler and broader access to politicians' track records made possible by improvements in technology and the rise of social media might have led to an increase in policy distortions. This result is related to the idea that transparency can be damaging for political accountability: adding a noisy reporting media to the model is in fact similar to relaxing the assumption of full action transparency. This links the present paper to the study of Prat (2005), among others. Looking at the quote from the CFOI (1997) found at the beginning of this paper [4], which also features in the article by Prat, we can see a nuance that my model seems to be particularly apt at highlighting: that is the dynamic aspect of the distortion caused by the disclosure of actions, which, the document maintains, can affect future actions.

As far as the commentator media is concerned, the analysis delivers one particularly counterintuitive result. Making the commentator better informed does not guarantee a decrease in the flip-flop avoidance distortion. If i) persistence is high ii) incompetent types have a sufficiently informative signal and iii) most politicians are competent, then increasing the level of informativeness of the commentator media can actually increase the extent to which incompetent politicians posture by avoiding flip-flops. The reason is that in those circumstances, increasing the informativeness of media commentary can substantially increase the gains from posturing and receiving an endorsement from the media; in other words, faking consistency with more informative media is a riskier lottery, but it has a larger upside, which can lead politicians to increase the distortion of their policy choices.

The main takeaway is that the returns to having more accurate media might be low if not even negative. This could have relevant implications, for example with respect to the public subsidization of media outlets.

[4] The quote reads: *Internal discussion and advice can only be withheld where disclosure of the information in question would be harmful to the frankness and candour of future discussions.*

2.2 Related Literature

This paper is related to several streams of literature. The electoral concerns model that I consider builds on Canes-Wrone, Herron, and Shotts (2001), Prat (2005) and Ashworth and Shotts (2010). Besides some differences in how the model is built, of secondary importance for the results, the main novelty of my model is to introduce an additional period in which the incumbent takes an action. This feature is central for the contribution of the paper, since it allows me to delve into the intrinsically dynamic nature of flip-flopping. Moreover, by adding a *stage-zero* to an electoral concerns model, my analysis also allows for a simple endogenous interpretation of the concept of pandering towards a popular action described in electoral concerns models. In my model, as a matter of fact, the popular action is simply the previous policy choice. In this sense, my work contributes to the political economy literature on conformity (sometimes also denoted as pandering or herding): along with the aforementioned Prat (2005), one of the seminal contributions is by Maskin and Tirole (2004) - who compare the welfare properties of representative democracy, direct democracy and judicial power; Levy (2007) considers a committee of career concerned decision makers and Frisell (2009) shows that voters' beliefs about the politician can have self-fulfilling consequences. Levy (2004), on the other hand, shows that in a similar setting also anti-herding (anti-conformism) can take place.

Connected to the idea of pandering towards a popular action is that of status-quo bias and propensity to reform: examples of models with endogenously status-quo biased politicians include Fu and Li (2014), where career-concerned policy makers undertake reform with lower than optimal probability, and Dewan and Hortala-Vallve (2014), in whose model voters learn through either the success of a reform or the information provided by a rival candidate. My work, therefore, links the concepts of pandering and status quo-bias: the past actions of the politician influence the voters' beliefs to which the politician has an incentive to conform.

As far as the discussion on media is concerned, Prat (2005) and Ashworth and Shotts (2010) are the closest references. The introduction of a commentator media sector in the model draws mostly on Ashworth and Shotts (2010): my contribution lies in the characterization of partially truthful equilibria (not just finding conditions for a truthful equilibrium) and in the comparative statics analysis which shows how increasing the accuracy of the commentator media need not decrease the distortion of politicians' behaviour. These comparative statics also innovates with respect to the paper of Gentzkow and Shapiro (2006): in their paper, the decision maker is a media outlet, which panders towards the prior belief of citizens in order to form a reputation of accuracy. Whereas Gentzkow and Shapiro show that the more likely it is for voters to learn the true state of the world, the lower the pandering by the media, I show that introducing an informative media might actually lead politicians to act in a more distorted way.

The analysis of the reporting media, on the other hand, is related to the discussion of action versus consequence transparency in Prat (2005). Since for many policy choices the assumption of action secrecy is not a realistic alternative, I show that the accuracy of the reporting media plays a similar role, and I demonstrate that the optimal arrangement is to have some but not perfect information on the action taken by the politician.

The single term limit rule I mention in the institutional design section, on the other hand, is related to the comparison between representative democracy and judicial power carried out by Maskin and Tirole (2004). A single term limit rule has similar effects to those of delegating decision making to a judicial power not subject to elections.

The consequences of reputation concerns on expert behaviour have also been studied outside the electoral environment. In particular, repeated action by experts has been analyzed by Prendergast and Stole (1996), Li (2007) and in a recent paper by Aghion and Jackson (2016): in addition to many differences in the modelling strategy, the main conceptual difference between my analysis and that of Prendergast and Stole has to do with the fact that I consider a changing environment rather than a fixed state. Aghion and Jackson, on the other hand, consider repeated action over a sequence of independent and identically distributed states (whereas I allow for correlation) and with action consequences being observable to the principal (whereas they are unobservable in my baseline model). The defining feature of Li's paper is to assume that decision makers' information becomes more accurate in the second period. In her model, a (fully reputation motivated) agent has to do make two reports before being evaluated by the principal. The state of the world is fixed and observable before the evaluation takes place, but the agent can be paid a wage that only depends on beliefs about her ability. In such a setup it is possible for agents to improve their reputation by changing their mind, unlike in my model. In other situations, the premium for consistency leads low competence agents to gamble on being proved consistently right, similarly to what happens in my model with the commentator media. The idea of gambling on a policy likely to be proved wrong is also at the heart of the paper by Majumdar and Mukand (2004): theirs is a model of experimentation, in which an incumbent can choose to implement a risky policy and has to decide whether to continue the project after a potentially unsuccessful trial. In their model, low type governments are inefficiently reluctant to abandon bad projects, gambling on the small probability of success that would boost their reputation.

Another paper which shares part of the mechanism with my work is Patacconi and Vikander (2015), in which a policy maker receives two signals from an agency and a conflict between voters and policy-maker can arise when the signals are mixed, similarly to what happens in my model after a mixed history of signals. However, whereas in their study voters protest to get an unbiased policy enforced, in the environment I describe there are no protests but elections to select the most competent candidate.

Since my paper explains the reputational stigma of flip-flopping, it is also worth mentioning some political economy works on this topic. Whereas my paper is the first, to the best of my knowledge, to provide a theory of flip-flopping by a politician taking repeated decisions over an issue, there exists some theoretical work by Agranov (2016) and Hummel (2010) dealing with flip-flops between primary and general elections, hence with a completely different objective than that of my analysis.

On the empirical side, finally, there are several papers assessing how voters react to politicians flip-flopping. Doherty, Dowling, and Miller (2015), for example, find that flip-flopping affects the perception voters have of a politician. They show that voters are more forgiving of flip-flops on complex issues or issues which are far away in time. These predictions are consistent with my model, in which flip-flops are a bad signal the higher the state persistence (and we can think that persistence is lower over long periods) and the worse is the signal that voters have. In the same paper, the authors also show that the reputational cost of a flip-flop is compensated by the fact that the new position taken by the politician will be seen favourably by some voters, creating a trade-off for the politician. Similarly, Tomz and Van Houweling (2012) conduct survey experiments showing that candidates repositioning affects their support not only in terms of commitment to an ideological issue, but also in terms of perceived valence. The valence aspect is important in particular for issues that are not too salient for voters.

Another issue that has been studied is whether the effects of a flip-flop differ between issues of principle (as abortion or gay marriage) versus pragmatic issues related to a specific policy. Tavits (2007) shows that flip-flopping on pragmatic issues is seen less badly than flip-flopping on ideological issues. On the other hand, Tomz and Houweling (2012) do not find any difference among issues: independent of the issue, candidates who reposition perform worse. In my model there is no difference between types of issues: I assume that one decision is correct depending on the state of the world and that citizens would all agree if perfectly informed. Tomz and Houweling (2012) also argue that the bad perception of flip-flopping will deter candidates away from it: this is exactly what I happens in the formal model I present, where flip-flopping is not bad per se, but politicians avoid it because it carries a bad signal for their reputations.

Levendusky and Horowitz (2012) show experimentally that in the context of international relations, leaders who make threats and subsequently back down pay a cost in terms of electoral support and reputation (which in the international relations literature is called *audience cost*): one of the main reasons for this effect is that a leader changing his mind is seen as less competent than one who stays coherent. What is more, they show that partisanship does not play a significant role in the determination of audience costs. This evidence seems to capture a mechanism lying very close to the one I present in my model. As a matter of fact, given the presence of asymmetric information between politicians and voters and the fact that politicians are often evaluated for their foreign policy conduct before the consequences of their actions are fully known, foreign policy issues are among those where the theory I develop should have more bite.

2.3 The Model

There is an incumbent politician, a set of identical voters and two periods $t \in \{1, 2\}$. The politician's job is to take an action $a_t \in \{0, 1\}$ in each period. The action's payoff depends on the underlying state of the world, which can take two values $\omega \in \{0, 1\}$. The initial probability of each state is equal to $p_1 = \frac{1}{2}$; moreover, the state is persistent so that for $j \in \{0, 1\}$, $Pr(\omega_2 = j|\omega_1 = j) = \gamma > \frac{1}{2}$. The utility for taking the right action is normalized to 1 and it applies to both voters and politicians. At the beginning of period one, an incumbent is randomly drawn. Incumbents can be of two types: competent and incompetent. In each period, both types receive an informative signal $s_t \in \{0, 1\}$ on the state of the world, but the accuracy of the signal, $Pr(s_t = j|\omega_t = j) = q_\theta$ depends on the politician's type $\theta \in \{H, L\}$: $\frac{1}{2} < q_L < q_H \leq 1$. To simplify exposition I fix $q_H = 1$ (competent politicians perfectly observe the state) and therefore I drop the subscript from q_L, which will be simply denoted by q. Notice that since the initial prior is $\frac{1}{2}$, for all $\gamma < 1$ the signal received by any politician is always decision-relevant. A signal is decision-relevant when the probability of matching the action to the state is maximized by $a_t = s_t$. I denote by $\rho_t = Pr(\omega_t = s_t|s_t, \theta, s_{t-1})$ the posterior probability that the state is equal to the signal after observing realization s_t. Since the posterior of the perfectly informed competent politician is always equal to 1, ρ will denote, when not further specified, the posterior of the incompetent politician. In particular, ρ_2 will be used as short form for $\rho_2(\omega_2 = s_2|s_1, s_2 \neq s_1, \theta = L)$, whereas $\bar{\rho}_2$ will indicate $\rho_2(\omega_2 = s_2|s_1, s_2 = s_1, \theta = L)$.

Politicians know their competence, and the signals they receive are private information: competent politicians represent a fraction λ of incumbents. At the end of period $t = 2$ there is an election, in which the representative voter decides whether to retain (r) or fire (f) the incumbent politician (denote the decision $e \in \{r, f\}$). Voters know the statistical process governing the economy and they observe the track-record of the incumbent politician, i.e. the actions taken over the two periods, denoted by $\tau = (a_1, a_2)$. Track-records are used to form beliefs $\mu(a_1, a_2) = Pr(\theta = H|a_1, a_2)$ over a politician's competence, which I call reputation. Before the election takes place, a challenger appears. The challenger is competent with probability λ_O. Moreover, as I will describe shortly, a draw from a uniform distribution in $v \in [-b, b]$, observed before the election takes place and independent of the competence of incumbent and challenger, determines the relative valence of the challenger versus the incumbent. The representative voter's utility depends on whether the politician's action matches the state of the world (plus the valence draw in case the challenger is elected); moreover, electing a competent politician gives voters a utility of b (denote by θ_e the type of the election winner). Hence, $U_c = \sum_{t=1}^{2} 1_{a_t=\omega_t} + 1_{e=f}z + 1_{\theta_e=H}b$. Politicians derive utility both from taking the right action while in office and winning the election, with $\frac{\phi}{2}$ denoting the additional utility received if re-elected. Formally: $U_p = \sum_{t=1}^{2} 1_{a_t=\omega_t} + 1_{e=r}\frac{\phi}{2}$.

As far as players' strategies are concerned, I'll start with politicians. A politician's information set, or history, denoted by h_t, includes all actions up to $t-1$ and signals up to t. The strategy of the incumbent is a mapping Ψ from any history h_t to any probability distribution of actions $a_t \in \{0, 1\}$. In particular, since signals are decision-relevant, it is useful to express the incumbent's strategy as the probability $\sigma(s_t)$ of choosing a policy a_t in accordance to the realization at time t of the signal, denoted by s_t.

As far as the voter's strategy is concerned, it implies retaining or firing the politician in the election at the end of $t = 2$. The voter's decision is $e \in \{r, f\}$ and it depends on the incumbent's track record $\tau = (a_1, a_2)$. Voters choose each politician with probability $\frac{1}{2}$ when indifferent [5].

The equilibrium concept I use is Perfect Bayesian (PBE), but I do not consider pooling equilibria where at each time t, both types play the same action with probability 1. These equilibria are ruled out by a simple trembling-hand perfection refinement [6].

[5] Notice, however, that this is a zero probability event, since challengers are distributed according to an atomless distribution.

[6] For an example and the explanation of how the trembling-hand refinement eliminates these equilibria, refer to the Appendix.

2.4 Results

I start the analysis from the election in which the voter chooses between the incumbent and the challenger. Just before the election, the valence draw v is realized. As a result, voting for the challenger gives the voter a utility of $v + \lambda_O b$. Voting for the incumbent gives instead the voter a utility of $\mu(a_1, a_2)b$. Therefore, the incumbent is re-elected if $v \leq (\mu(a_1, a_2) - \lambda_O)b$, and $Pr(v \leq (\mu(a_1, a_2) - \lambda_O)b) = \frac{(\mu(a_1,a_2)-\lambda_O)b-(-b)}{2b} = \frac{1}{2} + \frac{\mu(a_1,a_2)-\lambda_O}{2}$. In other words, the re-election probability is linearly increasing in reputation.

Having described the election stage, go back one step and consider the decisions of the incumbent in periods $t = 1$ and $t = 2$. Signals are always decision-relevant, so maximizing the probability to match the action to the state requires that politicians follow their signals. If that happens in equilibrium, then I call the equilibrium truthful.

Definition 1. *An equilibrium is truthful if and only if $\sigma(s_t) = 1$ for each s_t at any t and for each type θ.*

Let's start with a simple observation about the reputation of different track records under truthful play, which I will denote by $\mu^T(a_1, a_2)$ in order to stress the fact that $a_t = s_t$. There are four possible track records, $\tau \in \{(0,0), (0,1), (1,0), (1,1)\}$. Given the symmetric initial prior, however, the probability of obtaining each of the two consistent and flip-flopping signal sequence is the same, and therefore, as the next claim will prove, $\mu^T(0,0) = \mu^T(1,1)$ and $\mu^T(0,1) = \mu^T(1,0)$. In other words, all that matters is whether the politician played the same action over the two periods or changed his mind. In particular, I will call the former *consistent* track-records and the latter *flip-flopping* track-records, and indicate them as $\tau = C$ and $\tau = F$. I will therefore indicate with μ^T_C and μ^T_F consistent and flip-flopping track-records under truthful play.

Fact 1. *Under truthful play, the reputation of a consistent track-record is strictly larger than that of a flip-flopping track-record: $\mu^T_C > \mu^T_F$.*

Proof. Take any flip-flopping track record. Under truthful play, $a_t = s_t$, so that $Pr(a_2, a_1|\theta) = Pr(s_2, s_1|\theta)$. Denote by $\frac{A(q_\theta)}{2}$ the probability that a politician receives a flip-flopping sequence of signals:

$$\frac{A(q_\theta)}{2} = \frac{1-\gamma}{2}(q_\theta^2 + (1-q_\theta)^2) + \frac{\gamma}{2} 2q_\theta(1-q_\theta).$$

First notice that since $p_1 = \frac{1}{2}$, the above expression holds for both types of flip-flopping sequences of signals $(0,1)$ and $(1,0)$. It can be noted that $\frac{A(q_\theta)}{2} = \frac{1-\gamma}{2} + \frac{2\gamma-1}{2} 2q_\theta(1-q_\theta)$

and since $\gamma > \frac{1}{2}$, $2\gamma - 1 > 0$ and $A(q)$ is decreasing in q. Hence, $A(1) = 1 - \gamma < A(q)$. Finally, let's construct the reputations from the 4 possible track-records. Since $\frac{A(q)}{2}$ only depends on whether the politician received a consistent or flip-flopping sequence of signals, there are only two possible levels of reputation: one from flip-flopping and one from being consistent. To see that reputation from flip-flopping is lower than from consistent play consider the following inequality:

$$\mu_F^T = \frac{\lambda A(1)}{\lambda A(1) + (1-\lambda)A(q)} < \frac{\lambda(1 - A(1))}{\lambda(1 - A(1)) + (1-\lambda)(1 - A(q))} = \mu_C^T$$

From now on, I will simply write $1 - \gamma$ for $A(1)$ and denote by $A \equiv A(q)$. □

The result I just stated is very important for the development of the whole paper since it highlights the reason that leads incumbents to distort their actions: the bad reputation associated with flip-flopping. The stigma of flip-flopping, as a matter of fact, puts an office motivated incumbent in front of a trade off. Whenever receiving a second signal that contradicts the first one, i.e. $s_2 \neq s_1$, a politician knows that doing the right thing for society will get him a worse reputation. If the flip-flopping stigma is large enough compared to the benefit from following his signal, the politician will choose to act contradicting his information.

In light of this, it can be shown that a necessary and sufficient condition for the existence of a truthful equilibrium is that incumbents have the incentive to follow their signals after receiving a non-constant stream of signals. Moreover, a single crossing property makes it sufficient to simply look at the incentives of incompetent incumbents. The reason is that whereas both competent and incompetent politicians enjoy the same benefit from avoiding a flip-flop, they do not sustain the same costs: having a less accurate signal, as a matter of fact, means having a larger probability of matching the state of the world when playing $a_t \neq s_t$. Lying is therefore cheaper for incompetent politicians. This property is very useful for the characterization of the equilibrium.

Fact 2 (**Single Crossing Property**). *The cost of acting against one's signal is* $\rho_t(\theta) - (1 - \rho_t(\theta)) = 2\rho_t(\theta) - 1$, *and since* $\rho_t(\theta = L) < \rho_t(\theta = H) = 1$, *contradicting the signal is more costly for the competent politician.*

It follows that a truthful equilibrium is only sustainable as long as incompetent politicians are willing to follow their signal at $t = 2$ after receiving a stream of non-constant signals. This in turn requires the office motivation parameter to be low enough, because as ϕ grows larger, the benefits from holding office progressively dwarf the utility from matching the action to the state of the world.

Proposition 1. *A truthful equilibrium is sustainable as long as* $\phi \leq \bar{\phi}$*, where*

$$\bar{\phi} = \frac{2\rho_2 - 1}{\mu_C^T - \mu_F^T}$$

Proof. See Appendix 2.A. □

The parameter $\bar{\phi}$ is therefore the upper-bound on electoral rents[7] under which a truthful equilibrium is sustainable. The question now is: what happens when ϕ is larger than $\bar{\phi}$? Moreover, does the game allow for a unique equilibrium? The answer to these questions is given by the following theorem:

Theorem 1. *The game always has a unique non-pooling Perfect Bayesian Equilibrium. For* $\phi \leq \bar{\phi}$*, the unique equilibrium is the truthful equilibrium. For* $\phi > \bar{\phi}$*, the unique equilibrium is partially truthful, meaning that:*
$\sigma(s_1) = \sigma(s_2|\theta = H) = \sigma(s_2 = a_1|\theta = L) = 1$ *and* $\sigma(s_2 \neq a_1|\theta = L) = \sigma^* < 1$.

Proof. See Appendix 2.A. □

This theorem proves that the game always has a unique non-pooling equilibrium. It must therefore be the case that when $\phi \leq \bar{\phi}$, the unique equilibrium is truthful. When $\phi > \bar{\phi}$, on the other hand, the unique equilibrium is only partially truthful, since whenever the signal received by politicians in the second period suggests flip-flopping, incompetent politicians mix between following their signal and pandering towards their previous action.

Theorem 1 has two interesting implications: the first is that for any level of ϕ, flip-flopping always decreases the incumbent's reputation (since the two reputations average at λ, a bad reputation is always below λ). However, the larger the flip-flopping avoidance distortion caused by incompetent politicians not following their signal, the smaller is the reputation gap between consistent play and flip-flopping, i.e. the less bad is the reputation from flip-flopping. As ϕ tends to infinity and politicians only care about re-election, the reputation gap between consistency and flip-flopping approaches zero.

The second (and related) implication is that when the equilibrium is partially truthful, there is an insufficient amount of flip-flopping compared to the truthful equilibrium. In other words, voters stigmatize flip-flopping but at the same time they would be better off if more flip-flopping took place. As a matter of fact, it could even happen that voters are more confident about the policy being correct after seeing a flip-flop rather than consistent policy: a flip-flop sends a bad signal on the incumbent's type but is always earnest, whereas

[7] The actual rents for the politician are $\frac{\phi}{2}$, but that only serves the purpose of simplifying calculations and has no other consequence on the model.

a consistent policy is a good sign on the politician's type but not necessarily earnest. I summarize these insights in the following corollary:

In equilibrium, flip-flopping gives a bad reputation compared to truthful play: $\mu_C^* - \mu_F^* > 0$ for any ϕ. Equivalently, $\mu_C^* > \lambda > \mu_F^*$. In a partially truthful equilibrium, there is insufficient flip-flopping compared to the truthful equilibrium.

Notice that another interesting implication of the model is that change hurts incumbents: when the state of the world changes, the reputation of incumbents is likely to fall (because of flip-flopping) and this means the incumbent is more likely to get replaced. At the same time, this also means that conditional on a change in leader, leaders who rise to power after a change in the state are worse than those who get in office after a period of stability. On the other hand, conditional on having a bad leader in office, a change in state increases the chances of having a better leader in the following period. In other words, improvements in leadership are more likely after a change in the state of the world.

Fact 3. *A leadership change is more likely after the state of the world changed. Conditional on a leadership change, leaders who rise to power after a change in the state of the world are worse; however, a change in the state of the world increases the chance of having a better leader in the following period.*

Proof. See Appendix 2.A. □

2.4.1 Comparative Statics and Welfare

First of all, the definition of welfare in this model is based on the expected utility of voters, calculated as of time $t = 0$ (i.e. before randomly picking the incumbent); as such, welfare does not account for the utility of politicians [8].

Definition 2. *Social welfare W is defined as:*

$$W = \mathbb{E}_0 \left[\sum_{t=1}^{2} 1_{a_t = \omega_t} \right] + \mathbb{E}_0[v|e = f] + 1_{\theta_e = H} b$$

Welfare can be decomposed in two parts, accountability and selection. Accountability indicates whether the incumbent acts in the best possible manner for society, which in this case means to follow the signal: accountability welfare is therefore the probability that the incumbent chooses the optimal policy, formally $\sum_{t=1}^{2} 1_{a_t=\omega_t}$. Selection-welfare, on the

[8] This is standard in political economy, as we consider politicians too small a fraction of the population for their welfare to matter in the aggregate.

other hand, indicates the utility the voter derives from the election winner, also accounting for the valence shock: formally $1_{\theta_e=H}b + 1_{e=f}v$.

Let's denote by $d \equiv \lambda - \lambda_O$ and by $\tilde{q} = q - A(1-\sigma^*)(2\rho_2 - 1)$ the accuracy of the incompetent's politician policy choice, taking into account that not following the signal changes its accuracy across the two states of the world. The expression for welfare can be rewritten in the following way:

$$W = [\lambda + (1-\lambda)q] + [\lambda + (1-\lambda)\tilde{q}] + b\left[\frac{1}{4} + \frac{\lambda_O^2}{4} - \frac{\lambda\lambda_O}{2} + \frac{\lambda+\lambda_O}{2} + \frac{\mathbb{E}\mu^2}{4}\right] \tag{2.1}$$

Proof. As far as $t = 1$ and $t = 2$ are concerned, since the competent politician follows his signal, which is perfectly accurate, he always takes the right decision. The incompetent politician, instead, follows the signal in the first period, taking the right decision with probability q, but in the second period, if the signal indicates flip-flopping as the optimal action, he contradicts it with probability $1-\sigma^*$. As a result, the accuracy of the incompetent incumbent's signal is $(1-A)\bar{\rho}_2 + A(\sigma^*\rho_2 + (1-\sigma^*)(1-\rho_2)) = q - (1-\sigma^*)(2\rho_2 - 1)A$, to get which I used the fact that $(1-A)\bar{\rho}_2 + A\rho_2 = q$, since $q = Pr(\omega = s_1|s_1)$. As far as the valence draw is concerned, the expression for expected welfare is the following: $\mathbb{E}[v|e = f] = \sum_{\tau\in\{C,F\}} Pr(\tau)Pr(e = f|\tau)\mathbb{E}[v|v\ b(\mu_\tau - \lambda_O)]$. Now, $\mathbb{E}[v|v\ b(\mu - \lambda_O)] = \frac{b}{2}[1 + (\mu_F - \lambda_0)]$ and $Pr(e = f|\mu) = \frac{1}{2} - \frac{\mu_F - \lambda_0}{2}$. So $\mathbb{E}[v|e = f] = \mathbb{E}_\mu \frac{b}{4}[1 - (\mu - \lambda_O)^2] = \mathbb{E}_\mu \frac{b}{4}[1 - d^2]$. Finally, $Pr(e = f) = \mathbb{E}_\mu\left[\frac{1}{2} - \frac{\mu - \lambda_O}{2}\right]$ whereas $Pr(e = r|\theta = H) = \frac{1}{2} - \frac{\lambda_O}{2} + \frac{Pr(\tau=C|H)\mu_C + Pr(\tau=F|H)\mu_F}{2}$, where $Pr(\tau = C|H) = \gamma$ and $Pr(\tau = F|H) = 1-\gamma$. Taking expectations with respect to μ where necessary, and considering that $\mathbb{E}\mu = Pr(\tau = C)\mu_C + Pr(\tau = F)\mu_F = \lambda$, the welfare expression can be rewritten as W=$[\lambda + (1-\lambda)q] + [\lambda + (1-\lambda)\tilde{q}] + \frac{b}{4}[1 - \mathbb{E}_\mu(\mu - \lambda_O)^2] + b\lambda_O\left(\frac{1}{2} - \frac{\lambda - \lambda_O}{2}\right) + b\lambda\left[\frac{1}{2} - \frac{\lambda_O}{2} + \frac{\gamma\mu_C + (1-\gamma)\mu_F}{2}\right]$. Consider now the expression $\mathbb{E}_\mu\mu^2$, where the expectation is taken over the distribution of μ:

$$\mathbb{E}\mu^2 = Pr(\tau = C)\mu_C^2 + Pr(\tau = F)\mu_F^2$$

Using the fact that $Pr(\tau = C) = \frac{\lambda\gamma}{\mu_C}$ and $Pr(\tau = F) = \frac{\lambda(1-\gamma)}{\mu_F}$ I can rewrite the former expression to get:

$$\mathbb{E}\mu^2 = \lambda(\gamma\mu_C + (1-\gamma)\mu_F).$$

This is very useful to simplify the result since we can see that the last term is in fact equal to: $\frac{b}{2}\mathbb{E}\mu^2$. Doing this substitution and multiplying out all the terms, the welfare expression can finally be expressed as (2.1).

□

Since from the analysis of the equilibrium I get a condition on $\mu_C - \mu_F$, however, it is useful to know whether the gap between the good reputation from consistent play and the bad reputation from flip-flopping is sufficient to conclude something on selection welfare. The following lemma tackles this question:

Lemma 1. *Selection welfare increases when μ_C and the reputation gap $\mu_C - \mu_F$ increase.*

Proof. See Appendix 2.A. □

Lemma 1 will be helpful to prove some of the welfare results, since it tells us that whenever the reputation gap gets larger, knowing that the reputation from consistent play increased is enough to tell that selection welfare increased. As a result, let's now move on to stating how variations in the model parameters affect welfare.

Fact 4. *Increasing ϕ weakly decreases welfare.*

Proof. When ϕ increases, the benefits from office increase and this increases the accountability distortion. This happens because $\mu_C - \mu_F$ is an increasing function of σ; when ϕ increases, therefore, σ^* decreases up to the point where $\mu_C - \mu_F$ is equal to the new costs of deviating for the signal. This reasoning can be easily verified from the equilibrium condition, noting that the right-hand side increases as σ^* increases:

$$2\rho_2 - 1\phi = \frac{\lambda\gamma}{\lambda\gamma + (1-\lambda)(1-A\sigma^*)} - \frac{\lambda(1-\gamma)}{\lambda(1-\gamma) + (1-\lambda)A\sigma^*}$$

The decrease in σ^* negatively impacts welfare in the second period. In terms of selection of politicians, since a lower σ^* decreases μ_C and increases μ_F, then from Lemma 1 we know that this means that the second moment will decrease, hence also selection welfare worsens. □

It has to be kept in mind, of course, that this model abstracts from all those reasons why it might be a good idea to offer electoral incentives to politicians (for example to improve the share of competent politicians in the pool); as a matter of fact, in this model politicians' interests are aligned with those of citizens except for electoral incentives.

Another conclusion from the analysis is that having a wide competence gap between the two politicians' types is bad for accountability: as a result, increasing q always increases welfare, both because politicians pander less (i.e. σ^* increases) and because incompetent politicians have better information when they choose a policy.

Fact 5. *As long as $\sigma^* < 1$, increasing q strictly increases welfare.*

Proof. Assume that q increases to q'. An increase in q moves $A = (1-\gamma)(q^2 + (1-q)^2) + 2\gamma q(1-q)$ down and $\frac{2\rho_2 - 1}{\phi}$ up. As long as both $\sigma^*(q)$ and $\sigma^*(q')$ are strictly less than 1, then in equilibrium $\frac{2\rho_2-1}{\phi} = \mu_C - \mu_F$ and therefore an increase of q to q' leads to a larger equilibrium value of $\mu_C - \mu_F$. Moreover, since a decrease in A lowers μ_C and raises μ_F ceteris paribus, σ^* has to increase in order to make $\mu_C - \mu_F$ larger. In particular, it has to be the case that $A\sigma^*$ increases. It follows that $\mu_C(q') > \mu_C(q)$, and as a result from Lemma 1 we know that selection welfare improves. In terms of accountability welfare, $\sigma^*(q') > \sigma^*(q)$ and $q' > q$, so not only incompetent politicians are better, but they also act in a less distorted way. Hence, accountability welfare and therefore total welfare increases. Notice that once in a truthful equilibrium, an increase in q decreases $\mu_C - \mu_F$, since A keeps decreasing but σ cannot increase any further. As a result, the second moment decreases and the probability of having a competent politician in office in the second period decreases. However, bad politicians are better so the effect on selection welfare is ambiguous.

□

Notice that when q is high, i.e. politicians are in general competent, then μ_F is lower, i.e. flip-flopping hurts more. This might be one of the reasons why flip-flopping can hurt candidates in a race such as the US presidential election, despite the fact that electoral incentives are high. In other words, flip-flopping is worse for a candidate's reputation with more homogeneity between competent and incompetent candidates [9].

When λ increases, both μ_C and μ_F increase. However, $\mu_C - \mu_F$ can either increase or decrease. As a result, σ^* can also move in either direction. This means that there exist cases in which having a better pool of politicians increases the distortion generated by incompetent politicians avoiding flip-flops. In terms of welfare from the selection of politicians, however, an increase in λ is always beneficial, at least whenever the starting point is a partially truthful equilibrium. The reason is that the equilibrium level of $\mu_C - \mu_F$ remains constant. This means that both μ_C and μ_F increase, and from Lemma 1 we know that this means the selection of politicians improves.

Fact 6. *Increasing λ in a partially truthful equilibrium improves selection welfare, but it can either increase or decrease accountability welfare.*

Proof. Assume that the game has a partially truthful equilibrium. The change in λ will therefore lead to either a truthful equilibrium or the game will remain in a partially truthful

[9] Notice that a bad reputation in this setup is just the probability of being of the incompetent type, no matter how bad the incompetence is.

equilibrium. In the first scenario, by definition the policy distortion decreases, because σ^* increases to 1. Moreover, given that $\mathbb{E}\mu$ and since in a truthful equilibrium (compared to a partially truthful one) $\mu_C - \mu_F$ is larger, then welfare improves, given that also μ_C increases for sure, given that $\mathbb{E}\mu$ has to increase.

Let's now consider the case in which the game remains in a partially truthful equilibrium. Accountability welfare can either increase or decrease. The reason is that, denoting by $D_C \equiv \lambda\gamma + (1-\lambda)(1-A\sigma)$ and $D_F = \lambda(1-\gamma) + (1-\lambda)A\sigma$, the expression $\frac{\partial(\mu_C - \mu_F))}{\partial\lambda} = \frac{A\sigma^*}{1-\gamma}\frac{1}{D_F^2} - \frac{1-A\sigma^*}{\gamma}\frac{1}{D_C^2}$ can move in both directions following an increase in λ. As a consequence, σ^* can either increase or decrease in order to move $\mu_C - \mu_F$ back to the equilibrium level. In other words, despite more competent politicians being there, it is possible for the increasingly distortive behaviour of incompetent politicians to decrease accountability welfare.

In terms of selection welfare, on the other hand, the average μ increases despite $\mu_C - \mu_F$ remaining constant. This means that either μ_C and μ_F increase, or that at least $Pr(\tau = C)$ has to increase. However, we can check using $Pr(\tau = C) = \frac{\lambda\gamma}{\mu_C}$ that μ_C has to increase, because otherwise both $Pr(\tau = C)$ and $Pr(\tau = F)$ would increase, which is a contradiction. It follows by Lemma 1 that selection welfare increases.

□

Things get more complicated when evaluating a change in the persistence parameter γ. With more persistence there are several possible cases: first of all, σ^* can be either increasing or decreasing in γ. In the former case, accountability improves with higher persistence, whereas in the latter it could go either way. In terms of selection, however, it can be shown that a more persistent state of the world decreases the effectiveness of elections in selecting competent politicians:

Fact 7. *The probability of selecting a competent politician through elections decreases as γ increases.*

Proof. When γ increases, the equilibrium value of $\mu_C - \mu_F$ decreases. From Lemma 1, we know that in such a situation selection only improves when both μ_C and μ_F increase. However, since it has to be that $Pr(\tau = C)\mu_C + Pr(\tau = F)\mu_F = \lambda$, then if both μ_C and μ_F were to increase, $Pr(\tau = F)$ would have to increase, too. However, this is not possible, because if γ increases and $Pr(\tau = F)$ also increases, then μ_F decreases. This means that selection always worsens when γ increases. □

This result is interesting since it tells us that elections perform worse in a less variable world. This can seem surprising, especially given that conditional on a leadership change, elected leaders are worse after a change in the state. However, the result is driven by the fact that whenever a bad leader is in office, a change in state increases the chances of having a competent one after the election.

Lastly, although I do not include politicians in the calculation of social welfare, it is nonetheless important to understand whether they are made better or worse off by their strategic behaviour.

Fact 8. *Incompetent politicians can be worse off in a partially truthful equilibrium in which they distort their actions compared to the truthful benchmark.*

Proof. Writing the utility of incompetent politicians under the two scenarios and using the envelope theorem we get to the following expression to evaluate:

$$\begin{aligned}(1-A)&\left[\frac{\lambda\gamma}{\lambda\gamma+(1-\lambda)(1-A)}-\frac{\lambda\gamma}{\lambda\gamma+(1-\lambda)(1-A\sigma^*)}\right]+\\ -A&\left[\frac{\lambda(1-\gamma)}{\lambda(1-\gamma)+(1-\lambda)A\sigma^*}-\frac{\lambda(1-\gamma)}{\lambda(1-\gamma)+(1-\lambda)A}\right]\end{aligned}$$

Since the result can go in both directions, it is enough to provide two examples. First of all, let's consider the case in which incompetent politicians are better off under the equilibrium with distortions. This just requires a small enough γ: it can be checked numerically that if γ is sufficiently low, then for any $\sigma^* \in [0,1)$, as well as λ and q, the expression is negative. In order to show that it is possible for the expression to be positive, check numerically that for sufficiently large γ and sufficiently low q and λ there is an interval $\sigma \in [\underline{s},1)$ in which the expression is strictly positive. Then take ϕ such that a truthful equilibrium is sustainable and increase it: since σ^* is continuous in ϕ, then for any $\hat{\sigma}$ in an ϵ-interval around 1 there exists a $\hat{\phi}$ such that $\sigma^*(\hat{\phi}) = \hat{\sigma}$. Since we know that the expression is positive at $\hat{\sigma}$, then there necessarily exist values of γ, q and λ such that the incompetent incumbent would be better off if he were able to commit to not distorting his actions. □

Another way to describe this result is that an institution forbidding incumbents to distort their actions (which would always benefit society as a whole) would also be supported by incompetent politicians when they make up a large enough share of the political class, their information is very bad and the world is more stable. This last part is particularly interesting, since it might suggest that such a reform could be easier to achieve when the environment is more stable.

2.5 Institutional Design

So far I have shown how the bad reputation from flip-flopping can give (incompetent) politicians an incentive to distort their actions. In the baseline model I have presented, there are three fundamental ingredients leading to the result: the first is the fact that politicians face an election at the end of $t = 2$, because being re-elected gives them a utility of $\frac{\phi}{2}$; the second is the fact that voters have no information on what constitutes the appropriate policy in each period and are therefore only able to evaluate incumbents based on their track record in office; third, voters are perfectly able to observe the action taken by the politician, in such a way that a flip-flop is immediately caught and used to form reputations. My aim in this section is to relax these assumptions by introducing new institutional features to the baseline model.

2.5.1 Single Term Limit

In the baseline model I analyzed above, the implicit assumption I make is that the incumbent can serve up to two terms in office, whereas the challenger can at most serve once, since the game ends after the third period. As a result, it is interesting to see what would happen under a single term limit rule, in which all politicians can serve only one term in office [10]. The single term limit for the President is an institution in several Latin American countries [11], Israel and South Korea, as well as for the head of the European Central Bank, among others. In the setup I describe in this model, the single term limit is a blunt yet effective instrument to eliminate all distortions due to politicians' willingness to avoid flip-flopping. At the same time, however, having a single term limit also means forgoing the possibility to condition the reelection decision on the beliefs about the incumbent's type. It follows that banning reelections is only welfare improving if the accountability distortion from the flip-flop avoidance is large and the upside from retaining incumbents with a good reputation is low.

Proposition 2. *A single term limit is welfare improving if and only if:*

$$b \leq 2\frac{(1-\lambda)A(1-\sigma^*)(2\rho_2 - 1)}{\frac{1+\lambda_O^2+\mathbb{E}\mu^2}{2} + \lambda - \lambda_O - \lambda\lambda_O}$$

[10] Notice that my baseline model is not exactly a model of a two-term limit economy; if the challenger was able to also serve two terms, then voters would expect him to be subject to electoral concerns in his first period in office, meaning that he could avoid flip-flopping. As a result, they will always prefer the incumbent when their reputations are the same: in other words, such a model would feature an endogenous incumbency advantage.

[11] El Salvador, Mexico, Honduras and a few others.

Intuitively, the condition states that the single-term limit rule is beneficial if the benefit from selecting a competent politician is low enough. Moreover, notice that if $\lambda \to 1$, the right-hand side goes to 0, meaning that if the incumbent is competent with a sufficiently high probability, even a very small benefit b is sufficient to prefer the re-electability of the incumbent. The same happens if $q \to \frac{1}{2}$ or as $\phi \to +\infty$, because in these cases lying will be so prevalent to offset most of the learning about the incumbent.

Finally, notice that implementing a single term limit rule requires the ability to commit (for example through a constitution) not to re-elect an incumbent thought to be competent with a high probability [12].

2.5.2 Transparency of Actions: Reporting Media

Voters usually rely on the media to learn the policies chosen by politicians. In some circumstances, for example when bills containing multiple prescriptions are voted [13], it is not so straightforward to understand whether the incumbent politician flip-flopped or played consistently. The same might happen when voting in a committee is secret and knowing the result only enables to make probabilistic statements on whether a member voted in a certain direction.

In this section I therefore relax the assumption of full observability of the incumbent's actions and evaluate its impact on social welfare. The fact that action transparency is not always beneficial is well known in the literature [14]. What I am going to show here is that an imperfectly accurate reporting media is always optimal (or, in terms of transparency, partial action transparency is always optimal), unless the equilibrium is truthful with a fully accurate media in the first place.

Assume that the incumbent's track record is only observable through the report of a media company, and that the media company's reporting technology is not perfectly accurate: given a true incumbent track record τ that displays flip-flopping, with probability $1-g$ the media company sends out a report indicating that the politician acted consistently: the

[12] The idea of commitment to a single term limit is developped in a rather hyperbolic form by Calvino (1969) in a short story in which he describes a hypothetical society in which leaders are beheaded at the end of their term in office.

[13] An example of a situation in which it was not simple to label a policy choice as a flip-flop is the vote by Bernie Sanders against the auto-industry bailout in January 2009: Sanders had actually supported the bailout previously and supported it afterwards, but after having voted in favour of it, he voted against the release of that tranche of aid since it also contained financial aid for the banking sector, which Sanders was not in favour of bailing out. In the recent presidential primary election, Hillary Clinton used this alleged flip-flop to attack Sanders.

[14] Prat (2005) again as the main example.

same happens when the true track record is consistent [15]. For example, a voter observing a consistent track-record infers that the actual track-record of the incumbent is consistent with probability:

$$p_C = Pr(\tau = C|\tilde{\tau} = C) = \frac{gPr(\tau = C)}{gPr(\tau = C) + (1-g)Pr(\tau = F)}$$

while an analogous expression, denoted by p_F, represents the probability that the true track record is flip flopping given an observed flip-flopping record. Given her guess of the incumbent's strategy, the voter updates her beliefs and assigns a reputation to each true track-record. Since she does not observe the true track record, the actual reputation that she assigns to the politician is simply the weighted average of the reputations following each track-record, weighted by the probability that the observed track-record is of each type conditional on the observed media report. As a result, if $g = 1$ there is no noise (full transparency) in the reporting media and we recover the baseline model, whereas if $g = \frac{1}{2}$ there is no transparency and the reputation associated to each track-record is simply λ.

Since the lower g, the less voters can learn about the incumbent independently of his behaviour, it is intuitive that truthful play can be restored for g low enough. It is also clear that if this g, which I'll denote by g^*, is larger than $\frac{1}{2}$, then having $g < g^*$ is never optimal, because once truthful play has been restored, lowering g only makes learning worse. What is not straightforward is the fact that increasing g starting from g^* is never optimal. The following proposition shows that the optimal reporting accuracy is always the maximum g sustaining truthful play.

Proposition 3. *The optimal media reporting accuracy is $g^* \leq 1$. The value of g^* is the largest possible such that incumbents play truthfully. Therefore, $g^* < 1$ whenever a truthful equilibrium is not sustainable in the baseline model.*

Proof. See Appendix 2.A. □

2.5.3 Transparency of Consequences: Commentator Media

So far I assumed that voters have no feedback on the state of the world before elections. This serves the purpose of creating an environment in which all learning about the incumbent is done through his track-record. In many situations, however, voters have some

[15] Notice that I do not need to fully specify whether the noise comes from the report of the second or the first period action, or a mix of both: all I need is that noise is symmetric across the actions played by politicians.

information about the right policy choice. A fundamental role in this respect is again played by the media.

In this section, therefore, I augment the baseline model with an additional player, which I call commentator media to follow up on the contribution by Ashworth and Shotts (2010). To keep the analysis as simple as possible I assume that the media is only active at the end of $t = 2$, i.e. just before elections: this also reflects the fact that the coverage of politics in the vicinity of elections is particularly salient. I assume that the media is endowed with a signalling technology of accuracy $q_M > \frac{1}{2}$ (conditionally independent of the signal that incumbents receive) and I abstract from strategic consideration on the part of the media assuming that before election, the media truthfully reveals the realization of its signal s_M.

What is going to happen is simply that voters now have another piece of information to use when updating their beliefs on the incumbent type: reputation will not only depend on whether the incumbent flip-flopped or played consistently, but also on whether the media report endorses his second period action. In such a setup it seems natural that having an informative signal on the state of the world will act as disciplining device for the incumbent. The main result of this section, however, is to show that things need not work this way.

Proposition 4. *Increasing the accuracy of the commentator media can increase flip-flopping avoidance.*

Proof. See Appendix 2.A. □

It is interesting to describe in some detail the circumstances in which an increase in media accuracy increases the distortion to accountability. What is needed is a highly persistent environment (high γ) and high level of competence of politicians (both high q and high λ). In this context, increasing q_M from the lower bound of $\frac{1}{2}$ is initially marginally beneficial, but as q_M increases the effect reverses and further increasing q_M (up to a point) increases the distortion. When q_M starts becoming very precise, however, there is a large benefit in further increases of q_M. In other words, in these environments a marginally informative media is slightly beneficial, an informative but not very precise media is detrimental but a very precise media is again beneficial.

The intuition behind this result is what we might define as politicians gambling on the endorsement by the media. The force driving the result is the fact that when λ, γ and q are large, then an increase in media accuracy q_M sharply increases the payoff value of a successful gamble, which is the difference in reputation between being consistent and endorsed by the media and being a flip-flopper and opposed by the media, denoted by $\mu_{C,E} - \mu_{F,O}$. On the other hand, the the payoff from a failed gamble, which is the difference in reputation between a consistent politician opposed by the media and a flip-flopper endorsed by the media, or $\mu_{C,O} - \mu_{F,E}$, decreases more slowly (but it suddenly drops when q_M is sufficiently high, meaning that when the media is very well informed, a non-endorsement is very costly in terms of reputation). Moreover, as long as the media is not too informative, the probability S that the media signal matches the politician's flip-flopping signal remains close to $\frac{1}{2}$, making it likely for a politician avoiding a flip-flop to gamble successfully. This result relies heavily on λ being high, which is what gives the necessary curvature to the payoff from gambling on avoiding the flip-flop: in other words, when most politicians are competent, increasing the informativeness of the media increases the value of an endorsement significantly.

A related result is that when persistence γ, the share of competent politicians λ and electoral concerns ϕ are all very high, it might be the case that incompetent politicians distort their behaviour even under a perfect media signal $q_M = 1$. In this case, flip-flopping is such a strong signal of incompetence that politicians are willing to take a gamble in which they lose the election for sure unless they receive a media endorsement.

The existence of potentially non-monotone responses of policy distortions to media accuracy can have interesting implications for issues such as the public subsidization of media outlets. In persistent environment with a prevalence of competent politicians, subsidizing media is only beneficial if very precise commentary is achieved. This could for example suggest that concentrating resources into one or few high-quality outlets is better than subsidizing many average ones. In particular, once the media is sufficiently informative, returns to small increases in informativeness can be very large: in other words, in these situations *the devil is in the details*.

Notice that when an increased accuracy of the commentator media increases flip-flopping avoidance, accountability welfare decreases but selection welfare might still improve, so I are not able to conclude that a more accurate commentator media is detrimental to welfare *tout court*.

2.6 Predictability, Opportunism and Flip-flopping

The central aim of this paper is to provide a rationale of why voters dislike flip-flopping politicians. The explanation I offer in this paper has to do with signalling competence: since the information available to badly informed politicians is more likely to change, flip-flopping is associated with incompetence.

Whereas information is certainly a component to the reasons why flip-floppers are disliked, other explanations that are often brought forward in the political discourse have to do with predictability and opportunism. In this section I am going to show that the structure of my model can also be used to shed some light on these alternative explanations of flip-flopping.

Let's begin with predictability. One of the reasons voters dislike flip-floppers, as a matter of fact, is because flip-floppers are seen to be less predictable candidates, and predictability can be desirable when for example politicians lack commitment. In my model, when voters see a flip-flop, they not only update negatively on the politician's competence, but they also assign a higher probability to a further flip-flop happening in the future period. In other words, the model is consistent with the idea that flip-flopping politicians are unpredictable (although predictability has no value per se for the voter in my model).

Proposition 5. *Politicians who flip-flop before elections are also more likely to flip-flop again if re-elected.*

Proof. See Appendix 2.A. □

What about opportunism? My model is a model of opportunistic avoidance of flip-flops rather than one of opportunistic flip-flops, and there is no readily available relabelling of the model to fully catch that feature. However, if the signal politicians receive in my model is interpreted as a signal concerning the opinion of the majority, then we can reinterpret the model as saying that flip-floppers have a bad reputation because they are more likely to be unskilled at understanding the opinion of majority of voters. In a world in which all politicians are opportunists, in other words, flip-floppers are more likely to be bad opportunists.

In future research I would like to address the question of opportunistic flip-flops more in detail.

2.7 Conclusion

This paper describes circumstances in which career concerned politicians have the incentive to inefficiently stick to their previous policy positions in order to avoid the reputational stigma of flip-flopping. This happens because voters are aware that policy shifts are more likely to be performed by incompetent leaders. The incentive to avoid efficient policy shifts damages voters' welfare both in terms of policy effectiveness and selection of competent candidates through elections.

In other words, this paper confirms the conventional wisdom that flip-flopping is bad for the reputation of a politician, but at the same time it suggests that if electoral concerns are strong enough, then the equilibrium displays insufficient flip-flopping. An additional interesting implication of the mechanism described in the paper is that changes in the fundamentals driving policy choices are likely to bring to leadership change, but politicians substituting an incumbent after a change of state has occurred are on average less competent than politicians entering power in stable times.

To sum up, my analysis provides a new example of how electoral competition can decrease welfare: elections enable voters to retain the politicians they believe to be competent, but at the same time they give potentially distortive incentives to politicians. In this context, therefore, it might be optimal not to subject politicians to electoral incentives, by either setting up a single term limit rule or handing decision power to a judiciary.

The problem highlighted in my paper might also be framed in terms of disclosure of politicians' actions. Going back to the excerpt from the British CFOI (1997) quoted at the beginning of the paper, the statement "*[...] where disclosure of the information in question would be harmful to the frankness and candour of future discussions*" makes a specific reference for the distortion to what is denoted as "future frankness and candour", which is precisely what happens in my model: the policy taken in the first period by the politician biases his second period action, making it potentially an expression of posturing rather than candour.

Given that the distortion caused by electoral incentives works through the observability of the politician's actions, another way to eliminate the incentives for inefficient policy choices is to relax the ability of voters to observe policy choices. In the realm of public policy choices, the observability of a politician's actions is to a large extent determined by the media reporting. In such a context, I show that a less than fully accurate media can partially insulate politicians from the flip-flopping stigma: for this reason having some noise in the media is always optimal. This result suggests that the increasingly widespread access to politicians' track records made possible by the internet might in fact have a negative impact on the accountability of politicians.

Finally, I also show that when considering media with the function of endorsing the policy choice of a politician, an increase in the accuracy of the media signal can in some circumstances give politicians an even larger incentive to distort their choices avoiding efficient policy shifts. In other words, giving voters more information to evaluate politicians' track records can backfire, which has implications on for example the extent to which the media should be subsidized.

Appendix 2.A Proofs

Proof of Proposition 1

Proof. Let's consider the choice of action at $t = 2$, politicians have already taken action a_1 and they know that if re-elected, which happens with probability $\frac{1}{2} + \frac{\mu(a_1,a_2)-\lambda_O}{2}$, they will get 2ϕ. On top of that, politicians get utility of 1 whenever they match the state of the world, the probability of which depends on the posterior belief, denoted by $Pr(\omega_2 = s_2|s_1, q_\theta) = \rho_2(s_2, s_1, q_\theta)$. In order to slightly simplify notation, denote the probability of winning given reputation μ by $r(\mu)$. So $r(\mu) = \frac{1}{2} + \frac{\mu-\lambda_O}{2}$ as $r(\mu)$. Given that there are only two actions available, the politicians will calculate the reputation associated to each action and follow his signal if and only if:

$$\rho_2(s_2, s_1, q_\theta) + r(\mu(a_2 = s_2, s_1))2\phi \geq [1 - \rho_2(s_2, s_1, q_\theta)] + r(\mu(a_2 \neq s_2, s_1))2\phi.$$

It follows that in order to have $\sigma(s_2) = 1$, the above condition needs to hold for both types and both signal realizations given any of the two possible choices a_1, which results in a set of 8 inequalities.

However, it is immediate to notice that whenever $a_2 = s_2$ is the most reputable action, i.e. $\mu(a_2 = a_1, a_1) \geq \mu(a_2 \neq a_1, a_1)$, following the signal is unquestionably optimal. Since consistency has a better reputation than flip-flopping, it follows that whenever $s_2 = a_1$, $a_2 = s_2$, for each a_1 and each type. This means that we are left with 4 conditions. Moreover, since $\mu^T(0,0) = \mu^T(1,1)$ and $\mu^T(1,0) = \mu^T(0,1)$ and the same holds for $\rho_2(1,0) = \rho_2(0,1)$, we are left with only two conditions, one for each type:

$$\rho_2(s_2 \neq s_1, q_\theta) + r(\mu_F^T)2\phi \geq [1 - \rho_2(s_2 \neq s_1, q_\theta)] + r(\mu_C^T)2\phi,$$

which can be rearranged to:

$$\frac{2\rho_2(s_2 \neq s_1, q_\theta) - 1}{\phi} \geq \mu_C^T - \mu_F^T.$$

Thanks to the single-crossing property, the binding constraint for a truthful equilibrium is the condition concerning the incompetent politician: whenever the incompetent politician follows his signal, or is at least indifferent, the competent politician does, too. Conversely, if the competent politician is indifferent or he doesn't follow his signal, the incompetent will also not follow it.

As a result, we know that if $\frac{2\rho_2(s_2 \neq s_1, q)-1}{\phi} \geq \mu_C^T - \mu_F^T$, politicians will follow their signal at $t = 2$. Rearranging we get the condition on ϕ, $\phi \leq \frac{2\rho_2(s_2 \neq s_1, q) - 1}{\mu_C^T - \mu_F^T}$.

Let's now see what happens at $t = 1$. If politicians know that they are going to follow their signal at $t = 2$, then the dominant strategy at $t = 1$ is to follow their signal. Since all that matters is being consistent versus flip-flopping, then given the persistence of the state, it is more likely to end up in the favourable situation of playing consistently by following one's signal in the first period.

□

Trembling-hand perfection refinement eliminates pooling equilibria

Proof. In a pooling equilibrium, both politicians play the same track-record with probability one independent of the signals received. In other words, in these equilibria an action in every period remains off-equilibrium, hence Bayes' rule cannot restrict beliefs on these actions in any way. It follows that if the reputation attached to any off equilibrium track-record is sufficiently bad, then when electoral concerns are high enough politicians will not have any incentive to deviate from the pooling track-record. Sufficient conditions for any pooling equilibrium to be sustainable are that:

$$2\rho_1 - 1 + Pr(s_2 = s_1|s_1)(2\bar{\rho}_2 - 1) < (r(\mu = \lambda) - r(\mu = 0))2\phi$$

and

$$2\bar{\rho}_2 - 1 < r(\mu = \lambda) - r(\mu = 0))2\phi$$

Let's now introduce the trembling-hand perfection requirement. Assume that with probability $\epsilon > 0$ close to zero, a politician willing to play action a will instead play action a'. Assume that ϵ is the same for both types. Take a pooling equilibrium. In any period, with probability ϵ the voter observes an action different from that on the pooling track-record. Since both politicians have the same strategy and ϵ is the same for both types, then the reputation the voter must attach to actions outside the pooling track-record has to be λ, the same as the reputation of the pooling track-record. However, this cannot happen in equilibrium, because if the reputation of any track-record is the same, then incumbents have an incentive to always follow their signal. In other words, no pooling equilibrium can survive the trembling hand perfection requirement. □

Proof of Theorem 1

Proof. I split the proof in several parts. First of all I characterize the symmetric partially truthful equilibrium.

Claim 1: Existence of Partially Truthful Equilibrium

In this equilibrium, $\mu(0,0) = \mu(1,1) \equiv \mu_C$ and $\mu(0,1) = \mu(1,0) \equiv \mu_F$ by symmetry and $\mu_C > \mu_F$. Moreover, assume for now that incumbents always follows their signal at $t = 1$. Consider now period $t = 2$. After following their signal in period 1, in period 2 the incumbent has to decide whether to follow his signal or not. When $s_2 = a_1$, the incumbent always follows his signal, because:

$$\rho_2(s_1, s_2 = s_1, \theta) + r(\mu_C)2\phi > (1 - \rho_2(s_1, s_2 = s_1, \theta) + r(\mu_F)2\phi$$

since $\mu_C > \mu_F$ insures that $r(\mu_C) > r(\mu_F)$ and $\rho_2 > 1 - \rho_2$ by the decision relevance of signals. However, when $s_2 \neq a_1$, the incumbent has a tradeoff. From Proposition 1, a truthful equilibrium requires that:

$$\rho_2(s_1, s_2 \neq s_1, L) + r(\mu_F)2\phi \geq (1 - \rho_2(s_1, s_2 \neq s_1, L)) + r(\mu_C)2\phi,$$

from which the upper bound $\bar{\phi}$ was derived. Moreover, we know from Fact 2 (single-crossing property), that the following holds:

$$\rho_2(s_1, s_2 \neq s_1, L) + r(\mu'_F) \geq 1 - \rho_2(s_1, s_2 \neq s_1, L) + r(\mu'_C) \Rightarrow$$
$$\rho_2(s_1, s_2 \neq s_1, H) + r(\mu'_F) > 1 - \rho_2(s_1, s_2 \neq s_1, H) + r(\mu'_C)$$

and

$$\rho_2(s_1, s_2 \neq s_1, H) + r(\mu'_F) \leq 1 - \rho_2(s_1, s_2 \neq s_1, H) + r(\mu'_C) \Rightarrow$$
$$\rho_2(s_1, s_2 \neq s_1, L) + r(\mu'_F) > 1 - \rho_2(s_1, s_2 \neq s_1, L) + r(\mu'_C)$$

This means that we are left with three possibilities: both politicians play $a_2 \neq s_2$ when that involves flip-flopping, or the high type mixes between $a_2 = s_2$ and $a_2 = s_2$, or the high type always plays $a_2 = s_2$ and the low type mixes. I will now prove that the first two cannot be part of an equilibrium. Assume that both politicians play $a_2 \neq s_2$ when $s_2 \neq a_1$. Then, nobody would flip-flop and therefore, in a candidate equilibrium $\mu_C = \mu_F$; in this case, however, the optimal strategy for the incumbent is to follow his signal. Consider now the other case, i.e. that when $s_2 \neq a_1$, the high type mixes between $a_2 = s_2$ and $a_2 \neq s_2$ while the low type always plays $a_2 \neq s_2$. If this were an equilibrium, then flip-flopping would reveal the high type, and therefore $\mu_F = 1 > \mu_C$. In such a situation, however, the incumbent would always flip-flop. It follows that the only possibility when $s_2 \neq a_1$ is that the high type follows his signal whereas the low type mixes between the two actions. The low type mixes when the following holds:

$$\frac{2\rho_2(s_1, s_2 \neq s_1, L) - 1}{\phi} = \underbrace{\frac{\lambda\gamma}{\lambda\gamma + (1-\lambda)(1 - A\sigma^*)}}_{\mu_C} - \underbrace{\frac{\lambda(1-\gamma)}{\lambda(1-\gamma) + (1-\lambda)A\sigma^*}}_{\mu_F} \tag{2.2}$$

In order to show existence and uniqueness of such an equilibrium, consider $\sigma^* \in [0,1]$. It holds from Proposition 1 that at $\sigma^* = 1$, $\frac{2\rho_2(s_1,s_2\neq s_1,L)-1}{\phi} < \mu_c - \mu_F$. At the same time, at $\sigma^* = 0$ it has to be that $\frac{2\rho_2(s_1,s_2\neq s_1,L)-1}{\phi} > \mu_c - \mu_F$, since in that case $\mu_F = 1$. By continuity of $\mu_C - \mu_F$, an equilibrium with $\sigma^* \in (0,1)$ exists. Moreover, notice that μ_C is strictly increasing in σ^* while μ_F is strictly decreasing in σ^*, and therefore the σ^* such that (2.2) is unique. Finally, consider the action choice at $t = 1$. It holds that $\mu_C > \mu_F$ and the incumbent knows he is going to follow his signal when $s_2 = s_1$. Moreover, the incumbent knows that, because of the persistence of the state of the world, given s_1 it is more likely for him to receive $s_2 = s_1$. However, the utility of receiving $s_2 = s_1$ is higher when $a_1 = s_1$ than when $a_1 \neq s_1$. As a result, following the signal at $t = 1$ is optimal both in terms of instantaneous payoff and in terms of future payoff. Mathematically, denote by $\pi(\theta) = Pr(s_2 = s_1|s_1)$. Notice that since $\gamma > \frac{1}{2}$, $\pi > \frac{1}{2}$ and notice that $\pi(\theta)\rho_2(s_1, s_2 = s_1, \theta) + (1-\pi(\theta))\rho_2(s_1, s_2 \neq s_1, \theta) = \rho_1(s_1,\theta)\gamma + (1-\rho_1(s_1,\theta))(1-\gamma)$. Denote also by $\bar{\rho}_2 = \rho_2(s_1, s_2 = s_1, \theta)$ and by $\underline{\rho}_2 = \rho_2(s_1, s_2 \neq s_1, \theta)$. With this notation, following one's signal at t_1 requires the following condition to hold:

$$\rho_1 + \pi(\bar{\rho}_2 + r_C 2\phi) + (1-\pi)(\underline{\rho}_2 + r_F 2\phi) \geq (1-\rho_1) + \pi(\bar{\rho}_2 + r_F 2\phi) + (1-\pi)(\underline{\rho}_2 + r_C 2\phi)$$

which can be rearranged to

$$(2\rho_1 - 1) \geq -(2\pi - 1)(r_C - r_F)2\phi$$

which always holds. Hence both types follow their signal at $t = 1$.

To sum up, I have showed the existence of a unique symmetric partially truthful equilibrium (it would be truthful if $\phi \leq \bar{\phi}$) with and $\mu_C > \mu_F$. Both politicians follow their signal at $t = 1$. At $t = 2$, the high type always follows his signal, whereas the low type follows his signal with probability 1 when if $s_2 = s_1$ and mixes playing $a_2 = s_2$ with probability $\sigma^* \in (0, 1)$ when $s_2 \neq s_1$.

Claim 2: Uniqueness of Symmetric Non-Pooling Equilibrium

In Claim 1 I have shown that if $\mu_C > \mu_F$, then the equilibrium is the partially truthful one I characterized (or truthful, depending on ϕ). I will now prove that remaining in the realm of symmetric equilibria, there exists no equilibrium where $\mu_F > \mu_C$. By claim 1, if $a_1 = s_1$ for both politicians the unique equilibrium is the (partially) truthful one. Let's then assume that $(2\rho_1 - 1) < (2\pi - 1)(r_F - r_C)2\phi$ for some incumbent type. Assume that the incumbent played an action $a_1 \neq s_1$. Since $\mu_F > \mu_C$, at $t = 2$ the optimal choice is to follow the signal when $s_2 = s_1$ whereas a tradeoff arises when $s_2 \neq s_1$. However, notice that $\frac{2\rho_1 - 1}{2\pi - 1} > 2\rho_2 - 1$, because $\pi = \gamma q + (1-\gamma)(1-q) < q$ given that $q > \frac{1}{2}$. Therefore, whenever the incumbent plays $a_1 \neq s_1$, at $t = 2$ he would always play $a_2 \neq a_1$. However, this cannot be part of an equilibrium. If the incumbent knows he is always going to flip flop, as a matter of fact, then the optimal strategy is to follow the first signal and then deviate if $s_2 = s_1$. The reason is that the such a strategy gives the incumbent the same reputation but it is less expensive in terms of policy costs, since $\pi(2\bar{\rho}_2 - 1) < (2\rho_1 - 1) + (1-\pi)(2\underline{\rho}_2 - 1)$. This can in fact be rearranged to yield:

$$[2(\rho_1\gamma + (1-\rho_1)(1-\gamma)) - 1] - (1-\pi)(2\underline{\rho}_2 - 1) < 2\rho_1 - 1 + (1-\pi)(2\underline{\rho}_2 - 1).$$

Claim 3: Non-Existence of Asymmetric Equilibria

I prove this result in several substeps.

Step 1: If $a_1 = s_1$, the equilibrium in the subgame starting at $t = 2$ is either truthful or partially truthful.

Proof. This was proved in Claim 1. □

Step 2: If for some type $a_2 = s_2$ with strictly positive probability, then for that type $a_1 = s_1$.

Proof. Denote by α and β the gap between the consistent and flip-flopping reputation after action $a_1 = 1$ and $a_1 = 0$ respectively. Moreover, denote by $c_1 = \frac{2\rho_1 - 1}{\phi}$. Denote by $\pi = Prob(s_2 = s_1|s_1)$. If some type contradicts the signal at $t = 1$ with some probability after both signal realizations, and then plays $a_2 = s_2$ with some positive probability at $t = 2$, the following has to hold:

$$\mu(1,0) - \mu(0,1) = c_1 + \pi[\mu(0,0) - \mu(0,1)] - (1-\pi)[\mu(1,1) - \mu(1,0)]$$
$$\mu(0,1) - \mu(1,0) = c_1 + \pi[\mu(1,1) - \mu(1,0)] - (1-\pi)[\mu(0,0) - \mu(0,1)]$$
$$\mu(0,0) - \mu(0,1) = \beta$$
$$\mu(1,1) - \mu(1,0) = \alpha$$

Substituting the last two equations into the first two, and then the second into the first, we obtain the following condition:

$$c_1 = -(2\pi - 1)\frac{\alpha + \beta}{2}.$$

Now, whenever negative, $\frac{\alpha+\beta}{2} \in [-1, 0]$. However, $\frac{c_1}{2\pi - 1} > 1$ since $\pi < \rho_1$. So this equality can never hold. This means that it is never the case that an incumbent contradicts the signal with some probability after both realizations of s_1.

Consider the case of a potential equilibrium in which in the first period, the incompetent politician mixes after receiving one of the signals. Let's consider, without loss of generality, a candidate equilibrium in which $Pr(a_1 = 1|s_1 = 1) = \sigma_1 < 1$. When the voter observes $a_1 = 1$, however, she knows that it is a genuine realization of $s_1 = 1$. As a result, the game play after $a_1 = 1$ is the same as that in the partially truthful equilibrium. Hence, $\mu(1,1) > \mu(1,0)$ and $\mu(1,1) - \mu(1,0) \leq c_2$. There are now therefore two cases: $\mu(0,0) > \mu(1,0)$ and $\mu(0,0) < \mu(1,0)$. If $\mu(0,0) > \mu(0,1)$, then since we are considering equilibria in which there is at least partialy truthful play at $t = 2$, it has to be that $\mu(0,0) - \mu(0,1) \leq c_2$. Since the average reputation after $a_1 = 1$ is larger than after $a_1 = 0$, because just incompetents are deviating, then it has to be that $\mu(1,1) \geq \mu(0,1)$. However, in order for the incumbent to have the incentive to deviate from $s_1 = 1$ it has to be the case that $c_1 + (\pi\mu(1,1) + (1-\pi)\mu(1,0)) \leq \pi\mu(0,1) + (1-\pi)\mu(0,0)$. However, $\pi\mu(0,1) + (1-\pi)\mu(0,0) - (\pi\mu(1,1) + (1-\pi)\mu(1,0))$ can be at most $2(1-\pi)c_2$ without the condition that $\mu(1,1) \geq \mu(0,1)$ being violated. Since $c_1 > 2(1-\pi)c_2$, then this distance is not enough to give the politician the incentive to deviate after $s_1 = 1$.

If $\mu(0,0) > \mu(0,1)$, and $\mu(0,0) - \mu(0,1) \leq c_2$, consider that since the average reputation after $a_1 = 1$ is larger than after $a_1 = 0$, because just incompetents are deviating, then it has to be that $\mu(1,1) \geq \mu(0,1)$. However, in order for the incumbent to have the incentive to deviate from $s_1 = 1$ it has to be the case that $c_1 + (\pi\mu(1,1) + (1-\pi)\mu(1,0)) \leq \pi\mu(0,1) + (1-\pi)\mu(0,0)$. However, $\pi\mu(0,1) + (1-\pi)\mu(0,0) - (\pi\mu(1,1) + (1-\pi)\mu(1,0))$ can be at most $2(1-\pi)c_2$ without the condition that $\mu(1,1) \geq \mu(0,1)$ being violated. Since $c_1 > 2(1-\pi)c_2$, then this distance is not enough to give the politician the incentive to deviate after $s_1 = 1$.

Let's now consider the case in which $\mu(0,1) > \mu(0,0)$. If $\mu(0,1) - \mu(0,0) \leq c_2$, then we can immediately see that $\pi\mu(0,1)+(1-\pi)\mu(0,0)-(\pi\mu(1,1)+(1-\pi)\mu(1,0)) \leq c_2$, but since $c_2 < c_1$, again the incumbent cannot have the incentive to deviate without $\mu(0,0) > \mu(1,1)$; however, if $\mu(0,0) > \mu(1,1)$ the reputation after action $a_1 = 0$ is strictly larger than after $a_1 = 1$, which is a contradiction.

□

Step 3: there is no equilibrium in which some type does not follow the signal at $t = 1$ with probability 1 and then contradicts one realization of the signal at $t = 2$.

Proof. Impossible for a high type to contradict the signal at $t = 2$, unless that is a pooling equilibrium (which we remove). Therefore, assume a low type contradicted s_1. Without loss of generality, assume that a low type who received signal $s_1 = 0$ played instead $a_1 = 1$. If $\mu(1,1) > \mu(1,0)$, then there are two possibilities: if action $a_1 = 1$ is only played by high types and low types who received a signal $s_1 = 0$, then contradicting signal $s_2 = 0$ would lead to $\tau = (1,0)$ to be revealing of the high type, which cannot happen in equilibrium. If $a_1 = 1$ were also to be played by low types receiving signal $s_1 = 1$, then if the low types who received $s_1 = 0$ contradict their signal $s_2 = 0$, also the low types receiving $s_2 = 0$ after having received $s_1 = 1$ and played $a_1 = 1$ would do that. And therefore $\tau = (1,0)$ would be revealing of the high type, which cannot hapen in equilibrium. So contradicting the signal $s_1 0$ at $t = 1$ requires $\mu(1,0) > \mu(1,1)$. Since the high type plays truthfully at $t = 2$, however, conditional on $a_1 = 1$ he plays $a_2 = 1$ more often than $a_2 = 0$. Therefore, $\mu(1,0)$ cannot be larger than $\mu(1,1)$.

□

□

Proof of Lemma 1

Proof. The average reputation is constant in equilibrium and equal to λ, hence $Pr(\tau = C)\mu_C + Pr(\tau = F)\mu_F = \lambda$. This allows us to rewrite $Pr(\tau = C) = \frac{\lambda - \mu_F}{\mu_C - \mu_F}$. From the model we know that selection improves if and only if the second moment m_2 of the distribution of reputation, $m_2 = Pr(\tau = C)\mu_C^2 + (1 - Pr(\tau = C))\mu_F^2$, increases. Moreover, in equilibrium it holds that $\mu_C \geq \mu_F$ and $Pr(\tau = C) > Pr(\tau = F)$. Taking the expression for the second moment and substituting in the restriction due to the constant mean yields:

$$m_2 = \frac{\lambda - \mu_F}{\mu_C - \mu_F}\mu_C^2 + [1 - \frac{\lambda - \mu_F}{\mu_C - \mu_F}]\mu_F^2$$

Let's now write the expression of a contour line, in the (μ_C, μ_F) plane, along which the value of the second moment is constant; using implicit differentiation one gets:

$$\frac{\partial \mu_F}{\partial \mu_C} = \frac{\lambda - \mu_F}{\mu_C - \lambda}$$

It can be seen that this contour line is increasing in μ_C and concave. This means that the second moment increases by increasing μ_C and decreasing μ_F. Moreover, notice that since in equilibrium $\mu_C > \mu_F$, $Pr(\mu_C) > Pr(\mu_F)$ and $Pr(\mu_C)\mu_C + Pr(\mu_F)\mu_F = \lambda$, then $\mu_C - \lambda < \lambda - \mu_F$. So $\frac{\partial \mu_F}{\partial \mu_C} \geq 1$ in equilibrium. As a result, therefore, whenever $\mu_C - \mu_F$ increases and μ_C increases, the second moment m_2 has to increase and with it also selection welfare increases. Thus can be immediately be seen graphically, since the contour line of the second moment is increasing and concave and the line along which $\mu_C - \mu_F$ is constant is increasing with a slope of 1. In other words, knowing the size of $\mu_C - \mu_F$ is sufficient to compare the selection welfare across different equilibria. □

Proof of Proposition 2

Proof. The welfare expression (2.1) derived in the text reads:

$$W = [\lambda + (1 - \lambda)q] + [\lambda + (1 - \lambda)\tilde{q}] + b\left[\frac{1}{4} + \frac{\lambda_O^2}{4} - \frac{\lambda\lambda_O}{2} + \frac{\lambda + \lambda_O}{2} + \frac{\mathbb{E}\mu^2}{4}\right]$$

With a single term limit, politicians can do no better than following their signal, since there is no re-election possibility. This generates a gain of

$$(1 - \lambda)A(1 - \sigma^*)(2\rho_2 - 1)$$

At the same time, however, a single term limit corresponds to a commitment to choosing the challenger no matter what the belief about the incumbent is. Therefore, society gets the benefit b with probability λ_O, plus $\mathbb{E}v = 0$. The utility from the selection of the right type of politician is therefore $\lambda_O b$ rather than $b\left[\frac{1}{4} + \frac{\lambda_O^2}{4} - \frac{\lambda\lambda_O}{2} + \frac{\lambda+\lambda_O}{2} + \frac{\mathbb{E}\mu^2}{4}\right]$ This generates a loss in terms of selection welfare expressed by:

$$b\left(\frac{1}{4}+\frac{\lambda_O^2}{4}-\frac{\lambda\lambda_O}{2}+\frac{\lambda-\lambda_O}{2}+\frac{\mathbb{E}\mu^2}{4}\right)$$

It follows that having a single term limit is beneficial if the following inequality holds:

$$(1-\lambda)A(1-\sigma^*)(2\rho_2-1)\geq b\left(\frac{1}{4}+\frac{\lambda_O^2}{4}-\frac{\lambda\lambda_O}{2}+\frac{\lambda-\lambda_O}{2}+\frac{\mathbb{E}\mu^2}{4}\right)$$

□

Lemma 2: p_C+p_F increasing in σ

Proof. $p_C=\frac{gPr(\tau=C)}{gPr(\tau=C)+(1-g)Pr(\tau=F)}$ and $p_F=\frac{gPr(\tau=F)}{gPr(\tau=F)+(1-g)Pr(\tau=C)}$. Now, remember that $\mu_C=\frac{\lambda\gamma}{\lambda\gamma+(1-\lambda)(1-A\sigma)}=\frac{\lambda\gamma}{Pr(\tau=C)}$ and similarly, $\mu_F=\frac{\lambda(1-\gamma)}{Pr(\tau=F)}$. Substituting these into the expression for p_C and p_F one gets the following expressions:

$$p_C=\frac{g\gamma\mu_F}{g\gamma\mu_F+(1-g)(1-\gamma)\mu_C}$$

and

$$p_F=\frac{g(1-\gamma)\mu_C}{g(1-\gamma)\mu_C+(1-g)(1-\gamma)\mu_F}.$$

Let's now denote by D_C and D_F the denominators of μ_C and μ_F respectively and by Den_C and Den_F the denominator of p_C and p_F respectively. First of all, let's establish that $Den_F<Den_C$. This clearly holds since $g(1-\gamma)\mu_C+(1-g)\gamma\mu_F<g\gamma\mu_F+(1-g)(1-\gamma)\mu_C$ can be rearranged to yield $D_C>D_F$, which is always satisfied. Let's now look at the numerator of the derivatives of p_C and p_F with respect to σ. The former is:

$$-\frac{g\gamma(1-\gamma)\lambda(1-\lambda)A}{D_F^2}Den_C-\lambda(1-\lambda)A\gamma(1-\gamma)A[\frac{(1-g)}{D_C^2}-\frac{g}{D_F^2}]g\gamma\mu_F$$

and the latter is:

$$\frac{g\gamma(1-\gamma)\lambda(1-\lambda)A}{D_C^2}Den_F-\lambda(1-\lambda)A\gamma(1-\gamma)A[\frac{g}{D_C^2}-\frac{(1-g)}{D_F^2}]g(1-\gamma)\mu_C$$

Substituting for Den_F and Den_C and rearranging one can see that the numerators only differ in the sign: p_C decreases as σ increases, whereas p_F increases. However, since $Den_F<Den_C$, then the positive effect from p_F dominates. □

Proof of Proposition 3

Proof. In terms of first period behaviour, nothing changes with respect to the baseline model: it can therefore be proved that all incumbents follow their signal at $t = 1$ by using the same arguments provided for the baseline model (refer to the proof of Theorem 1). Let's now move to the decision to be taken at $t = 2$. An incumbent receiving a signal that suggests him to flip-flop knows that if he does flip-flop, with probability g voters will observe a media report indicating he has been consistent, and viceversa if he goes against his signal and avoids the flip-flop there is a probability $1 - g$ that his track record will come across as flip-flopping. It immediate to see, therefore, that given any two reputation levels, the incentive to strategically avoid the flip-flop is smaller when the reporting media is noisy. However, the voter is aware of the noisy signal received from the media and therefore she also adjusts her beliefs: whenever observing a flip-flopping track-record, for example, the voter knows that with some probability, the politician actually played consistently and viceversa. In particular, let's fix the strategy played by incumbents, and in particular let's assume that incompetent incumbents follow a flip-flopping signal with probability σ. If the voter were able to observe track-records perfectly, beliefs would be $\mu_C(\sigma)$ and $\mu_F(\sigma)$. Given the noise, the belief the incumbent is competent when observing a consistent track-record is $\tilde{\mu}_C = p_C\mu_C(\sigma) + (1 - p_C)\mu_F(\sigma)$, where $p_C = \frac{gPr(\tau=C)}{gPr(\tau=C)+(1-g)Pr(\tau=F)}$. Analogous expressions, denoted by $\tilde{\mu}_F$ and p_F, hold for the case in which the voter observes a flip-flopping track-record and are derived by simply swapping C with F. Denote now for simplicity by $\rho \equiv \rho_2(s_2 \neq s_1, q)$. It follows that the incumbent prefers to follow a flip-flopping signal whenever the following inequality holds:

$$\rho + gr(\tilde{\mu}_C)2\phi + (1 - g)r(\tilde{\mu}_F)2\phi \geq 1 - \rho + gr(\tilde{\mu}_C)2\phi + (1 - g)r(\tilde{\mu}_F)2\phi$$

Using the definitions of $\tilde{\mu}_C$ and $\tilde{\mu}_F$, this expression can be rearranged to yield the following:

$$\frac{2\rho - 1}{\phi} \geq (2g - 1)(\tilde{\mu}_C - \tilde{\mu}_F)$$

and further to get:

$$\frac{2\rho - 1}{\phi} \geq (2g - 1)(p_C + p_F - 1)(\mu_C - \mu_F)$$

Notice that if $g = 1$, i.e. no noise, then one gets back to the expression from the baseline model, given also that $p_C(g = 1) = 1 = p_F(g = 1)$. Let's now consider the case in which a truthful equilibrium is not sustainable when $g = 1$, since $\frac{2\rho-1}{\phi} < (\mu_C^T - \mu_F^T)$. When noise is introduces, the right hand side becomes $(2g - 1)(p_C + p_F - 1)(\mu_C^T - \mu_F^T)$. It is immediate to check that given μ_C^T and μ_F^T, the right-hand side strictly decreases as g becomes smaller than one. This means that there is a level of noise $g^* \in \left(\frac{1}{2}, 1\right)$ such that

$$\frac{2\rho - 1}{\phi} = (2g^* - 1)(p_C(g^*, \mu^T) + p_F(g^*, \mu^T) - 1)(\mu_C^T - \mu_F^T).$$

Moreover, remember that $\mu_C(\sigma) - \mu_F(\sigma)$ is strictly increasing in σ and by Lemma 2 it follows that $p_C(\hat{g}, \mu^T) + p_F(\hat{g}, \mu^T) - 1$ is also strictly increasing in σ. As a result, then, decreasing g always increases the equilibrium level of σ. This means that as long as $\sigma < 1$, decreasing g will alleviate the accountability distortion caused by incumbents avoiding flip-flops. Since the equilibrium level of $\tilde{\mu}_C - \tilde{\mu}_F$ increases as g decreases (as long as $\sigma < 1$), because in equilibrium $\frac{2\rho-1}{\phi} \geq (2g-1)(\tilde{\mu}_C - \tilde{\mu}_F)$, then the selection of politicians also improves thanks to Lemma 1. As a result, decreasing g improves welfare as long as it crowds out the lies of politicians. Decreasing g below g^*, however, $\mu_C - \mu_F$ cannot increase further and hence incompetent incumbents start having a strict preference towards following their signal when it prescribes a flip-flop. Therefore, decreasing g further will hurt learning and have no benefit on accountability. It follows that g^* is the optimal level of news informativeness. □

Proof of Proposition 4

Proof. Let's start the analysis by writing the modified reputations. Given that there are now two signals (the track record and the media signal) we now have four different reputations. I denote by $\mu_{C,E}$ and $\mu_{F,O}$ the reputation from consistent play given that the media endorses (E) or opposes (O) the politician's decision. Given this notation, the reputation expressions can be written in the following way:

$$\mu_{C,E} = \frac{\lambda\gamma q_M}{\lambda\gamma q_M + (1-\lambda)[(p_2 q q_M + (1-p_2)(1-q)(1-q_M)) + ((1-p_2)q(1-q_M) + p_2(1-q)q_M)(1-\sigma)]}$$

$$\mu_{F,E} = \frac{\lambda(1-\gamma)q_M}{\lambda(1-\gamma)q_M + (1-\lambda)((1-p_2)q(1-q_M) + p_2(1-q)q_M)\sigma}$$

$$\mu_{C,O} = \frac{\lambda\gamma(1-q_M)}{\lambda\gamma(1-q_M) + (1-\lambda)[((1-p_2)q(1-q_M) + p_2(1-q)q_M) + ((1-p_2)q q_M + p_2(1-q)(1-q_M))(1-\sigma)]}$$

$$\mu_{F,O} = \frac{\lambda(1-\gamma)(1-q_M)}{\lambda(1-\gamma)(1-q_M) + (1-\lambda)((1-p_2)q(1-q_M) + p_2(1-q)q_M)\sigma}$$

Notice that if $q_M = \frac{1}{2}$, then we get back to the expressions used in the baseline model. Moreover, notice that compared to the reputations from the baseline model, $\mu_{C,E} > \mu_C > \mu_{C,O}$ and the analogous inequality holds for the flip-flopping reputations. Now, denote by $S = Pr(s_M = F | s_P = F) = \rho_2 q_M + (1 - \rho_2)(1 - q_M)$, i.e. the probability that, given the incumbent's signal is flip-flopping, the media signal also endorses a flip-flop. It turns out that when they receive a flip-flopping signal, incompetent incumbents follow their signal if the following inequality holds:

$$\frac{2\rho_2 - 1}{\phi + q} = [S\mu_{C,O} + (1-S)\mu_{C,E}] - [S\mu_{F,E} + (1-S)\mu_{F,O}]$$

The expression can be rearranged to yield:

$$\frac{2\rho_2 - 1}{\phi + q} = S[\mu_{C,O} - \mu_{F,E}] + (1-S)[\mu_{C,E} - \mu_{F,O}]$$

Using implicit differentiation one can see that σ^* can decrease when q_M increases. In such a situation, accountability welfare decreases. In fact, it is sufficient to look at is the derivative of the right-hand side of the expression above with respect to q_M. In particular, if the right hand side increases in q_M, then σ^* needs to decrease for equilibrium to be restored. Differentiating the right-hand side we get the following condition for an increase in q_M to decrease σ^*:

$$(2\rho_2-1)[(\mu_{C,O}-\mu_{F,E})-(\mu_{C,E}-\mu_{F,O})]+(1-S)\left[\frac{\partial\mu_{C,E}}{\partial q_M} - \frac{\partial\mu_{F,O}}{\partial q_M}\right]-S\left[\frac{\partial\mu_{F,E}}{\partial q_M} - \frac{\partial\mu_{C,O}}{\partial q_M}\right] > 0$$

It turns out that the above inequality holds when γ and q are both high (with γ potentially higher than q), λ is sufficiently high and q_M is not too high. In this situation, increasing the informativeness of the commentator signal has the effect of increasing the distortion to accountability.

□

Proof of Proposition 5

Proof. The proof just consists of simple algebra. Denote by $p_3 = \gamma\rho_2 + (1-\gamma)(1-\rho_2)$ and $\bar{p}_3 = \gamma\bar{\rho}_2 + (1-\gamma)(1-\bar{\rho}_2)$. We have:

$$Pr(a_3 = a_2 | a_2 = a_1) =$$
$$= \mu_C\gamma + (1-\mu_C)[Pr(s_2 = s_1|a_2 = a_1, L)(q\bar{p}_3 + (1-q)(1-\bar{p}_3))+$$
$$(+1 - Pr(s_2 = s_1|a_2 = a_1, L))((1-q)p_3 + q(1-p_3))]$$

and

$$Pr(a_3 = a_2|a_2 = a_1) = \mu_F\gamma + (1-\mu_F)(qp_3 + (1-q)(1-p_3))$$

Moreover, $Pr(s_2 = s_1|a_2 = a_1, L) = \frac{1-A}{1-A\sigma^*}$. All I want to show is that $Pr(a_3 = a_2|a_2 =$

$a_1) \geq Pr(a_3 = a_2 | a_2 = a_1)$. Using the expressions for the probabilities and rearranging yields:

$$\frac{1-A}{1-A\sigma^*} \geq \frac{2\bar{p}_3 - 1}{\bar{p}_3 + p_3 - 1}$$

Since only the left-hand side depends on σ^*, and it is strictly increasing in σ^*, then evaluating the inequality at the lower bound on σ^*, i.e. $\underline{\sigma} = \frac{1-\gamma}{A}$, works as a sufficient condition. Substituting in for $\bar{p}_3$ and p_3 and $\underline{\sigma}$ yields:

$$\frac{2(1-\gamma) - 1 + 2(2\gamma - 1)\rho_2}{2(1-\gamma) - 1 + (2\gamma - 1)(\rho_2 + \bar{\rho}_2)} \geq \frac{1-\gamma}{\gamma} 2q(1-q) + (q^2 + (1-q)^2)$$

and it can be checked numerically that this condition is always satisfied for the parameter values of interest for the model. □

3 Candidates, Leaks and Media

This is joint work with Antoni Italo De Moragas

We present a model of a media market in which a set of news outlets compete to break a news. In our model, each media receives some information on whether a politician in office is corrupt. Media outlets can decide whether to break the story immediately or wait and fact-check, taking into account that if another media breaks the news, the profit opportunity disappears. We show that as the number of competitors increases, each outlet becomes more likely to break the news without fact-checking. Therefore, as the number of media increases, the incumbent politician is more likely to be accused of corruption by the media: this makes the re-election of incumbents more difficult and increases political turnover. In particular, we show that if voters consult with higher priority the media outlets that report about a scandal, increasing the number of competitors decreases the probability of having an honest politician in office.

3.1 Introduction

The news media play a fundamental role in providing political information to citizens and keeping candidates and elected officials accountable to public opinion (Strömberg, 2015). In order to fulfill their role, news media act as a filter between the information they receive from their sources and the information they transmit to the public. At the heart of this process lies what is often called fact-checking, which means verifying the claims of a source before publishing a report. A crucial question, therefore, is whether competition gives media outlets the incentive to undertake this very important task.

In this paper we show that this might not be the case: in our model, competition between media outlets crowds out fact-checking and leads to faster but more inaccurate reporting. In order to investigate what broader effects this might have on society, we introduce an electoral choice and describe how the less fact-checked information provided by the media to voters can distort it.

In our model, we consider a set of news outlets competing to break a news. Each media receives some information on whether a politician in office is corrupt. Media outlets can decide whether to break the story immediately or wait and fact-check. The benefit of fact-checking is to allow a media outlet to be sure about the validity of the rumour. Since publishing a fake story is costly for the media outlet [1], fact-checking therefore prevents fake scandals from making the news and affecting a reader's electoral choice. The cost of fact-checking consists in having to put the publication of a story on hold, thus giving other firms the opportunity to break the news, leaving the firm that decided to fact check with nothing.

We show that as the number of competitors increases, each outlet becomes more likely to break the story without fact checking. This, in turn, makes incumbent politicians increasingly likely to be accused of corruption, hurting their chances of re-election. This does not only affect corrupt incumbents, but also honest incumbents which might be ousted from office and replaced with corrupt challengers.

Interestingly, we show that if there is a *pecking order* such that whenever some media outlet reports the scandal, readers are exposed to that news (for example because they preferably buy the newspaper whose headline mentions the scandal, or because they google search for scandals involving the politician) then increasing the number of media makes it less likely for an honest politician to win the election.

The changes in the media landscape that happened in recent years, especially due to the ever growing importance of the internet, seem to well fit the conditions implied by our model for a decrease in news quality (and the resulting effect on the quality of elected politicians).

First of all, the internet has dramatically decreased the entry barriers into the news media sector: setting up a blog does not require significant capital or expertise, but gives anybody access to a potentially vast market of readers [2]. Secondly, the role of the internet and social media in the spread of news has transformed the way in which media outlets compete. In a world of around the clock news, being the first to cover an event is fundamental to increase traffic and earn through advertisements. Moreover, news of scandals travel very fast on social media. Since not reporting on a scandal does not go viral, readers are more likely to be exposed to the outlets that talk about a scandal rather than to those that don't.

In light of this, the mechanism described in our model might help interpreting the consistent decrease in the level of trust in the accuracy and fairness of media that took place in

---

[1] The channel through which publishing a fake scandal is costly for a media outlet is not made endogenous in our model. Two explanations can be libel lawsuit and reputation. However, how these two channel interact with the dynamics of publication and with the number of outlets is left unanswered by our model and could be object of future research.

[2] The trend towards an increased number of media outlets is not limited to the internet: see for example Cagé (2016).

the last two decades. According to a poll by Gallup in 2017 in the United States, confidence in printed newspapers stands at around 27%, whereas it is even lower for other media such as television news (24%) and internet news (16%)[3].

Moreover, our work can pose a caveat on the idea that pluralism should bolster the quality of news, which is the backbone of theories such as the *marketplace for ideas* and that is also predicted by models such as Besley and Prat (2006). In other words, whilst pluralism and competition might insure against the risk of capture by interest groups, the decrease in fact-checking might act as a countervailing effect.

3.1.1 Historical Evidence: Yellow Journalism

Looking at the history of journalism there are several examples of how competition can lead to a lowering of publication standards: an interesting case in point is the so called Yellow Journalism period in the United States towards the end of the nineteenth century. As the number of media outlets increased, prices decreased and competition started to be centred on circulation, especially in large cities such as New York, where entrepreneurial and ambitious media owners such as Joseph Pulitzer and William Randolph Hearst led the industry. Competition was fierce and newspapers battled to attract potential buyers on the streets with enlarged headlines mentioning sensational, scandal-ripe and often completely unsubstantiated stories [4].

Consistently with this view, Zaller (1999) underlines how the first half of the twentieth century saw both a toning down from the sensationalism and muckraking of the yellow journalism era and a dramatic fall in competition.

One of the consequences of the aggressive reporting style of the yellow journalism era can be found in the Spanish-American war of 1898. The newspapers led by Pulitzer and Hearst, as a matter of fact, played a decisive role in making public opinion call for a war. One of the highlights of the media campaign against Spain was the stream of accusations (mostly not backed up by evidence) following the sinking of the USS Maine ship in the Havana Harbour. As the historian Allan Keller wrote: "*Had these publishing titans not decided to slug it out toe to toe, the efforts of the downtrodden Cubans to throw off the yoke of Spanish oppression might never have burgeoned into a war between Spain and the United States*".

The story of the Spanish-American war also brings to mind the much more recent case of Iraq and the alleged weapons of mass destruction possessed by Saddam Hussein's regime;

[3] http://www.gallup.com/poll/212852/confidence-newspapers-low-rising.aspx

[4] For an account of yellow journalism and circulation war, refer for example to https://publicdomainreview.org/collections/yellow-journalism-the-fake-news-of-the-19th-century/

that is another important example of a competitive and pluralistic media failing to debunk a fake story, which then led to a tragic and costly war.

3.1.2 Selection of Related Literature

The potentially negative effects of media competition have been picked up by media scholars. This is a quote from a book by Thompson (2013): "*The pressure to run a story before one's competitors acts as an incentive to disclose information that could spark off a scandal, or which could fuel a scandal which is already underway*" [5].

From an empirical point of view, the question of what are the consequences of a more pluralistic media market has been addressed by several scholars. For example, Gentzkow, Shapiro, and Sinkinson (2011) use a long time series of newspaper entry and exit to study its effects on political participation and electoral competition, focussing on the years 1869-1928. They find that newspaper entry increases turnout but they find it has no significant effects on incumbency advantage. Despite not being statistically significant, their point estimates of the effect of an additional newspaper on incumbency advantage are negative, i.e. in the direction predicted by our model. Drago, Nannicini, and Sobbrio (2014) carry out a similar exercise with data on Italian local newspapers. They find, in line with Gentzkow, Shapiro, and Sinkinson (2011), a positive effect of the number of newspapers on voters' participation in elections. In terms of incumbency advantage, they find an increase in the re-election probability of mayors who decide to rerun (they find no significant difference in the probability for incumbents to run for re-election). The authors claim that the positive effect on incumbency advantage is mostly due to increased incentives rather than selection: in fact, they find that an increase in the number of newspapers has no effect on the characteristics of elected officials, but it positively affects the efficiency of public policy.

Another paper addressing the effects of an increase in the number of media outlets is the above-mentioned Besley and Prat (2006). Their model shows that media pluralism decreases the risk of capture by corrupt politicians. The main idea is that as the number of media increases, a corrupt politician or interest group would have to pay monopolist profits to each outlet in order to prevent the publication of a scandal: therefore, the larger the number of media, the more expensive it is for interest groups to prevent the publication

[5] In a similar way, Garrard and Newell (2006) claim that: "[...] *modern scandals are mediated, shaped to varying degrees by the priorities of those reporting them. This has rightly led some commentators to wonder whether the priorities of capitalist (even public-service) media competition have produced behaviour disfunctional for the liberal democracies that modern industrial capitalism tends to produce. [...] Whilst the latter require the spread of serious information and debate, the competitive priorities of the former, particularly mass-circulation tabloids, point increasingly to sensationalism, titillation, entertainment and trivialisation.*"

of a corruption scandal and the better for voters. Our model shows that if the concern is not capture but reporting accuracy, competition can instead be detrimental for voters welfare.

Our work is also related to Cagé (2017). Her model is based on vertical product differentiation and it shows that the effect of entry on quality depends on the heterogeneity of readers: with no heterogeneity, there is no change in quality but simply a splitting of the market, whereas with heterogeneous readers newspapers differentiate on quality in order to soften price competition. Moreover, with heterogeneity on more dimensions, duopolists reduce quality on the less heterogeneous dimension. In our model, on the other hand, readers are homogeneous, but nonetheless we get that an increase in the number of firms leads to news of lower quality. Whereas we abstract from price competition, firms compete on breaking the news, and the cost of quality is represented by increased time to publication: the increased competition on being first on a news is what leads to a decrease in quality as the number of firms increases. Finally, compared to Cage's model, our model can deal with any number of firms and not just monopoly versus duopoly.

Finally, Gratton, Holden, and Kolotilin (2016) present a model of strategic leak timing which is also connected with out work. In their model, good and bad leakers (who respectively have a true or a fake piece of information on a political scandal), try to influence the outcome of an election. In their model the media is not specifically modeled, but once a leak is released, a learning process takes place, which can uncover the truth. In our model, instead, the initial leak reaches all media at the same time and we focus on the gatekeeping role of profit maximizing media outlets in deciding whether to release the information, with the objective of media being profit.

3.2 The Model

Consider a media market composed by N media outlets playing across 2 periods. In each period, a state of the world $\omega \in \{0, 1\}$ is independently drawn such that $Pr(\omega = 1) = p$. When the realization of the state is $\omega = 1$ there is a political scandal, whereas $\omega = 0$ means that there is no scandal. We will later addition this simple model of media competition without politicians with an electoral choice between an incumbent and a challenger in order to evaluate the effects of news reporting on elections.

Each media i receives a signal s_i about ω, distributed according to some full support density function $f_\omega(s)$ for each state of the world ω, with cumulative distribution function $F_\omega(s)$. Let $\psi(s) = \frac{f_1(s)}{f_0(s)}$ denote the likelihood ratio at s. We will assume that $\psi(s)$ is increasing in s which implies $F_1(s) > F_0(s)$ for all s (first order stochastic dominance).

Notice that this assumption means that higher values of the signal are more likely in case of scandal. Furthermore, we will also assume that $\lim_{s \to +\infty} \psi(s) = +\infty$ and $\lim_{s \to -\infty} \psi(s) = 0$, so that posterior beliefs converge to 0 and 1 when s goes to $-\infty$ and $+\infty$ respectively.

After observing s_i, each media company simultaneously decides whether to publish announcing a scandal or whether to fact-check the information received with a new signal. We assume that fact-checking allows the media to receive a fully informative signal of the state of the world before deciding whether to publish the scandal or not.

The size of the media market is normalized to 1 and we assume that the revenues from publishing a scandal are equally split among the media outlets who published the scandal first. In particular, this means that the revenue from publishing a fact-checked scandal that was already published by another media outlet without fact-checking is 0.

Publishing fake scandals is costly for media because at the end of the first period, the state of the world is exogenously revealed and the media outlets that published a fake scandal are replaced by an equal number of identical ones.

In the second period, the game is repeated. We assume that the value of the market in the second period is $R > 1$. The reason for this assumption is that we think of the second period as a reduced form for all future periods in an infinitely repeated game.

3.3 Analysis

Let's analyze the model starting from the second period. In the second period there is no disciplining effect from the possibility of being replaced. Therefore, all media publish, no matter what the state of the world is. It follows that the utility from staying in the market in period 2, or the opportunity cost of publishing a fake scandal, is $c = \frac{R}{N}$.

Let's now move to period 1: each media infers the state of the world conditional on the signal they received using Bayesian updating. Given the prior p that there is a scandal, each media updates according to the posterior:

$$\hat{p}(s) = \frac{pf_1(s)}{pf_1(s) + (1-p)f_0(s)}, \tag{3.1}$$

Lemma 1. *$\hat{p}(s)$ is increasing in s and the image of $\hat{p}$ is $(0, 1)$.*

Proof. All proofs can be found in the appendix. □

Let's for a moment focus on a single media company. Denote by r_j the revenue from publishing without fact-checking conditional on the state $\omega = j$ for $j \in \{0, 1\}$. Finally, let's define by $\gamma \in [0, 1]$ the probability that none of the other media publish a scandal without fact checking conditional on the scandal being true. Notice that these quantities depend on the equilibrium behaviour of the media firms and will be made endogenous in the following pages. For now, assume that $R > N$, meaning that c is large enough such that $c > 1$: in other words, publishing a certainly fake news is worse than not publishing; moreover, notice that by construction $r_1 \geq \frac{1}{N}$, since at worst all media publish without fact-checking and the revenue is $\frac{1}{N}$.

Lemma 2. *When $N = 1$, the monopolist media always fact-checks the scandal. When $N > 1$, in any equilibrium a media outlet uses a strategy characterized by a cut-off point s^*, such that the media outlet publishes if $s > s^*$ and fact-checks otherwise.*

From the perspective of the monopolist, fact checking is always better than publishing directly, because she does not face the risk that another media publishes without fact-checking, leaving her without market revenues. On the contrary, when there are more than two firms, they have to trade-off the informational gain of fact-checking with the probability of having less revenues either because another media published without fact-checking or because if they publish after fact-checking the revenues are always split with all other media.

We still need to prove the existence of the equilibrium. In particular, let's consider a symmetric equilibrium. From the previous lemma we know that if s^* is the threshold that characterizes the equilibrium strategy of a media outlet, it has to be that $\gamma = F_1(s^*)^{N-1}$. Moreover, in a symmetric equilibrium, we can also rewrite the expressions of r_j as functions of s^*. In particular:

$$r_j = \sum_{k=0}^{N-1} \frac{1}{k+1} \binom{N-1}{k} (1 - F_j(s^*))^k F_j(s^*)^{N-1-k}$$

First of all, we will prove that the expected revenues of publishing increase in the threshold s^* used by the opponents. In other words, r_j is an increasing function of s^*.

Lemma 3. $r_j = \frac{1}{N} \frac{1-F_j(s^*)^N}{1-F_j(s^*)}$ *and r_j is strictly increasing in s^*.*

The fact that r_j is strictly increasing in s^* means that the revenue from publishing without fact-checking is higher if the other media require a higher threshold for publishing. This gives the media outlet an incentive to publish without fact checking. However, a larger s^* also translates into a lower probability of direct publishing for the other media. Therefore, also fact checking becomes more profitable, since it becomes less likely that one of the competitors published without fact-checking.

Having characterized the revenues from publishing, we can go back to our equilibrium in cutoff strategies. A symmetric equilibrium requires the following fixed-point equation to hold:

$$s^* = \hat{p}^{-1}\left(\frac{c - r_0}{c - r_0 + r_1 - \frac{F_1(s^*)^{N-1}}{N}}\right).$$

Theorem 1. *The game has a unique symmetric equilibrium, in which all media outlets publish the news without fact-checking if $s > s^*$ and fact-check if $s \leq s^*$.*

The fact that $\frac{c-r_0}{c-r_0+r_1-\frac{F_1(s^*)^{N-1}}{N}}$ is decreasing in s^* means that from the perspective of an individual media outlet, higher standards in the industry (i.e. a higher s^*) mean a lower indifference point in terms $\hat{p}(s)$ to publish without fact checking. In other words, this model describes an environment in which there is an incentive to free ride on the high fact-checking standards of the media industry.

What can we say about the threshold s^* that media outlets use to decide whether to fact check or publish? A natural questions concerns the amount of information that will make media indifferent between the two options. The following lemma finds sufficient conditions for $\hat{p}(s^*)$ to be larger or smaller than p. In the case of $\hat{p}(s^*) > p$, media outlets only ever report a scandal when the information they receive makes them more confident than the prior about the existence of the scandal. In other words, the bar for publication is higher than the prior. If instead $\hat{p}(s^*) < p$, there are situations in which the evidence is against the scandal but the outlet nonetheless decides to publish. Whereas we rule out *fully* fake news with the assumption that $R < N$, publishing a scandal despite evidence going against it might also be considered a (slightly milder) version of fake news.

Lemma 4. *In equilibrium, if $N < (1-p)R+p$, then $\hat{p}(s^*) > p$. If $p > \frac{R-1}{R}$, then $\hat{p}(s^*) < p$.*

We will see in the next section that this property is important for the welfare implications of the model.

3.3.1 Fact-checking and competition

So far we have proved that there is more fact checking under monopoly than when there are two or more firms competing. In the next proposition we generalize this result to an arbitrary increase in the number of firms:

Proposition 1. *Increasing the number of firms decreases fact-checking, i.e. s^* decreases in N.*

The intuition of the result is the following: an increase in the number of firms makes fact-checking less profitable because it increases the probability that another firm publishes without fact-checking and it also increases the number of firms to share the revenues with in case no other media publishes without fact-checking. Moreover, sharing the market with a larger number of firms decreases the value of being in the market in the second period. As a result, increasing N leads media outlets to have lower standards for publishing a scandal. This result can be seen as a caveat to the reliability of competitive markets to deliver informative and fact-based media commentary, as maintained by the proponents of the theory of the marketplace of ideas.

3.3.2 Large N scenarios

The analysis above rests on the assumption that $N < R$. This assumption makes sure that the profits from publishing a fake news as a monopolist, given by 1, are smaller than the value of remaining in the market, given by $\frac{R}{N}$. Therefore, when $R < N$ no fake news in the most strict sense are published: conditional on being certain about the scandal, media outlets only publish true scandals. In other words, once fact-checking has taken place, only true news are published.

What happens if $N > R$? If nobody published fake news after the fact-checking stage, since $N > R$ any firm would have an incentive to unilaterally deviate and publish independently of the results of fact checking. Similarly, it cannot be the case that all firms publish with probability one independently of the fact-checking outcome. If that were the case, in fact, it would always be optimal to publish after receiving the leak rather than waiting. It follows that in any equilibrium with $N > R$, media outlets must mix when fact-checking reveals the scandal to be fake: this in turn means that in equilibrium, the expected return from publishing a fake news has to be equal to the cost $\frac{R}{N}$. Other than that, the main features of the equilibrium remain the same as in the $N < R$ case, and in particular the characterization of the strategy in the first stage does not change. We can summarize this in the following proposition:

Proposition 2. *When $N > R$, media outlets follow a cutoff strategy in the first stage, with s^* being described by the same condition as in the game with $N < R$. Conditional on fact-checking, media outlets always publish true scandals whereas they publish fake scandals with probability $1 - \sigma$ increasing in N.*

Notice that as N goes to infinity, σ converges to zero, meaning that fake scandals are published with higher and higher probability even after fact-checking. Moreover, notice that independently of the mixing, even in the $N > R$ game news published at the fact checking stage are more informative of a scandal than news published without fact checking, since true scandals are always published and fake scandals are always published with lower probability after fact checking compared to before fact checking. The reason for this is that, in equilibrium, $\frac{1-F_0(s^*)^N}{1-F_0(s^*)} = R - \frac{1-F_1(s^*)^{N-1}}{1-F_1(s^*)}\frac{\hat{p}}{1-\hat{p}} < R$. As a result, it has to be the case that $\sigma > F_0(s^*)$, meaning that publication when the news is fake is more likely in the first stage.

3.4 Fact-checking and Political Accountability

This section adds an electoral choice to our model of media competition in order to evaluate how media behaviour influences the choices of voters, therefore influencing whether politicians are re-elected or ousted from office.

There are two candidates. Each of them can be *corrupt* or *honest*, depending on whether he is involved in a scandal or not. The utility from electing a clean candidate is 1, that from electing a dirty one is normalized at zero. Assume that both candidates have the same ex ante quality, meaning that they have the same unconditional probability of being dirty, and for the time being let's assume that voters and media outlets assign the same prior probability to the candidate being dirty, denoted by p as in the above analysis. Moreover, let's assume that only the incumbent can be involved in a scandal newspapers can write about (for example because scandals involve their behaviour in office, or because the scandal of the incumbent will only be realized if he or she are elected). We will assume that voters are fully rational and update their prior by both reading about a scandal and not reading about a scandal (in other words, no news is good news).

In this section we will assume that readers only consult one media outlet. Since politicians are characterized by the same prior p, all that matters for the electoral choice is the direction of the update and not the size. Therefore, the electoral decision is the same independent of whether the reader consumes a fact-checked or a non-fact checked piece of news: any informative news of a scandal will lead to the dismissal of the incumbent in favour of the challenger [6] One explanation for that might be that readers do not know the timing of the leak and therefore cannot infer from the timing of publication whether the news is fact-checked or not [7].

[6] An extension of the model with different priors p for challenger and incumbent would make the intensity of information important.

[7] A model exploring the dynamics connected to the timing of leaks is Gratton (2016), in which the credibility of news depends on the timing of the release.

For maximum simplicity, let's first assume that the economy is composed of only one reader [8]. The first scenario we are going to analyze is one in which the reader picks randomly one of the media outlets. The outlet can be the same or change across the two stages of the game. The question we would like to answer is whether, in this economy, more media lead to a higher or lower probability of electing an honest politician (which we sometimes denote as welfare).

As we know from the analysis in the previous section, increasing N decreases the threshold s^* that each media uses to decide whether to publish without fact checking. As a result, the reader is more likely to encounter a scandal when consulting the news media and therefore she is less likely to vote for the incumbent. This means that dirty incumbents are less likely to be re-elected, but at the same time also clean incumbents are less likely to be re-elected. The following proposition proves that the trade-off can be resolved in both ways.

Proposition 3. *When the reader selects one outlet randomly, the probability of having a clean politician in office can increase or decrease with N.*

A necessary (but not sufficient) condition for an increase in N to be welfare improving is therefore that $f_1(s^*) > f_0(s^*) + Nf_1(s^*)F_1(s^*)^{N-1}$. In other words, the equilibrium threshold for publication $\hat{p}(s^*)$ has to be sufficiently larger than p, meaning that media outlets switch from fact-checking to publishing only when sufficiently bad news about the scandal arrives. From Lemma 5 we know that this can only happen if $N < (1-p)R + p$. Notice that if fact checking were not possible, the optimal threshold for publication would be exactly the one distinguishing bad news from good news, i.e. $\hat{p}(s^*) = p$. As a result, as we increase N, the probability of having fact-checking decreases (both mechanically by the increase in N and indirectly through the decrease in s^*), but at the same time the decrease in s^* might be welfare improving given the decreased probability of having fact checking.

In order to consider a case in which fact-checking is not relevant for welfare except that through s^*, let's now assume that elections are imminent and that any fact checked news will therefore necessarily arrive after the new leader has been elected. In this situation, non-fact checked news actually might serve a socially beneficial purpose, i.e. that of providing information on a scandal in time for the electoral choice, and in fact we show that increasing N always increases welfare:

Proposition 4. *With imminent elections, increasing N increases welfare if and only if*

[8] With multiple readers, the results of this section would have to account for the probability that a majority of readers read a media publishing or not publishing the scandal in the first period. However, the intuition of the results would remain very similar.

$\hat{p}(s^*) > p$.

The intuition of this result is that if fact checking is not useful for electoral purposes, the optimal publication threshold s^* is finite. In particular, given the symmetry of the problem, a media outlet trying to maximize the voter's welfare would report the scandal if $\hat{p}(s^*) > p$ and withhold it if $\hat{p}(s^*) < p$. As we have seen from Lemma 5, market competition can make s^* lie both above and below this threshold. Therefore, when elections are imminent increasing the number of media is welfare improving if the market equilibrium makes media too reluctant to report news of a scandal.

Let's now consider a different case, in which the reader, instead of selecting a newspaper at random independently of whether it mentions the scandal or not, reads one among the outlets (if any) which published the news of the scandal. In this case, compared to the previous one, the reader never misses a non-fact checked news. This means that if the politician is corrupt, the reader always ends up knowing it, either through a non-fact checked or through a fact checked news. As a result, all corrupt leaders are voted out of office, but some clean ones are, too. The expression for welfare becomes the following:

$$1 - p^2 - p(1-p)(1 - F_0(s^*)^N)$$

From this expression, it is immediate to see that as N increases, welfare decreases, since $F_0(s^*)$ decreases.

Proposition 5. *If readers read one outlet among those (if any) which talk about the corruption scandal, then welfare decreases as N increases.*

This results tells us that as N increases, welfare increases only if the standards for publication of a scandal become stricter, i.e. s^* increases. However, we showed that what happens is exactly the opposite. The intuition for the welfare decrease is that with many media, it is more likely that one will get a high enough signal for publication. As a result, a reader will be very likely to find a scandal in the news and therefore to vote for the challenger even if the incumbent is clean.

Notice that both our welfare results point out that in the internet age, in which decreased entry barriers for media led to an increase in the number of outlets N, the resulting competitive pressure might be detrimental for welfare. The reason is that fact-checking is more likely to be election-relevant and that, at the same time, readers are more easily exposed to media mentioning a scandal (for example through social media). As a result, when evaluating the effect of the number of media on welfare, details such as the electoral relevance of fact checks and readers selection of the outlet to consult play a fundamental role.

3.5 Discussion

In this section we will discuss some key elements of the model and possible future extensions.

The first fundamental ingredient is the initial signal, that we assume all media receive at the same time. A particularly interesting extension of our model would be to have leaks being spread by interest groups acting strategically. For example, an interest group siding with the challenger might have the interest to spread scandals concerning the incumbent. In this extended game, leakers could have either the timing of the information (in a similar fashion as in Gratton, Holden, and Kolotilin (2016)), or the number of outlets receiving the leak, or even the realization of the signal s.

Another possible extension would be to allow media outlets to invest in fact-checking (for example by employing investigative journalists). This would be reminiscent of industrial organization models of investment in the quality of products with varying degrees of competition.

The other key element of our model is the cost for publishing a fake scandal, which we consider exogenous (it might represent for example libel lawsuits). In real journalism, however, it is often up to competitors to expose as fake the scandal raised by another media outlet. Allowing for similar dynamics could be a significant addition to our model.

Finally, our political accountability results rest on the assumption that politicians have a fixed type, either honest or corrupt. However, it would be interesting to model corruption as an endogenous choice of politicians. In an environment where many fake scandals make the news, politicians might have a stronger incentive to become corrupt, as in a self-fulfilling prophecy.

3.6 Conclusion

This paper shows that increasing the competitive pressure to break a news can lead media outlets to be less demanding in the amount of evidence required to publish a story. In particular, we consider a case in which media outlets can accuse a politician of being involved in a scandal prior to an election: we show that when readers consult with priority one (if any) of the media talking about a scandal, then increasing the number of competitors decreases the probability of having a clean politician in office. Our results aim to pose a caveat to the claim that media pluralism always benefits democracy, suggesting that an increase in competition in the media sector might be socially damaging.

Appendix 3.A Proofs

Lemma 1. *$\hat{p}(s)$ is increasing in s and the image of $\hat{p}$ is $(0,1)$.*

Proof.

$$\hat{p}(s) = \frac{pf_1(s)}{pf_1(s) + (1-p)f_0(s)} = \frac{p}{p + \frac{1-p}{\psi(s)}}$$

And

$$\hat{p}'(s) = \frac{p(1-p)}{(p\psi(s) + (1-p))^2}\psi'(s) > 0$$

Finally, the image of $\hat{p}$ is the set $(0,1)$ because $\hat{p}$ is a continuous function and the $\lim_{s\to+\infty}\hat{p}(s) = 1$ and $\lim_{s\to-\infty}\hat{p}(s) = 0$. □

Lemma 2. *When $N = 1$, the monopolist media always fact-checks the scandal. When $N > 1$, in any equilibrium media uses a strategy characterized by a cut-off point s^*, such that the media publishes if $s \geq s^*$ and fact-checks otherwise.*

Proof. The pay-off from publishing the news is $\Pi_1 = \hat{p}(s)r_1 + (1-\hat{p}(s))(r_0 - c)$. The pay-off from fact-checking is $\Pi_0 = \hat{p}(s)\frac{\gamma}{N}$. The indifference point is such that $\Pi_1 = \Pi_0$ and it yields the following condition:

$$\hat{p}(s^*) = \frac{c - r_0}{c - r_0 + r_1 - \frac{\gamma}{N}}. \tag{3.2}$$

The right-hand-side is a constant and bounded by 0 and 1. To see that it is higher than 0 notice that both the numerator and the denominator are positive. The numerator because by definition $r_0 \leq 1$ and $c > 1$, the denominator because $r_1 \geq \frac{1}{N} \geq \frac{\gamma}{N}$. It is lower than one because the numerator is always lower than the denominator (strictly if $N > 1$).

The left-hand-side is increasing in s and the range is $(0,1)$. Therefore, there always exists a unique s^* that solves the indifference condition and, in any equilibrium, media publishes the scandal if $s \geq s^*$ and fact checks otherwise. □

Lemma 3. *$r_j = \frac{1}{N}\frac{1-F_j(s^*)^N}{1-F_j(s^*)}$ and r_j is strictly increasing in s^*.*

Proof. First notice that

$$\frac{1}{k+1}\binom{N-1}{k} = \frac{1}{k+1}\frac{(N-1)!}{k!(N-1-k)!} = \frac{1}{N}\frac{N!}{(k+1)!(N-1-k)!} = \frac{1}{N}\binom{N}{k+1}.$$

Plugging-in this identity in r_j and multiplying and dividing by $1 - F_j(s^*)$ we get:

$$r_j = \frac{1}{N}\frac{1}{1-F_j(s^*)}\sum_{k=0}^{N-1}\binom{N}{k+1}(1-F_j(s^*))^{k+1}F_j(s^*)^{N-1-k}$$

Let's then add and subtract $\frac{1}{N}\frac{F_j(s^*)^N}{1-F_j(s^*)}$. We can now rewrite the expression as:

$$r_j = \frac{1}{N}\frac{1}{1-F_j(s^*)}\sum_{k=-1}^{N-1}\binom{N}{k+1}(1-F_j(s^*))^{k+1}F_j(s^*)^{N-1-k} - \frac{1}{N}\frac{F_j(s^*)^N}{1-F_j(s^*)}.$$

Now, substituting $k = k' - 1$, we have that the summation of the previous equation is simply the sum of the probabilities of all possible events of a discrete binomial distribution which have to sum 1. Therefore we are left with

$$r_j = \frac{1}{N}\frac{1-F_j(s^*)^N}{1-F_j(s^*)}$$

In order to show that this expression is increasing in s^* we can use the well known formula for the summation of a geometric series to get:

$$r_j = \frac{1}{N}\sum_{k=0}^{N-1}F_j(s^*)^k$$

and it is immediate to verify that this is increasing in s^*, given that for any $k \geq 0$, $F_j(s^*)$ is increasing in s^*.

□

Theorem 1. *The game has a unique symmetric equilibrium, in which all media publish the news without fact-checking if $s > s^*$ and fact-check if $s \leq s^*$.*

Proof. In equilibrium, the following has to hold:

$$\hat{p}(s^*) = \frac{c - r_0}{c - r_0 + r_1 - \frac{F_1(s^*)^{N-1}}{N}}$$

We know that $\hat{p}(s^*)$ is strictly increasing in s^*, approaching 0 as s^* goes to $-\infty$ and approaching 1 as s^* goes to $+\infty$. As far as the right hand side is concerned, we can rewrite it as: $\frac{1}{1+\frac{r_1 - F_1(s^*)^{N-1}/N}{c-r_0}}$. We know that as s^* goes to $+\infty$, r_j and $F_j(s^*)$ go to 1. If s^* goes to $-\infty$, on the other hand, r_j goes to $\frac{1}{N}$ and $F_j(s^*)$ goes to 0. It follows that $\lim_{s^*\to+\infty} \frac{c-r_0}{c-r_0+r_1-\frac{F_1(s^*)^{N-1}}{N}} = \frac{c-1}{c-1/N} < 1$ and $\lim_{s^*\to-\infty} \frac{c-r_0}{c-r_0+r_1-\frac{F_1(s^*)^{N-1}}{N}} = \frac{c-1/N}{c} > 0$ and we can see that $\forall N > 1$ $\frac{c-1}{c-1/N} < \frac{c-1/N}{c}$, since $1 + (N-2)cN > 0$. This means that

$\lim_{s^*\to-\infty} \frac{c-r_0}{c-r_0+r_1-\frac{F_1(s^*)^{N-1}}{N}} > \lim_{s^*\to+\infty} \frac{c-r_0}{c-r_0+r_1-\frac{F_1(s^*)^{N-1}}{N}}$. Moreover, it is immediate to verify that $\frac{c-r_0}{c-r_0+r_1-\frac{F_1(s^*)^{N-1}}{N}}$ is continuous. In order to show that it is strictly decreasing, let's focus on the ratio $\frac{r_1-F_1(s^*)^{N-1}/N}{c-r_0}$. It is immediate to verify that the denominator is decreasing in s^*, since r_0 increases in s^*. As far as the numerator is concerned, we can rewrite it as $\frac{1}{N}\frac{1-F_1(s^*)^N}{1-F_1(s^*)} - \frac{1}{N}F_1(s^*)^{N-1}$. This can be rearranged into $\frac{N-1}{N}\frac{1}{N-1}\frac{1-F_1(s^*)^{N-1}}{1-F_1(s^*)}$ which is just $\frac{N-1}{N}r_1(N-1)$, i.e. it is proportional to the revenue when the number of players is $N-1$ instead of N. Since r_1 is increasing in s^* for all N, this is also increasing in s^*. Therefore, $\frac{c-r_0}{c-r_0+r_1-\frac{F_1(s^*)^{N-1}}{N}}$ is strictly decreasing in s^*. However, since $\hat{p}$ is strictly increasing in s^* and since $\lim_{s^*\to-\infty}\hat{p}(s^*) < \lim_{s^*\to-\infty} \frac{c-r_0}{c-r_0+r_1-\frac{F_1(s^*)^{N-1}}{N}}$ and $\lim_{s^*\to+\infty}\hat{p}(s^*) > \lim_{s^*\to+\infty} \frac{c-r_0}{c-r_0+r_1-\frac{F_1(s^*)^{N-1}}{N}}$, there exists a unique s^* solving the above equation. Hence the symmetric equilibrium exists and is unique. □

Lemma 4. *In equilibrium, if $N < (1-p)R+p$, then $\hat{p}(s^*) > p$. If $p > \frac{R-1}{R}$, then $\hat{p}(s^*) < p$.*

Proof. Consider the indifference condition $\hat{p}(s^*) = \left(\frac{c-r_0}{c-r_0+r_1-\frac{F_1(s^*)^{N-1}}{N}}\right)$. The right-hand side takes values in $(\frac{c-1}{c-\frac{1}{N}}, \frac{c-1/N}{c})$ A sufficient condition for $\hat{p}(s^*) > p$ is therefore that $p < \frac{c-1}{c-\frac{1}{N}}$. We can rewrite this as $c > \frac{1-\frac{p}{N}}{1-p}$. On the other hand, a sufficient condition for $\hat{p}(s^*) < p$ is that $p > \frac{c-1/N}{c}$, which can be rewritten as $p > \frac{R-1}{R}$. □

Proposition 1. *Increasing the number of firms decreases fact-checking.*

Proof. Let s^* be the equilibrium threshold. Thus s^* solves the indifference condition:

$$\hat{p}(s^*) = \frac{c-r_0}{c-r_0+r_1-\frac{F_1(s^*)^{N-1}}{N}}$$

Now, let's keep s^* fixed. Notice that the left-hand-side of the indifference condition does not depend on N, which enters only on the right-hand side. The right-hand side of the expression can be rewritten as $\left\{1+\frac{\frac{1}{N}\frac{1-F_1(s^*)^{N-1}}{1-F_1(s^*)}}{\frac{R}{N}-\frac{1}{N}\frac{1-F_0(s^*)^N}{1-F_0(s^*)}}\right\}^{-1}$. Let's focus on the ratio contained in this term. We can cancel out $\frac{1}{N}$ and it is straightforward to check that, increasing N and fixing s^*, the numerator increases while the denominator decreases. Hence, the right hand side of the indifference condition decreases; therefore an increase in N, has to decrease $\hat{p}(s^*)$ and this happens only if s^* decreases. Thus, s^* is decreasing in N for all N. □

Proposition 2. *When the reader selects one outlet randomly, the probability of having a clean politician in office can increase or decrease with N.*

Proof. The probability of having a clean politician in office is $p((1-F_1(s^*)+F_1(s^*)^N)(1-p)+(1-p)(1-F_0(s^*))p+(1-p)F_0(s^*)$ This can be rearranged to yield:

$$W=(1-p)^2+p(1-p)[1-F_1(s^*)+F_1(s^*)^N+F_0(s^*)]$$

Taking the derivative with respect to N results in the following expression:

$$p(1-p)\left[-f_1(s^*)+f_0(s^*)+Nf_1(s^*)F_1(s^*)^{N-1}\right]\frac{\partial s^*}{\partial N}+F_1(s^*)^N\ln F_1(s^*),$$

which is not unambiguously positive or negative. □

Proposition 3. *With imminent elections, increasing N increases welfare if and only if $\hat{p}(s^*)>p$.*

Proof. In this scenario, welfare can be expressed in the following way:

$$(1-p)^2+p(1-p)(1+F_0(s^*)-F_1(s^*))$$

Notice that since $F_0(s^*)\geq F_1(s^*)$ and $\lim_{s^*\to\infty}F_0(s^*)=\lim_{s^*\to\infty}F_1(s^*)$, in this case welfare is maximized when $f_0(s^*)=f_1(s^*)$. Notice that by definition of $\hat{p}$, $f_1(s^*)=f_0(s^*)$ implies that $\hat{p}(s^*)=p$. Therefore, we can use Lemma 5 to characterize sufficient conditions for $\hat{p}(s^*)$ to lie above or below p. □

Proposition 4. *If readers read one outlet among those (if any) which talk about the corruption scandal, then welfare decreases as N increases.*

Proposition 5. *When $N>R$, media outlets follow a cutoff strategy in the first stage, with s^* being the same as in the game with $N<R$. Conditional on fact-checking, media outlets always publish true scandals whereas they publish fake scandals with probability $1-\sigma$.*

Proof. Let's start from the second stage, i.e. after fact-checking has taken place. If the scandal is true, all firms publish it. If the scandal is wrong and no firm publishes it, then each firm as the incentive to unilaterally deviate and publish it, since the monopolistic revenue 1 is larger than $\frac{R}{N}$. At the same time, if all firms were to publish the fake scandal, then it would be optimal to always publish the scandal in the first without fact-checking. It follows that media outlets must mix when the scandal is proved to be fake. Using an analogous formula to (3.2), where σ denotes the probability of not publishing the fake news after fact-checking, the expected revenue from publishing a fake news, given that all other media outlets use the same strategy, is $\frac{1}{N}\frac{1-\sigma^N}{1-\sigma}$. In equilibrium, this expected revenue has to equal the cost $\frac{R}{N}$. Notice that if $N > R$, there always exists a $\sigma < 1$ such that $\frac{1}{N}\frac{1-\sigma^N}{1-\sigma} = \frac{R}{N}$. To see this, rearrange the condition to get $\frac{1-\sigma^N}{1-\sigma} = R$. The left-hand side is a strictly increasing function of σ with image in $[1, N]$. As a result, there is a unique $\sigma \in (0, 1)$ such that the condition is satisfied. Let's now move to the analysis of the first stage, i.e. the decision to fact check. The only change compared to the case of $N < R$ is in the case of the news being fake. However, we just proved that the expected revenue from publishing conditional on fact-cehecking indicating that the scandal is fake is equal to zero. Therefore, the payoff for the media outlet conditional on fact-checking and the scandal being fake is the same as before. This means that the indifference condition determining s^* remains the same. In terms of existence of the equilibrium, since the indifference condition is the same, the only change concerns the limit of the right hand side $\frac{c-r_0}{c-r_0+r_1-\frac{F_1(s^*)^{N-1}}{N}}$ as s^* goes to infinity, represented by $\frac{c-1}{c-\frac{1}{N}}$. Whereas with $N < R$ this limit is strictly positive, since $c > 1$, as N grows larger c becomes smaller than 1 and $\frac{c-1}{c-\frac{1}{N}}$ tends to $-\infty$. In other words, compared to the case of $N < R$, $\hat{p}$ at the equilibrium s^* is no longer bounded below, meaning that as N grows, the cutoff s^* grows smaller and smaller. □

Bibliography

ABRAMOWITZ, A. I. (2010): "The Disappearing Center: Engaged Citizens, Polarization and the American Democracy," *Yale*.

ADAMS, J., AND S. MERRILL (2008): "Candidate and Party Strategies in Two-Stage Elections Beginning with a Primary," *American Journal of Political Science*.

AGHION, P., AND M. O. JACKSON (2016): "Inducing Leaders to Take Risky Decisions: Dismissal, Tenure, and Term Limits," *American Economic Journal: Microeconomics*, 8(3), 1–38.

AGRANOV, M. (2016): "Flip-Flopping, Primary Visibility, and the Selection of Candidates," *American Economic Journal: Microeconomics*, 8(2), 61–85.

ANSOLABEHERE, S., J. J. SNYDER, AND C. STEWART III (2001): "Candidate Positioning in U.S. House Elections," *American Journal of Political ...*, 45(1), 136–159.

ASHWORTH, S., AND K. W. SHOTTS (2010): "Does informative media commentary reduce politicians' incentives to pander?," *Journal of Public Economics*, 94(11-12), 838–847.

BESLEY, T., AND A. PRAT (2006): "Handcuffs for the grabbing hand? Media capture and government accountability," *The American Economic Review*, 96(3), 720–736.

BRADY, D., H. HAN, AND J. POPE (2007): "Elections and Candidate Ideology : Out of Step with the Primary Electorate ?," *Legislative Studies Quarterly*, 32(1), 79–105.

BULLOCK, W., AND J. CLINTON (2011): "More a Molehill than a Mountain: The Effects of the Blanket Primary on Elected Officialsâ Behavior from California," *The Journal of Politics*, 73(3), 915–930.

CAGÉ, J. (2016): *Saving the Media*. Harvard University Press.

——— (2017): "Media Competition, Information Provision and Political Participation: Evidence from French Local Newspapers and Elections, 1944-2014," Discussion paper, CEPR Discussion Papers.

Calvino, I. (1969): *Prima che tu dica pronto*. Mondadori.

Canes-Wrone, B., M. C. Herron, and K. W. Shotts (2001): "Leadership and Pandering: A Theory of Executive Policymaking," *American Journal of Political Science*, 45(3), 532–550.

Carey, J., and J. Polga-Hecimovic (2006): "Primary Elections and Candidate Strength in Latin America," *The Journal of Politics*, 68(3), 530–543.

Casas, A. (2014): "Ideological Extremism and Primaries," *Working Paper*, pp. 1–28.

CFOI (1997): "Freedom of Information:Key Issues," .

Cho, I., and D. Kreps (1987): "Signaling games and stable equilibria," *The Quarterly Journal of Economics*, 102(2), 179–221.

Dewan, T., and R. Hortala-Vallve (2014): "Electoral Competition, Control and Learning," *Working Paper*, pp. 1–31.

Doherty, D., C. M. Dowling, and M. G. Miller (2015): "When is Changing Policy Positions Costly for Politicians? Experimental Evidence," *Political Behavior*, 38(2), 455–484.

Drago, F., T. Nannicini, and F. Sobbrio (2014): "Meet the press: How voters and politicians respond to newspaper entry and exit," *American Economic Journal: Applied Economics*, 6(3), 159–188.

Elisabeth R. Gerber, R. B. M. (1998): "Primary Election Systems and Representation," *Journal of Law, Economics, and Organization*, 14(2), 304–324.

Fearon, J. D. (1999): "Electoral Accountability and the Control of Politicians: Selecting Good Types versus Sanctioning Poor Performance," in *Democracy, Accountability and Representation*, ed. by A. Przeworski, and S. C. Stokes, chap. 2, pp. 55–97. Cambridge University Press.

Frisell, L. (2009): "A theory of self-fulfilling political expectations," *Journal of Public Economics*, 93(5–6), 715 – 720.

Fu, Q., and M. Li (2014): "Reputation-concerned policy makers and institutional status quo bias," *Journal of Public Economics*, 110, 15–25.

Garrard, J., and J. L. Newell (2006): *Scandals in past and contemporary politics*. Manchester University Press.

Gentzkow, M., and J. M. Shapiro (2006): "Media Bias and Reputation," *Journal of Political Economy*, 114(2), 280–316.

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

Gratton, G., R. Holden, and A. Kolotilin (2016): "When to Drop a Bombshell," Discussion Papers 2016-13, School of Economics, The University of New South Wales.

Grofman, B., O. Troumpounis, and D. Xefteris (2016): "Electoral competition with primaries and quality asymmetries," pp. 1–25.

Hirano, S., J. M. J. Snyder, S. Ansolabehere, and J. M. Hansen (2010): "Primary Elections and Partisan Polarization in the U.S. Congress," *Quarterly Journal of Political Science*, 5(2), 169–191.

Hummel, P. (2010): "Flip-flopping from primaries to general elections," *Journal of Public Economics*, 94(11–12), 1020 – 1027.

——— (2013): "Candidate strategies in primaries and general elections with candidates of heterogeneous quality," *Games and Economic Behavior*, 78, 85–102.

Kartik, N., and R. McAfee (2007): "Signaling character in electoral competition," *The American economic review*.

Kaufmann, K., J. Gimpel, and A. Hoffman (2003): "A Promise Fulfilled? Open Primaries and Representation," *The Journal of Politics*, 65(2), 457–476.

Kemahlioglu, O., W.-S. R., and S. Hirano (2009): "Why Primaries in Latin American Presidential Elections?," *The Journal of Politics*, 71(1), 339–352.

Levendusky, M. S., and M. Horowitz (2012): "When Backing Down Is the Right Decision: Partisanship, New Information and Audience Costs," *Journal of Politics*, 74, 323–338.

Levy, G. (2004): "Anti-herding and strategic consultation," *European Economic Review*, 48(3), 503–525.

——— (2007): "Decision Making in Committees: Transparency, Reputation, and Voting Rules," *The American Economic Review*, 97(1), 150–168.

Li, W. (2007): "Changing One's Mind When the Facts Change: Incentives of Experts and the Design of Reporting Protocols," *The Review of Economic Studies*, 74(4), 1175–1194.

Majumdar, S., and S. W. Mukand (2004): "Policy Gambles," *American Economic Review*, 94(4), 1207–1222.

Maskin, E. S., and J. Tirole (2004): "The Politician and the Judge: Accountability in Government," *American Economic Review*, 94(4), 1034–1054.

McGhee, E., S. Masket, B. Shor, and N. McCarty (2014): "A Primary Cause of Partisanship? Nomination Systems and Legislator Ideology," *America Journal of Political Science*, 58(2), 337–351.

Patacconi, A., and N. Vikander (2015): "A model of public opinion management," *Journal of Public Economics*, 128, 73 – 83.

Prat, A. (2005): "The Wrong Kind of Transparency," *The American Economic Review*, 95(3), 862–877.

Prendergast, C., and L. Stole (1996): "Impetuous Youngsters and Jaded Old-Timers: Acquiring a Reputation for Learning," *The Journal of Political Economy*, 104(6), 1105–1134.

Serra, G. (2011): "Why primaries? The party's tradeoff between policy and valence," *Journal of Theoretical Politics*, pp. 1–43.

Seth J. Hill, C. T. (2015): "A Disconnect in Representation? Comparison of Trends in Congressional and Public Polarization," *The Journal of Politics*, 77(4), 1058–1075.

Snyder, J., and S. Hirano (2014): "Primary Elections and the Quality of Elected Officials," *Quarterly Journal of Political Science*, (4), 473–500.

Snyder, J. M. J., and M. M. Ting (2011): "Electoral Selection with Parties and Primaries," *American Journal of Political Science*, 55(4), 782–796.

Stone, W., and E. Simas (2010): "Candidate Valence and Ideological Positions in U . S . House Elections," *American Journal of Political Science*, 54(2), 371–388.

Strömberg, D. (2015): “Economic and Social Impacts of the Media,” *Handbook of Media Economics*, 1.

Takayama, S. (2014): “A Model of Two-stage Electoral Competition with Strategic Voters,” *Working Paper*, pp. 1–36.

Tavits, M. (2007): “Principle vs. pragmatism: Policy shifts and political competition,” *American Journal of Political Science*, 51(1), 151–165.

Thompson, J. B. (2013): *Political scandal: Power and visability in the media age*. John Wiley & Sons.

Tomz, B. M., and R. V. Houweling (2012): “Political Repositioning,” *Working Paper*, pp. 1–34.

Tomz, M., and R. P. Van Houweling (2012): “Candidate Repositioning,” pp. 1–43.

Zaller, J. (1999): “Market competition and news quality,” in *Annual Meetings of the American Political Science Association, Atlanta, GA.* Citeseer.