A Commitment Theory of Populism*

Massimo Morelli† Antonio Nicolò‡ Paolo Roberti §

September 25, 2021

Abstract

A reduced trust in the reliability of politicians is responsible for the demand of simple policies that are easy to monitor for voters. Disenchanted citizens prefer politicians who stick with their easy to check promises (committed delegates) to politicians who propose themselves as competent policy makers but without a specific policy commitment (trustees). In a two-party competition, an asymmetric equilibrium may exist where voters with lower interest for the common good select a committed delegate, while those with higher interest for the common good appoint a trustee. In this asymmetric equilibrium, we show that the committed candidate chooses also all the other strategies that are typically connected to populism in the literature. Hence, this paper can be considered a commitment theory of populism.

Keywords: Populism, Competence, Commitment, Information Acquisition, Interest Groups, Moral Universalism.

JEL Codes: D72, D78.

*We are grateful to Georgy Egorov, Ben Enke, Gabriele Gratton, Edoardo Grillo, Matt Jackson, Gilat Levy, Julien Manili, Dana Sisak, Jim Snyder, Ken Shotts, Konstantin Sonin for helpful comments and interesting discussions. Comments by the (virtual) seminar participants at MEDS, at the World Congress of the Econometric Society 2020 and Formal Theory Society in 2021 are also acknowledged. We are grateful to the European Research Council (Advanced Grant 694583) and to MIUR (for the PRIN grant prot. 2015EL3MRC). The usual disclaimer applies.

†Bocconi University, IGIER and CEPR. Email: massimo.morelli@unibocconi.it

‡University of Padua and University of Manchester

§Free University of Bozen-Bolzano

1 Introduction

The recent literature on the populism wave that emerged in the aftermath of the great recession has focused mostly on what may have caused it on the *demand* side, i.e., primarily looking at the changes in attitudes, perceptions and political actions of voters in the 21st century. A first set of economists and some political scientists have focused on economic insecurity as a main source of mistrust in traditional institutions and demand of protection policies (see, e.g., Algan et al., 2017, Guiso et al., 2019 and 2021, Ananyev and Guriev, 2018).[1] A parallel strand of political science literature focused instead on cultural causes of distrust in political institutions (see, e.g., Norris and Inglehart, 2019). Both the economic and cultural mechanisms, well summarized in the survey article by Guriev and Papaioannou (2021), may have played a significant role for the sharp increase of demand of populism. Such a demand of populism mainly takes the form of a demand of policies and credible promises *for the people* combined with strong anti-elite sentiments.[2] This can also be seen as a demand of *protection*, which in turn can be protection from the immigration or globalization or automation threat if the cause is mainly economic, or can take the form of identity and cultural protection otherwise.[3] These key observations – the increased anti-elite sentiment and lower trust in representative democracy on the one hand, and the greater demand of protection on the other – may together lead voters to desire unconditional policies: brexit, build walls, close harbors and borders, fight Chinese globalization threat, nationalism.

In this paper we take the changes in the key elements of voters' preferences and political trust mentioned above as given, and we focus instead on the *supply* side. More precisely, the question is what causes the entry and success in the political arena of candidates and parties who concentrate their rhetoric on anti-elite statements and focus on closeness to the people, simple

[1] The immigration threat (see, e.g., Laitin, 2018), the globalization threat (see, e.g., Autor et al., 2020, Rodrik 2018, Colantone and Stanig, 2018) and the automation threat (see e.g. Acemoglu and Restrepo, 2017) all increased economic insecurity of certain classes of citizens even before the financial crisis, but Guiso et al. (2021) show that the latter has been the real watershed.

[2] See, e.g., Mudde (2004) for the most recognized definition of populism in political science, which primarily focuses on the pure people against corrupt elite framing.

[3] The connection between antielitism and short term protection can be found in the definition of populism on the encyclopedia Britannica, (www.britannica.com/topic/populism).

protection commitments, anti-experts attitudes. We answer this general question with a new theory that, in a nutshell, establishes the following logical chain: reduced trust implies greater chance of electoral success for a candidate who commits to policies that are easy to understand or monitor; and once a candidate shifts from the traditional trustee model of political agency to this commitment strategy, all the other features of populist supply mentioned above follow as rational complementary strategies.

First, we present a simple principal-agent baseline model, where a representative voter (principal) selects a politician (agent) who chooses a policy. The optimal policy is uncertain at the time the politician is appointed, and depends on the realization of a state of nature. The voter selects either a politician who commits to implement a policy that is ex-ante determined (a committed delegate) or one who promises to choose the ex-post optimal policy for the voter (a trustee). In the latter case, whether the politician will actually choose or not the ex-post optimal policy depends on (i) her competence, (ii) the probability that a lobby distorts her choice, and, (iii) the effort that the voter devotes to monitor her. Given that a third player, i.e., an interest group, a lobby or even a supranational institution, can influence the non-committed politician, then it follows that the choice of commitment is the more likely the greater the expected bias and the probable strength or influence of the interest group, the lower is the expected ability of the potential uncommitted politician, and the lower are the expected benefits that an optimal ex-post policy could bring to the principal compared to the cost of monitoring. A lower trust in politicians' competence, an increased fear that lobbies could capture them, and disenchantment about the benefits that an optimal policy can really bring to her, all shift the voter's preference towards a committed delegate, who panders to the voter's ex-ante desiderata. Since pandering to the ex-ante voter's beliefs is a component of populism, in line with Acemoglu et al. (2013) we call the committed delegate a populist.

In the full-blown model, we introduce electoral competition between two parties. The state-contingent policy is common value, but the voters of the two parties may differ in terms of the weight that they assign to it compared to the non state-contingent component (ideology or private

benefits).[4] Importantly, while voters perfectly understand which policy favors their private interest, figuring out the optimal common value policy requires competence and information acquisition.

We show that there exists an asymmetric equilibrium in which the party of voters who assign a lower weight to the common value policy chooses a committed delegate while the party of voters who attach a higher weight to the common value policy chooses a trustee. When someone puts a low weight on the utility loss from not getting the common value policy right, hence placing a relatively higher weight on private benefits, (s)he is an individual who could be described as having "low moral universalism", using the terminology in Enke et al. (2020).[5] Similarly, someone who fears corruptibility of an uncommitted representative (perhaps because of a perception of the existence of a very strong elite or interest group influence) may end up putting low importance on her competence in the determination of the common value policy. Distrust on state functioning, wariness towards representative democracy, fear of globalization, may all cause lower moral universalism. In an asymmetric equilibrium it is always the party with lower moral universalism that turns populist, endorsing, e.g., nationalism, protectionism, closed border policies.

An important corollary of our main results is that moral polarization is relevant both as a precondition for an asymmetric equilibrium that generates populist pandering behavior, and as a potential consequence of the campaigning incentives in such an asymmetric (populist) equilibrium. Indeed, we show that with endogenous turnout, a populist candidate would like to push down the moral universalism of her supporters, while a non-populist candidate would like to increase the moral universalism of her voters. On the other hand, we show that ideological polarization may actually reduce the likelihood of a populist equilibrium.

Next, we show that the features that the political science literature has identified as constitutive elements of populism (beyond pandering) can be simultaneously rationalized in our asymmetric

[4]Private benefits could be anything from transfer promises to one's own group, protection of communal values, or even fixed values like being pro-choice or pro-life, all things that can be protected only conditional on the victory of one's own party and do not depend on information.

[5]In Enke et al. (2020) moral universalism refers explicitly to trust and altruism for all, so that an individual with low moral universalism can be thought of as someone who instead trusts and cares only about people in her party, broadly defined. Placing a low weight on the common good with respect to private benefits is simply a generalization of this dichotomy.

populist equilibrium: the committed delegate has clear incentives to use anti-elite rhetoric, fueling voters' fear and distrust towards competent politicians, while a trustee has an incentive to focus her campaign on her ability to choose the optimal policy ex-post. Moreover, fake news production, anti-media and anti-bureaucracy rhetoric can also easily emerge as complementary strategies for a committed delegate trying to defeat a non populist. We also show that commitment may be implicit and self-enforcing: in the absence of a commitment technology, an environment with low trust in politicians favors an incompetent candidate, who does not have more information than voters about the optimal policy. Once elected, she has to stick to the committed platform, for otherwise (s)he would prove to be biased. An additional corollary of the anti-expertise tendencies of a populist candidate is that, if and when (s)he is elected, (s)he should be expected to be in favor of firing expert bureaucrats who could undermine their campaign commitment.

The distinction between the committed delegate versus trustee model of political agency has first been introduced by Fox and Shotts (2009), in the context of optimal accountability of incumbents. [6] Kartik et al. (2017) show that in electoral competition the equilibrium degree of discretion left to a politician depends, like in our model, on the level of trust.[7] On the generalizability of our insights on commitment vs flexibility, see Amador et al. (2006). Most existing economic theories of populism focus on pandering (see, e.g., Acemoglu et al., 2013). Our paper shows that this component is complementary to all the other features of the populist strategies emphasized in the political science literature. Levy et al. (2020) depict populist policies as simplistic ones desired by unsophisticated voters, who sometimes win elections because of intense dislike of the status quo. Crutzen et al. (2020) show that if the people are divided in an informed minority and an uninformed majority, parties tend to cater more to the better informed or elite, and hence the common people develop disaffection for the traditional parties, leading to entry incentives for a populist third party. Sonin et al. (2021) present a model where an informed minority (the elite) can

[6] A similar distinction is present also in Ghosh and Tripathi (2012) and Bueno De Mesquita and Friedenberg (2011), but in their context the committed delegate is an "ideologue". In our paper ideology enters voters' utility function, but it is not a constitutive feature of the populist's strategy.

[7] On the importance of credibility, also see Van Weelden (2013).

advise the uninformed majority on candidates' competence, when one candidate is biased towards the elite and the other one is unbiased, and look at the conditions under which the uninformed majority follows the elite's advice. More than the partition of citizens along cognitive dimensions, we believe that the critical partition concerns the degree of moral universalism.[8] Prato and Wolton (2017) view populism as primarily political opportunism by incompetent politicians, but they do not link it to complexity of the environment, turnout incentives, disillusion for the common good. Gennaro et al. (2021) display robust evidence for the last two US House elections that anti-elite rhetoric is a complementary strategy employed by outsiders in close asymmetric races in districts with high economic insecurity.

A recent set of papers have emphasized the changes in social identification, making the national vs global identity become the most relevant cleavage, even more relevant than the standard left-right ideology cleavage (see, e.g., Gennaioli and Tabellini, 2019, Besley and Persson, 2019, and Shayo, 2009). Their socio-psychological analysis offers a demand-side interpretation of this phenomenon. Our paper offers a complementary supply-side interpretation: once it becomes rational for a number of parties and politicians in the new circumstances to choose a protection commitment strategy, such parties and politicians become, as a rational consequence, strategic suppliers of national and communal values protection messages. The lower trust in institutions and lower moral universalism of the audience of such political players imply that the left-right dimension drops in salience, since ideological cleavages are orthogonal to the governance issues that relate to trust.

The paper is organized as follows: in section 2 we present the simplest possible baseline model, and the consequent baseline results are described in section 3. Section 4 is the core of the paper, where we sequentially and incrementally display the most interesting implications of the model when we add endogenous dispersed information acquisition, endogenous participation, fake news production and anti-elite campaigning. Section 5 describes some implications of our model for

[8]Enke et al. (2020) show clear evidence about the fact that moral universalism is the single most important parameter characterizing the main political divide in western democracies. See also Enke (2021) for a more general analysis of the relevance of moral values for political economy.

empirical research and section 6 concludes. All proofs are relegated to the Appendix.

2 The baseline model

Let G be a principal (a party or a representative voter) who needs to choose a delegated agent (a politician) g for policy making. At the time in which the principal chooses her agent (time 0), there is uncertainty about the optimal policy q^* in a uni-dimensional policy space. The optimal policy at time 1 is going to be a realization from a distribution with differentiable and strictly positive density f on the reals (with cumulative F), with mean $\bar{q}$ and variance σ^2. The party can either ask her chosen agent to commit ex-ante to policy $\bar{q}$ at time 0,[9] or give the agent free hand to choose the policy ex-post, counting on the positive probability that the agent will have more information to choose the policy better than it could at time zero. Formally, the binary choice for principal G is between policy mandates $P^G = C$ (*committed delegate*) and $P^G = NC$ (non committed agent, or *trustee*). There are two potential agents to choose from, one with ability (type) $t^g = h$ and one with ability $t^g = l$, $1 \geq h > l \geq 0$. The ability $t^g = h, l$ will affect the probability with which the agent can observe the realization of the optimal policy at time 1. If G chooses g with $t^g = l$ there is no cost; whereas there is a cost $\epsilon > 0$ for G if she wants to select the agent with $t^g = h$.[10]

At time 1, if the agent is a committed delegate, the policy implemented is $\bar{q}$. If the agent is a trustee, she chooses the policy after a potential update on the state of the world. As already mentioned, before the trustee takes a final decision on policy, she observes the realization of the state of the world with probability t^g. Moreover, the principal can acquire information about the optimal policy: the principal chooses effort $s \in [0, 1]$, with cost $c(s)$ increasing and convex, $c'(0) = 0$, $\lim_{s \to 1} c'(s) = +\infty$. With probability s the principal learns the optimal policy q^*, and we assume that in such a case the trustee chooses q^*.[11]

[9] We assume that the policy advocated by the principal in case of commitment, $\bar{q}$, is optimal ex ante, without adding cognitive inabilities or any other ex-ante bias. The results of the paper would go through qualitatively even if $\bar{q}$ were not ex-ante optimal due to some bias.

[10] This cost represents the incentives needed to attract a high-ability agent, who has better options outside politics than a low-ability agent.

[11] A behavioral justification of this assumption would be that party keeps the politician as representative in the

The final crucial player is a lobby L, whose ideal policy q^L is drawn independently at time 1 from a distribution with differentiable and strictly positive density ϕ on the reals (with cumulative Φ), with mean $\bar{q}+\beta$, and variance τ^2 sufficiently small.[12] We also assume that $\lim_{x\to\pm\infty}\phi(x)x^2 = 0$, i.e., that the probability that the lobby's bliss point is extreme is sufficiently low.

At time 0 the principal is uncertain whether there is a lobby willing to bribe the politician. The lobby is present with probability $p \in [0,1]$. At time 1 the lobby will be active with probability $p(1-s)$. Being active means that the lobby makes a transfer offer to the uncommitted agent in exchange for choosing the lobby's ideal policy. With probability $(1-p)+ps$ the lobby will not be active, and hence will not interfere with policy making.[13]

When the lobby is active, both the lobby and the agent observe q^L, and the lobby makes a take-it-or-leave-it monetary offer $m > 0$ to the agent, and if the offer is accepted the chosen policy must be the one advocated by the lobby, q^L. When making the offer, neither the lobby nor the agent know what the optimal policy q^* for the principal is (recalling that the lobby is active only when q^* is not observed by the principal). After taking this decision, the agent observes the optimal policy with probability d^g and implements a policy.[14]

Let us now describe the principal G's payoff function:

$$U^G(q) = -\lambda^G\left(q-q^*\right)^2 + I^G - c(s) - \epsilon\mathbb{1}|_{t^g=h}.$$

future unless she is corrupt with probability 1, which would be clear if she chooses a policy different from q^* when the principal knows it.

[12] Formally, we assume that $\phi(q^L)$ is a translation by an amount β of a distribution function $\hat{\phi}(\cdot)$,$\phi_\beta(q^L) = \hat{\phi}(q^L+\beta)$, where distribution $\hat{\phi}(\cdot)$ has the same expected value as $f(\cdot)$. Hence for $\beta = 0$, q^L and q^* have the same expected value $\bar{q}$.

[13] Note that even when the lobby exists (probability p) it is assumed not to be active with probability s, because, when the optimal policy is known by the principal, we simply assume that the transfer needed to bribe the uncommitted agent to choose q^L is made unaffordable by the retrospective voting consideration mentioned in footnote 10.

[14] Assuming that the politician knows the optimal policy, before accepting the lobby's offer, leads to similar results as the ones obtained by this model, however with a more complex analysis. With this alternative assumption, it is shown that politicians with lower ability are more likely to accept the lobby's offer, further strengthening the results of this paper. In the working paper version (Morelli et al. 2021) an interested reader can find this different version of the model.

The principal's utility has four parts: a common value component that comes from the policy q; a non-contingent private value component I^G derived by having her own agent as policy-maker; the cost of acquiring information $c(s)$ and finally the cost of selection which is equal to ϵ if $t^g = h$, and equal to zero if $t^g = l$. The weight $\lambda^G > 0$ represents the intensity of the principal's preference for the common value policy. For completeness we introduce here the full utility of the principal, even though the I^G component will become relevant only with electoral competition. I^G could be interpreted as job protection or any private benefit accruing only to members of party G. I^G could alternatively be the value attached by G to having a pro-life (pro-choice) politician in power. In both interpretations I^G is not state dependent. In the first interpretation of I^G, λ^G would be close to the concept of *Moral Universalism* introduced by Enke et al. (2020).

The payoff function for the agent g is

$$U^g(q) = R - \lambda^g (q - q^*)^2 + m,$$

where R are the standard ego-rents from holding office, and m is the money that may be obtained from the lobby; λ^g denotes the expected moral universalism of the selected politician.

The payoff function for the lobby is

$$U^L(q) = -\lambda^L (q - q^L)^2 - m.$$

To summarize, the time-line is as follows:

Time 0: Principal G chooses $P^G \in \{C, NC\}$ and $t^g \in \{h, l\}$;

Time 1: Nature chooses q^* from f and q^L from ϕ. The principal decides how much information s to acquire. If g is a committed delegate then the policy is $\bar{q}$. If g is a trustee, then q^* is observed by G with probability s and q^* is implemented; the lobby becomes active with probability $p(1-s)$, and in that case makes a take-it-or-leave-it offer m to g as a quid-pro-quo for having q^L implemented, and g accepts or rejects, observes q^* with probability d^g and implements a policy.

3 Baseline equilibrium analysis

We start the analysis from the policy choices of the selected politician at the end of the game tree. If g has mandate $P^G = C$ then the committed policy choice is $\bar{q}$. If g has mandate $P^G = NC$, then with probability s the principal observes q^* and the agent, by assumption, chooses q^*; with probability $(1-s)$ the principal does not observe the optimal policy, and in this case with probability $1-p$ the lobby is not active and g chooses q^* with probability t^g and $\bar{q}$ otherwise; with probability p the lobby is active, and the following lemma provides a condition such that the lobby makes an offer that the trustee accepts.

Lemma 1. *When $\lambda^L > \lambda^g$, in equilibrium the lobby successfully bribes the trustee and implements q^L.*

We assume for the rest of the paper that $\lambda^L > \lambda^g$.

It is useful to define now the expected loss for the principal when the lobby is active:

$$\bar{L} := \int_{q^L \in \mathbb{R}} \int_{q^* \in \mathbb{R}} \left(q^* - q^L\right)^2 dF\left(q^*\right) d\Phi\left(q^L\right).$$

$\bar{L}$ depends on the two distributions of the lobby and the principal's optimal policies.

We are now ready to establish the baseline result:

Proposition 1. *1. For $\epsilon > 0$ but small enough, if G chooses $P^G = C$, it also chooses $t^g = l$; in contrast, if $P^G = NC$ then $t^g = h$.*

2. The principal is more likely to choose $P^G = C$ the higher β and p are, and the lower is h.

3. A higher λ^G reduces the net cost of information acquisition, hence makes $P^G = NC$ more likely.

4. The more information acquisition takes place, the smaller are the effects of p, β, h on the possibility of observing a committed delegate in equilibrium.

This baseline result establishes that pandering to what we expect to be right today (a form of populism according to multiple definitions) is indeed more appealing for the principal (party or voter) when corruption is high and costly (p, β are high), when even the best available politicians are not that great (h is low). A lower λ^G implies a lower benefit of having an ex-post optimal common-value policy compared to the private cost of information acquisition. If λ^G is high, it is possible that for the principal it is worth investing on monitoring ex post and hence endorse a non committed politician of type h.

It can be shown that the expected loss $\bar{L}$ increases with β, the parameter shifting the distribution of q^L away from the expected ex-ante optimal policy of citizens, $\bar{q}$.[15] To see what effects an increase of p or β has on the amount of information acquired in equilibrium by the principal, consider a set of parameters such that in equilibrium $P^G = NC$: an increase in either p or β increases the information acquisition, but a further increase of these parameters may induce a switch to $P^G = C$ and the information acquired in equilibrium drops to zero. Thus,

Corollary 1. *An increase in corruption threats (higher p and β) has a non monotonic effect on information acquisition.*

The results contained in Proposition 1 are consistent with the results of the retrospective accountability model by Fox and Schotts (2009). Also in their model, a politician may act as a trustee (choosing a policy based on her competence) or as a delegate (choosing a policy based on the representative voter's beliefs). Hence, their accountability logic and our delegation logic converge in terms of conditions for different types of equilibrium political agency. The baseline model offers also an important and novel observation: the principal chooses different types of delegation (commitment vs non commitment) also depending on the weight that she puts on the (state-contingent) policy dimension in her utility function, capturing the relative benefit of monitoring.

The baseline model offers a clear picture of the consequences of the decline of trust towards politicians. Voters demand a committed delegate and are willing to trade the possibility of getting

[15] This result is provided as part of the proof of Proposition 1.

an ex-post optimal policy for the safe option of clear and simple commitment that does not need any costly monitoring. In our baseline delegation model the principal fears the potential influence of interest groups, lobbies or elites, and this sets up the stage for our theory of populism, which will take its complete form when we introduce elections. The next section shows, among other things, that the ex ante pandering commitment component of populism is conducive to moral polarization, anti-elite rhetoric and fake news production incentives, anti-media and anti-bureaucracy campaigning. All these are components of populist politics often emphasized in the political science literature, and they are all and simultaneously rationalizable by our model. We show that only in an electoral competition between a committed delegate and a trustee we will observe the rise of a fully-fledged populist package. Such asymmetric equilibrium may emerge thanks to the difference in the moral universalism of the two parties.

4 Electoral competition, mobilization, and the rationalization of the full-fledged populism strategy

What happens when we augment the model by allowing for primaries and general elections rather than a simple principal-agent choice, and how do endogenous turnout and mobilization strategies modify the baseline result? What may explain the existence of an equilibrium in which one party chooses commitment but the other does not? And what role do the fundamentals play in such cases for the determination of the probability of winning for the populist party? Is the ex ante pandering commitment component of populism conducive to the other components of populist politics often emphasized in the political science literature? In this section we offer an answer to all these questions.

4.1 Main results

The virtue of a baseline model with a simple principal-agent delegation logic is simplicity, and it allowed us to already establish baseline results on the impact of corruption, expected competence of politicians, and intensity of common-value policy preferences on the possibility of a strategic populist commitment choice. We now extend the model to allow for two-party electoral competition and endogenous turnout, since all other aspects of populism appear when considering these realistic elements of the political game. Adding endogenous participation, ideology or private benefits I^G matter for turnout. Formally, we introduce the following modifications of the model:[16]

- **Voting:** There are two parties $G = A, B$, who need to choose their respective candidates for policy making, $g = a, b$, with the same tradeoffs as in the baseline model. Each party G is composed by χ^G citizens, who select candidate g in a primary. We denote by V the set of citizens who are members either of party A or party B. Without loss of generality, we assume $0 < \lambda^A \leq \lambda^B$. We keep the assumption that the general interest policy q that affects all citizens' welfare has common value and therefore the optimal policy q^* is the same for both parties.[17] To reduce the heterogeneity between the two parties, we assume that all candidates have the same λ which we denote by λ^w. Hence independently of the party the candidate belongs to, she faces the same trade-offs when dealing with the lobby.[18] Citizens in each party have heterogeneous cost of voting. Given common interest, in each party only members with negative cost of voting turn out in the primaries; but in the general elections, where a runs against b, also agents with positive cost of voting can turn out, depending on the assumptions made on rational voting decisions. To avoid standard problems of free-riding in collective action, we assume rule utilitarian voting (see Feddersen and Sandroni,

[16] For conciseness, the full description of the model is presented in the appendix.

[17] In section 4.3.1 we discuss which results go through if the voters in party A and B have different optimal policies, both ex ante and conditional on information ex post.

[18] If voters can choose the λ of their candidate, they clearly have the incentive to select a representative with the maximum possible λ^w, to reduce corruptibility. If the pools of candidates have the same features in the two parties, then both parties select a representative with the same (the largest) λ^w. We show in the extensions section that, when candidates have the same λs of their party, our results are even stronger.

2006a, and Coate and Conlin, 2004) – each voter maximizes their party's aggregate utility, and each group's optimal voting rule determines the threshold cost of voting below which all members are mobilized. We also assume that before the general elections there is a realization of a random shock γ^G to the cost (or benefit) of voting for citizens of each party G, only observed by party G's voters. This way the probability of winning remains interior like in probabilistic voting models, and we can then trace the effects of the fundamental parameters on the probability of winning of committed or uncommitted candidates.

- **Information acquisition**: Having replaced the single principal of the baseline model with a voting selection, we have to be consistent also in terms of information acquisition, assuming that if the elected policy-maker is not committed then the endogenous information acquisition efforts are by the citizens once again. Consistently with the assumption that citizens have heterogeneous cost of voting, we assume that they are heterogeneous in their cost of information acquisition. Let $s_G^j \in [0,1]$ denote the amount of information acquired by a generic citizen j of party G. The cost of information is a continuous increasing and convex function $k^j c\left(s_G^j\right)$ such that $c(0) = c'(0) = 0$, $\lim_{s\to 1} c'(s_G^j) = +\infty$, and k^j is a realization from a uniform distribution between $\underline{k} \geq 0$ and $\overline{k} > \underline{k}$. k^j realizes after elections and before citizens get informed. Again, party utility maximization determines the group-optimal information, denoted by s_G^{j*}, acquired by each citizen in line with the framework of Feddersen and Sandroni (2006b). Let us define $S = \int_{j\in V} s_G^{j*} dj$ as the aggregate information acquired by all citizens. Similarly to the baseline, S is the probability with which the optimal policy becomes common knowledge, which forces a trustee to implement it, no matter her personal interest.

With these two modifications, all the substantive results of Proposition 1 are essentially unchanged. Importantly, in our model a difference in the level of λ^G between parties may lead to an asymmetric equilibrium in which one candidate proposes a committed policy and the other candidate an uncommitted one.

Lemma 2. *Given $\lambda^A < \lambda^B$ it is possible to find parameter values under which $P^A = C$ and $P^B = NC$, while it is impossible to find an equilibrium under which $P^A = NC, P^B = C$.*

The lower is λ^G, the higher is the opportunity cost of being informed for a voter; and therefore there is a threshold under which members of party G prefer to elect a committed delegate, because they do not need to acquire information to monitor the policy-maker. As a consequence, when an equilibrium in which the candidates propose different policies exist, then the candidate who offers commitment is the representative of party A, given $\lambda^A < \lambda^B$. The existence of an asymmetric equilibrium in which candidate a is a committed delegate while b is a trustee is guaranteed under two sufficient conditions. The first is that the population sizes of the two groups are not too different. The second is that the expected welfare loss when the lobby is influencing the policy-maker is intermediate. In fact, if the bias were too high, then even those citizens who have a high λ^G would prefer a commitment policy. In the opposite case, if the welfare loss due to lobbies' influence is small, no citizen would be too worried about the lobbies' influence and all would prefer to take the risk of having a policy that is contingent on the realization of the state of the world, even if lobbies may distort policy-maker's choices.

Ideology or private benefits did not affect the results in the baseline principal-agent model, but in the full-blown model with electoral competition the I^G parameters will start to matter. I^G is a component of a voter's utility function that is activated only when the voter's preferred party wins, and the effect of I^G on citizens' welfare does not depend on the realization of the state of nature. Therefore it does not require any information acquisition by voters.[19] Leaving the technical details of the above modifications to the appendix, we summarize the substantive findings in the next proposition:

Proposition 2. *1. When the equilibrium is symmetric, displaying either both candidates with commitment or both without commitment, the probability of winning of the candidate of party G is increasing in I^G, χ^G and decreasing in I^{-G}, χ^{-G}.*

[19] To reiterate, pro-life vs pro-choice policy preferences are hardly dependent on realizations of any state of the world or scientific discovery; and similarly for private benefits interpretation: cash is cash.

2. *When the equilibrium is asymmetric with $P^A = C$ and $P^B = NC$, the probability of winning for candidate a (candidate b) is still increasing in I^A (I^B) and the size of the party, but it is also increasing (decreasing) in p, β and decreasing (increasing) in h, λ^A, λ^B.*

Proposition 2 offers relevant insights. First, in symmetric equilibria when both candidates are trustees or both are committed delegate, the probability of winning of each party G is increasing in I^G and therefore our model predicts that the ideological component of the policy platform plays a major role in the electoral campaign.[20] Second, in the asymmetric equilibrium the probability of winning depends also on the voters' expectation about the negative effect of corruption on politicians (p, β), on the ability of the trustee and, importantly, also on the weight that each party assigns to the common value policy. Electoral campaigns will be polarizing in multiple dimensions. A committed candidate benefits from convincing the electorate of the risk of politicians being captured by elites, of the difficulty of monitoring them, and of the complexity in understanding the state of the world, while a trustee benefits from convincing the electorate of her expertise. Finally, it is important to notice how the weights λ^A and λ^B affect the probability of winning of a candidate: for candidate a, the committed delegate, the lower is λ^A the higher is the probability of winning, because the higher λ^A the higher is the disutility of an ex-post inefficient policy; while for candidate b, the trustee candidate, the opposite holds: the higher λ^B, the higher is the probability of winning, because the higher λ^B the lower is the opportunity cost of monitoring. It follows that in an asymmetric equilibrium we predict a sort of *moral polarization*, with a candidate adopting a divisive electoral campaign that further reduces the moral universalism of her supporters and the other candidate standing for a more cohesive and inclusive campaign. Thus the model provides a novel link between moral universalism, introduced by Enke et al. (2020), and populism.

While *moral* polarization is definitely important, as precondition as well as potential consequence of the differential campaign incentives, *ideological* polarization can actually reduce the likelihood of a populist equilibrium, as the following proposition highlights.

[20] When both candidates are trustees their ability affects the probability of winning but our model predicts that in equilibrium they both have high ability.

Proposition 3. *An increase in I^A does* **not** *necessarily imply higher propensity towards commitment. In fact,*

1. *consider a set of parameters such that an asymmetric equilibrium exists. If λ^A decreases or λ^B increases, the asymmetric equilibrium still exists. When the difference between the two weights λ^A and λ^B is large, an increase in I^A* **shrinks** *the set of parameters such that the asymmetric equilibrium exists.*

2. *There exist pairs (λ^B, λ^A) such that, when $I^A = 0$ the equilibrium is asymmetric, while for I^A sufficiently large the equilibrium is symmetric and both candidates propose no commitment.*

A higher ideological component for party A does not necessarily imply a higher propensity towards populist commitment: If I^A increases (higher ideological polarization or higher value given to some type of party-level protection or communal value) party A's voters may prefer to increase the probability that their own party candidate wins the election even if this implies not having their preferred type of policy-maker (committed delegate). Intuitively, consider $I^B = 0$, $I^A = 0$, λ^A small and λ^B large, such that the equilibrium is asymmetric. Suppose now that I^A increases: party B's members still only care about which common value policy is implemented, while party A's members now care more about which politician is elected. It may then follow that in equilibrium party A chooses a NC representative, thus proposing the same policy that B supporters would like to have implemented, because the subsequent reduction in party B members' turnout is larger than the reduction in party A members' turnout.

What matters the most for the existence of a populist equilibrium is moral polarization, and, in particular, in an asymmetric equilibrium with only one populist it is the party with lower moral universalism that chooses a populist commitment.

4.2 The rationalization of the other components of the full-fledged populism strategy

The above analysis has shown that in an asymmetric equilibrium the effect of the parameters of the model on the probability of winning of each candidate goes in opposite directions, and therefore candidates will differentiate their electoral campaign. We now incorporate in our analysis two relevant aspects of political campaigns that have raised attention in recent years. Populism in western democracies has always been characterized by a rampant use of anti-elite rhetoric by populist candidates, as reflected in the definition itself of populism that prevails in political science. Second, with the new technologies we observe an increased diffusion of fake news and distorted information, often fomented by the same populist parties. We show that these phenomena are instrumental and fully rationalizable, since the use of anti-elite rhetoric, fake news production and mass media delegitimization, boost the probability of electoral success of a committed candidate in an asymmetric equilibrium

4.2.1 Fake news and mass media delegitimization

Suppose that each candidate $g \in \{a, b\}$ at the general election can exert a costly effort (or campaign budget) $n_g \in \mathbb{R}_+$ to circulate fake news, delegitimize and discredit national media outlets, and create doubts. Such efforts increase the cost of information acquisition by voters: $(k^j + n_a + n_b)c(s_G^j)$.[21] Let $\Psi(n)$ be the cost of circulating fake news, assumed the same for both candidates, with $\Psi(0) = 0$, strictly increasing and convex.

Proposition 4. *Suppose $P^A = C$ and $P^B = NC$. For R sufficiently large, candidate a exerts positive effort and candidate b zero effort in producing fake news.*

The intuition of this result is straightforward. In case candidates compete in the general

[21] While generally news are produced by media and not candidates, we abstract from modeling the market for news, to keep this extension close to the main model. Moreover, candidates typically talk about news in their speeches and political campaigns, increasing the salience of a subset of the whole menu of news offered by media. By choosing to talk overwhelmingly about fake news or to constantly attack the main news sources, a candidate makes it more difficult for citizens to acquire correct information.

election with asymmetric strategies, candidate a benefits from an increase in the cost of information acquisition because it decreases party B's voters turnout and increases party A's voters turnout. Candidate b benefits from an increase in the cost of information because a reduction in the amount of information acquired by voters increases the probability that she gets the money from the lobby; however the negative effect induced by a lower probability of winning the election overcomes the positive effect of an increase in the expected bribe, conditional on holding office.[22] It follows that in case fake news do not affect the probability of winning the election, as it is in case both candidates offer an uncommitted policy, candidates will exert a positive effort to lower citizens' monitoring.

Notice that if we assumed that fake news by a candidate increase the cost of information acquisition of the candidate's party more than for the other party (due perhaps to echo-chambers), our results would continue to hold, and it would follow that citizens with lower λ are also more likely to be affected, in line with the empirical results on "information inequality" by Angelucci and Prat (2021).

4.2.2 Anti-elite rhetoric

Populist rhetoric is characterized by a strong anti-elitism. In this subsection we suggest how anti-elite rhetoric aiming to affect voters' beliefs about the likelihood that lobbies and elites could distort politicians' actions is a natural complement of commitment policies. We assume here that the candidate who wins the primary of party G chooses how to allocate a unitary endowment of time between campaigning on competence and anti-elite rhetoric, as in Gennaro et al. (2021). Formally, the choice $\rho_g \in [0,1]$, $g = a, b$, that we may interpret as the time spent for anti-elite or competence campaigning, has the following impact on citizens' beliefs of competence and possibility of elite capture:

$$p(\rho_a, \rho_b) = p + \eta(\rho_a) + \eta(\rho_b), \text{ with } \eta(0) = 0, \eta' > 0,\ 2\eta(1) < (1-p);$$

[22] In our model the lobby has full bargaining power and makes a take or leave offer to the politician. If the politician had some bargaining power, a trustee would have higher incentives to increase the cost of information. Still, the general observation that in an asymmetric equilibrium the committed delegated exerts more effort than the trustee in producing fake news would hold. See the proof of Proposition 4 for more details.

$\hat{t}^g(\rho_g) = t^g + \theta(\rho_g)$, with $\theta(1) = 0,\ \theta' < 0, \theta(0) < (1 - t^g)$.

In words, the general election campaign effort by politician g of party G can be used to modify the prior of that party's citizens on either her competence or on the risk of elite capture.

Proposition 5. *A candidate $g \in \{a, b\}$ who proposes a commitment policy chooses $\rho_g = 1$, while a candidate g who proposes a non-commitment policy chooses $\rho_g = 0$.*

This proposition does not need a formal proof: it is evident that is a weakly dominant strategy for any committed delegate candidate to allocate all her time campaigning to increase voters' perception on the probability that the lobby is active. Only in case both candidates are committed delegates a candidate is indifferent on which topic to devote her campaign. For a trustee it is a dominant strategy to devote her campaign to increase voters' beliefs about her competence.[23]

4.2.3 Consequences for the bureaucracy

Another feature of populists' behavior that could easily be seen as complementary (or even consequent) to the choice of a committed delegate form of political agency, is the strong anti-buraucracy and anti-expert rhetoric (and action once in office) by populist politicians. Sasso and Morelli (2021) derive formally the implications of the committed delegate model of populism for bureaucratic turnover, bureaucrats' incentives and selection, and for some government quality effects. Bellodi et al. (2021) find strong and robust evidence that indeed once a populist enters office there is a significant increase in bureaucratic turnover, a significant reduction in the quality of the bureaucracy, and a significant reduction in government performance. The intuition for the mechanism and how it relates to our model is straightforward: a bureaucracy of experts can block or alter the (ex-post inefficient) policies that populists have committed to during the campaign, and populists do not want that.

[23] Notice that, even if we allow negative campaigning about the competence of the opponent, in an asymmetric equilibrium the committed delegate would still prefer to allocate her time on an anti-elite rhetoric as long as the mobilisation effect of this rhetoric is stronger than other effects. In a related paper, Gennaro et al. (2021) study the sensitivity of the anti-elite component of the populist campaign strategy in US elections to economic insecurity variation and district characteristics. The parameter λ in our model could be affected by economic and political conditions of a specific context.

4.3 Extensions

We now discuss a number of directions in which our results can be extended. First, we discuss the realistic possibility that the two parties have heterogeneous preferences over the policy that affects all citizens. Second, we consider the case in which voters of different parties care about different policies; and finally we discuss the relation between competence and endogenous commitment.

4.3.1 Heterogeneous Preferences

In the paper we have assumed that q is a common value policy. How do the results change when the policy that needs information acquisition affects all citizens but they may disagree on what is the optimal policy to implement? In this section we extend the baseline model to capture this case, ignoring first the effects of this heterogeneity on voters' turnout. We discuss at the end of the section what hurdles does endogenous turnout add to this extension with heterogeneous preferences.

Assume that the two agents selected by the two parties win with an exogenous probability 1/2. Assume that the two principals acquire information to learn the state of the world q^* but have heterogeneous preferences over the optimal policy to implement: $q^* - \delta$ is the optimal policy for principal A and $q^* + \delta$ is the optimal policy for principal B. The parameter $\delta \geq 0$ captures the degree of preference heterogeneity among principals. Consistently, a committed agent a would choose $\bar{q}_A$ and a committed agent b would choose $\bar{q}_B = \bar{q} + \delta > \bar{q}_A = \bar{q} - \delta$.[24] Let us briefly solve the game by backward induction, to underline the similarities and differences with respect to the main model.

The analysis of the subgame in which the lobby is active is equal to the one in the main model, with the differences that the trustee with probability $(1-p) + pS$ implements her optimal policy and $S = \frac{1}{2}(s_A + s_B)$: if the policy-maker is a, she implements $q_A^* := q^* - \delta$ while b implements

[24] For example, consider the case in which the single most important or salient national policy is openness towards immigrants or openness to globalization. In these two typical examples for the literature on populism, δ captures the cultural or socioeconomic differences between a left-wing and a right-wing voter even under full information. Yet, the optimal policy on globalization or immigration also depends on the realization of many economic and political factors.

$q_B^* := q^* + \delta$. Let $\bar{L}(G)$ denote the expected loss for principal $G \in \{A, B\}$ when the lobby successfully bribes the politician:

$$\bar{L}(G) := \int_{q^L \in \mathbb{R}} \int_{q^* \in \mathbb{R}} \left(q_G^* - q^L\right)^2 dF\left(q^*\right) d\Phi\left(q^L\right).$$

If the policy-maker w is a trustee, the expected utility of principal G is:

$$-\lambda^G (1-S)\left[p\bar{L}(G) + (1-p)(1-t^w)\left(\sigma^2 + \mathbb{1}_{|w\neq g} 4\delta^2\right)\right] - \lambda^G S \mathbb{1}_{|w\neq g} 4\delta^2 - c(s_G),$$

where $\mathbb{1}_{|w\neq g}$ takes value 1 if and only if the elected politician w does not belong to party G. The intensity of principal G's information acquisition s_G^* solves, therefore, the following first order condition:

$$\lambda^G\left[p\bar{L}(G) + (1-p)(1-t^w)\sigma^2 - [1-(1-p)(1-t^w)]\mathbb{1}_{|w\neq g} 4\delta^2\right] = 2c'\left(s_G^*\right).$$

With respect to the baseline model, when the agent selected by the other principal is the policy-maker, there is a lower incentive to get informed. Indeed, a larger 2δ, the distance between the preferences of the two principals, reduces the marginal benefit of information acquisition, because disciplining the policy-maker implies getting her preferred policy instead of the bias induced by lobbying (when the lobby is active). Only if the expected bias of the lobby is sufficiently large, $p\bar{L}(G) > [1-(1-p)(1-t^w)]\mathbb{1}_{|w\neq g} 4\delta^2 - (1-p)(1-t^w)\sigma^2$, a principal gets informed when the opposing agent is elected.

Finally, when selecting commitment or no commitment, each principal chooses a committed delegate when the utility from a commitment policy is higher than the utility from no-commitment:

$$-\lambda^G \sigma^2 > -\lambda^G (1-S)\left[p\bar{L}(G) + (1-p)(1-t^w)\sigma^2\right] - c(s_G).$$

The trade-off faced by each principal is the same as the one in the baseline model. Thus Proposition

1 still holds, namely a higher λ^G decreases the probability that principal G chooses a committed delegate.

As anticipated at the beginning of this extension section, some results of the model are altered when heterogeneous preferences are allowed and the probability of winning is endogenous due to endogenous turnout. Indeed, by introducing two important differences in the two groups' preferences (the λs and their optimal policies), the model naturally leads to a larger set of possible equilibrium outcomes. Lemma 2 states that, if there is an equilibrium where one principal chooses a committed delegate and her opponent chooses a trustee, then the latter is the one with the larger λ. Depending on parameter values, this Lemma could fail to apply to this extension. Indeed, by introducing heterogeneity in the principal's preferences, we also added an asymmetry in their expected disutility of the lobby's bias. One of the principal suffers a larger expected disutility from the lobby, because her optimal policy in expectation will be farther from the lobby's bliss point, with respect to her opponent. If such principal is the one with a larger λ, and the bias is sufficiently large, she could choose commitment to avoid the bias, while her opponent, with a lower λ, could choose no-commitment because her bias from lobbying is much smaller. Left-wing populism is actually characterized by a strong anti-elite sentiments, especially against financial elites who have opposing interests to those of the poor. This extension would then suggest that, even in the absence of a lower λ, left-wing parties could resort to commitment strategies because their optimal policy is very far from the one of special interests (see also discussion in section 5) and the probability of elite capture p is high.

4.3.2 Multidimensional policies

Consider now an extension in which there are two national policies to be decided, e.g., security q_a and welfare policy q_b. Suppose for simplicity that for party A $\lambda^A_{q_a} > 0$ but $\lambda^A_{q_b} = 0$, and vice versa for party B, so that even if both policies enter additively in the utility function each voter considers only the policy dimension most important for her. In this case once again all our results qualitatively continue to hold, with the little caveat that citizens of one group do not get

informed after elections if the opponent is elected, because they do not value the opponent's policy dimension. Moreover, our results clarify the role of moral universalism for the choice of a populist strategy within each party: for party A a lower $\lambda_{q_a}^A$ implies that committing to "America first" or disengagement becomes preferable to more flexible international policies; similarly, for a voter of party B a lower $\lambda_{q_b}^B$ implies that a commitment to a simple citizenship income can be preferred to more elaborate welfare policies. In fact, security at the more global level or welfare considerations for others have low relevance when moral universalism is low.

4.3.3 Heterogeneous λ of candidates

In the main model we assumed that all candidates attach the same weight to policy in their utility function. What happens if candidates have different λs? Let us assume, for simplicity, that candidate g has the same λ^G of her party members. Candidate a therefore, having a lower moral universalism than candidate b, is more willing to sell out to the lobby. Hence, voters of party A are more prone to select a candidate who proposes commitment, with respect to voters of party B, because the expected loss from lobbying is larger. Thus, adding this heterogeneity to our model further strengthens our results: citizens with lower value attached to policy are more likely to vote for a committed delegate.

4.3.4 Incompetence as a commitment device

Some final remarks concern the possibility of endogenous and partial commitment. First, observe that as long as the commitment is credible, voters do not acquire information about the state of the world, so they never push the policy-maker to change the ex-ante optimal policy. Still, a policy-maker may learn (thanks, for instance, to a competent bureaucracy) with positive probability what is the ex-post optimal policy and realize that it is very far from $\bar{q}$. If commitment is not full, a committed politician may propose to change ex-post the policy and voters face the dilemma whether this bid depends on her superior knowledge or lobby pressure. If voters can select in the primary a candidate g of any ability $t^g \in [0, 1]$, choosing a candidate with $t^g = 0$ is a commitment

device. A candidate with zero ability will never learn the optimal ex-post policy and voters should assign probability one that the policy-maker is derailed by the lobby if the commitment is not maintained. Di Tella and Rotemberg (2018) offer an explanation for the observation that voters sometimes seem to prefer incompetent politicians. In their model, voters are disappointment averse and more competent politicians are more likely to betray them. Our result offers an alternative explanation of why voters sometimes seem to prefer incompetent policy-makers. Consistently, Sasso and Morelli (2021) show theoretically and Bellodi et al. (2021) confirm empirically that populist leaders can lead to the firing of expert bureaucrats and decreased bureaucratic performance. Together with the current paper, they show that the consequences of populism include poor policy implementation and decreased voter welfare.

Finally, in our model candidates can be either committed or uncommitted. However, we might consider the possibility that candidates could propose a partial commitment platform, like in Kartik et al. (2017). We conjecture that if candidates could commit to choose a policy within an interval, the size of the interval would shrink under the same conditions that make commitment more desirable in our paper. Most likely, the interval would probably not be symmetric around $\bar{q}$, but rather larger on the opposite side of the lobby bias.

5 Empirical predictions

The model presented in this work has rich empirical implications. We list them now and discuss the available empirical evidence to support them.

1. *Higher distrust towards elites pushes towards populism.* There is a large literature showing that the demand for populism is positively impacted by a drop in trust in politicians, political parties, political institutions, and ruling classes in general. For European citizens, this can take the form of not trusting European institutions and a general suspicion that the political class will not be pursuing the interests of common people (Dustman et al. 2017, Algan et al., 2017, Foster and Frieden, 2017, and Guiso et al., 2021).

2. *Populist politicians have lower competence.* The model predicts that, on average, we should expect politicians that support commitment policies to have lower competence. A Gallup poll in 2019 showed that Trump scored relatively low among voters in terms of policy competence.[25] Moreover, in Italy, looking at municipal elections from 1998 to 2019, simple linear probability model estimates conditioning on the population size of the municipality confirm that being populist is associated with a decrease in the probability of having a university degree by 6 percentage points (see Bellodi et al. 2021).

3. *Citizens with lower moral universalism are more likely to vote for populist politicians.* Enke et al. (2020) show that right-wing citizens are generally characterized by a lower moral universalism, a feature that is persistent throughout the Western world. Broad confirmation of our empirical prediction comes from noticing that populism has been recently mostly a right-wing phenomenon. Notice that, in the cases in which populism had left-wing characteristics, leftist voters were the ones mostly worried about elite capture by lobbies. This fear, consistently with our model, pushed them to demand commitment policies even in the absence of lower moral universalism (see also extension 4.3.1 on heterogenous preferences). Consider the cases of Podemos in Spain, or Chavez's movement in Venezuela. Podemos rallied strongly against "la Casta financiera", deemed responsible for the economic crisis, while Chavez campaigned against the foreign powers controlling Venezuelan natural resources.

4. *Populist supporters put less effort in getting informative news about policy.* This empirical prediction closely relates to the quality of information that individuals get, because the theory models information acquisition as a costly action, done to receive a correct signal about the optimal policy. If a consumer of news claims to be informed but has incorrect political knowledge, because she did not put effort to check the source, this would classify in our model as not getting informed. Van Kessel et al. (2021), using survey data from nine European democracies, show that right-wing populist supporters claim to be politically

[25] See https://news.gallup.com/poll/260495/trump-seen-marginally-decisive-leader-not-honest.aspx

informed, whereas they in fact are less likely to possess the correct information. Similarly, Allcott et al. (2017) show that US Republican voters are less able, with respect to Democrats, to distinguish correct and incorrect political statements. Hameleers et al. (2017), using data on the Netherlands, show that populist voters have a tabloidized and entertainment-based media diet. Bovet et al. (2019) show that the activity of Trump supporters largely contributes to the diffusion of the top fake news on Twitter. Similarly Guess et al. (2020) show that, during the 2016 US electoral campaign, Trump supporters were more likely to visit websites publishing fake news.

5. *Populist politicians circulate more fake news* . There is large anecdotal evidence and increasing literature (see, e.g., Ross and Rivers 2018) showing that Donald Trump circulated fake news through his Twitter account and during his public speeches, especially during the 2016 and 2020 US electoral campaigns. Similarly, Jair Bolsonaro contributed to the diffusion of fake news both during the 2018 Brasilian Presidential campaign, and during the COVID crisis, see Ricard and Medeiros (2020).

6. *The turnout of populist (non populist) supporters is increasing (decreasing) in the distrust towards the elites.* The work by Guiso et al. (2021) shows, with European data, that turnout decreases when citizens trust less the ruling class, but conditional on voting, they choose populist candidates. These results support our empirical prediction, if the decline in turnout of non-populist supporters is larger than the increase in turnout of populist voters.

7. *Populists specialize on anti-elite rhetoric during electoral campaigns.* Gennaro et al. (2021) confirm empirically that Donald Trump resorted to anti-elite rhetoric during the 2016 electoral campaign, more than his opponent, Hillary Clinton, and more evidently in speaches made in districts with higher voters' disillusion due to economic insecurity. Moreover, using data on the 2018 and 2020 congressional elections, the authors show that politicians who run as outsiders in districts with high economic insecurity tend to specialize in anti-elite rhetoric.

8. *The average citizens' (quality of) information acquisition is non-monotonic with respect to the distrust towards elites.* This empirical prediction closely relates and follows logically from predictions 1 and 4. Thus, as far as we know, the empirical works discussed in relation to these two predictions provide potential indirect evidence to this claim. Testing in depth this prediction is left for future research.

9. *Ideological polarization is not necessarily a cause of increased populism.* We showed that, when the ideological component in the utilities of citizens is large, candidates do not propose commitment. This implies that candidates will also not implement all the complementary strategies analyzed in the paper (circulation of fake news, anti-elite rhetoric) and should focus instead on the ideological part of their proposals. Instead we expect to observe that, when the ideological component is low and there is large moral polarization (given by very different values of moral universalism), the whole bundle of populist features realizes. This empirical prediction thus speaks to the intense debate on the ideological content of populism (see, e.g., Stanley, 2008, and Freeden, 2016), suggesting a nuanced relationship, whose empirical test is an interesting venue for future research.

On top of the empirical implications listed above, we wish to reiterate the implications for bureaucratic efficiency and government performance: our model suggests that populists, when in office, should want to maximize the probability of keeping their promises, and in fact the word "promise" is prominent in the populism dictionary used in text analysis. Hence there should be a consistent tendency by populist incumbents to fire expert bureaucrats who want instead to match the state of the world. This prediction is fully worked out in Sasso and Morelli (2021) and empirically confirmed in Bellodi et al. (2021), where they also show that the high turnover and deterioration of the bureaucracy, which are effects of populism, have significant consequences in terms of quality of governance and government performance.

6 Concluding remarks

This paper shows, starting from a simple principal-agent model and then adding electoral competition among two parties, that a supply of fully-fledged populism only emerges in an asymmetric equilibrium when a delegated candidate faces a trustee: only in this case the commitment to an ex ante popular policy (pandering) is complemented by the other elements that characterize populism: anti-elite rhetoric with no emphasis on competence; production of fake news to make information acquisition and monitoring of politicians harder (lower accountability expected at election time); and equilibrium endogenous elite (lobby) activity consistent with the popular fear (possibly exacerbated by rhetoric). Moreover, we establish that this populist equilibrium is more likely to arise when there is moral polarization, while ideological polarization is not necessarily related to the populism phenomenon, since we show that the space of parameters such that a populist equilibrium exists may actually shrink when ideological polarization increases.

Adam Smith in the last chapter of the wealth of nations already noticed that stagnation and inequality, combined with excessive division of labor and fear of replacement by machines, may generate a collapse of the moral sentiments necessary for the good functioning of commercial society, namely ingenuity, frugality and prudence. But beside this decay of individual moral values that were functional to capitalism, as emphasized in Censolo and Morelli (2021), it is easy to imagine that also moral universalism may go down. Fear of globalization, distrust for politicians, increased fears of a zero-sum game, bring citizens to be less generous and more self-absorbed and therefore less prone to care about the utility of those distant from them, in space and intergenerationally, and hence to put a lower weight on common-value policies. The fact that the party with lower moral universalism is the most likely to turn populist in a two-party competition is consistent with the observation that populism in western democracies emerged mostly on the right of the political spectrum: an average right-wing voter (an average Republican in the US) typically cares less about (and trusts less) the public good provision by the state.

One important direction for future research concerns dynamics. While Levy, Razin and Young

(2020) study the dynamics of intensity of preferences of voters of two given groups, the sophisticated and the unsophisticated, the dynamics of learning that can be envisioned as a follow-up of our paper has as main components the interplay between learning on new policy dimensions and on the real power of the elite(s) in biasing policy-making. Under some conditions a long phase of non-populist equilibrium can be followed by short or long phases of populist equilibria or committed delegation on both sides, and with different exit patterns depending on the fundamentals. Learning dynamics both on the variance of the optimal policy and on the strength of the elite(s) may vary with the characteristics of the equilibrium supply (non populist or populist, symmetric or asymmetric) in non trivial ways. We conjecture that if the elite is monolitic or one-sided, then a populist equilibrium could last much longer than when there are multiple competing elites, because with expected interest groups' pressure on two different sides the expected equilibrium distortion may be lower. An additional direction of extension of our model could be that when a populist is in office we can learn about the costs of populism or on unexpectedly important new policy dimension. For example the cost of having incompetent bureaucrats could be updated upwards during a pandemic, and on newly relevant dimensions like logistics. Another important conjecture on the potential follow-up implications of our model is that if the lobby is not expected to be extreme, then if the policy-maker is more constrained (as it is the case for the policy-maker in a country with no independent monetary or fiscal policy) the value of uncommitted policy-makers goes down, and populist equilibria can surface.

References

Acemoglu, Daron, Georgy Egorov and Konstantin Sonin. 2013. "A Political Theory of Populism," *Quarterly Journal of Economics*, 128 (2): 771-805.

Acemoglu, Daron and Pasqual Restrepo. 2017. "Robots and Jobs: Evidence from US Labor Markets", *NBER Working Paper*, No. 23285.

Algan, Yann, Sergei Guriev, Elias Papaioannou, and Evgenia Passari. 2017. "The European

Trust Crisis and the Rise of Populism." *Brookings Papers on Economic Activity*, 309-82.

Allcott, Hunt, and Matthew Gentzkow. 2017. "Social Media and Fake News in the 2016 Election." *Journal of Economic Perspectives*, 31 (2): 211-36.

Amador, Manuel, Ivan Werning, and George Marios Angeletos. 2006. "Commitment vs. Flexibility." *Econometrica*, 74 (2): 365-96.

Ananyev, Maxim, and Sergei Guriev. 2018. "Effect of Income on Trust: Evidence from the 2009 Economic Crisis in Russia." *Economic Journal*, 129 (619): 1082-1118.

Angelucci, Charles and Andrea Prat. 2020 "Measuring voters' knowledge of political news". *CEPR*, DP 15222.

Autor, David , David Dorn, Gordon Hanson, and Kaveh Majlesi. 2020. "Importing Political Polarization? The Electoral Consequences of Rising Trade Exposure." *American Economic Review*, 110 (10): 3139-83.

Bellodi, Luca, Massimo Morelli and Matia Vannoni. 2021. "The Cost of Populism for Bureaucratic Efficiency and Government Performance". Available at http://www.massimomorelli.eu/

Besley, Timothy, and Torsten Persson. 2019. "The Rise of Identity Politics". Unpublished manuscript.

Bovet, Alexandre, and Hernan A. Makse. 2019. "Influence of fake news in Twitter during the 2016 US presidential election." *Nature Communications*, 10, 7.

Coate, Stephen, and Michael Conlin. 2004. "A group rule-utilitarian approach to voter turnout: Theory and evidence," *American Economic Review*, 94 (5): 1476-1504.

Censolo, Roberto and Massimo Morelli. 2021. "Adam Smith and the Moral Decline of the Middle Class.". Mimeo, Bocconi University.

Colantone, Italo, and Piero Stanig. 2018. "The Trade Origins of Economic Nationalism: Import Competition and Voting Behavior in Western Europe." *American Journal of Political Science*, 62 (4): 936-53.

Crutzen, Benoit, Dana Sisak and Otto Swank. 2020. "Left Behind Voters, Anti-Elitism and Popular Will," *Tinbergen Institute Discussion Paper*, 055.

Di Tella, Rafael, and Julio J. Rotemberg. 2018. "Populism and the return of the "Paranoid Style: Some evidence and a simple model of demand for incompetence as insurance against elite betrayal." *Journal of Comparative Economics*, 46 (4): 988-1005.

Dustmann, Christian, Barry Eichengreen, Sebastian Otten, André Sapir, Guido Tabellini, and Gylfi Zoega. 2017. "Europe's Trust Deficit: Causes and Remedies." CEPR Press.

Enke, Benjamin. 2021. "Moral Values and Voting." *Journal of Political Economy*. Forthcoming.

Enke, Benjamin, Ricardo Rodríguez-Padilla, and Florian Zimmermann. 2020. "Moral Universalism and the Structure of Ideology." *NBER Working Paper*, No. 27511, Working Paper Series, July.

Feddersen, Timothy, and Alvaro Sandroni. 2006a. "A theory of participation in elections." *American Economic Review*, 96-4: 1271-82.

Feddersen, Timothy, and Alvaro Sandroni. 2006b. "Ethical voters and costly information acquisition." *Quarterly Journal of Political Science*, 1 (3): 287-311.

Foster, Chase, and Jeffry Frieden. 2017. "Crisis of Trust: Socio-Economic Determinants of Europeans' Confidence in Government." *European Union Politics*, 18 (4): 511-35.

Fox, Justin, and Kenneth W. Shotts. 2009. "Delegates or trustees? A theory of political accountability." *The Journal of Politics*, 71 (4): 1225-37.

Freeden, Michael. 2016. "After the Brexit referendum: Revisiting populism as an ideology." *Journal of Political Ideologies*, 22: 1-11.

Gennaioli, Nicola and Guido Tabellini. 2020. "Identity, Beliefs, and Political Conflict." Unpublished manuscript.

Gennaro, Gloria, Giampaolo Lecce and Massimo Morelli. 2021. "Mobilization and the Strategy of Populism: Theory and Evidence from the United States." *CEPR discussion paper series*, DP 15686.

Guess, Andrew M., Brendan Nyhan. and Jason Reifler. 2020. "Exposure to untrustworthy websites in the 2016 US election." *Nature Human Behaviour*, 4: 472–480.

Guiso, Luigi, Helios Herrera, Massimo Morelli, and Tommaso Sonno. 2019. "Global Crises and

Populism: the Role of Eurozone Institutions." *Economic Policy*, 34 (97): 95-139.

Guiso, Luigi, Massimo Morelli, Tommaso Sonno and Helios Herrera. 2021. "The Financial Drivers of Populism in Europe," available at http://www.massimomorelli.eu/

Guriev, Sergei and Elias Papaioannou. 2021. "The political economy of populism." *Journal of Economic Literature*, forthcoming.

Hameleers, Michael, Linda Bos and Claes H. de Vreese. 2017. "The Appeal of Media Populism: The Media Preferences of Citizens with Populist Attitudes." *Mass Communication and Society* 20 (4): 481-504

Kartik, Navin, Richard van Weelden and Stephane Wolton. 2017. "Electoral Ambiguity and Political Representation". *American Journal of Political Science*, 61 (4): 958-70.

Laitin, David D. 2018. "Populism: the Muslim Connection." Keynote Address at the CESifo Conference *Globalization and Populism - Past and Present*, Venice, Italy, June 4-5.

Levy, Gilat, Ronny Razin and Alwin Young. 2020. "Misspecified Politics and the Recurrence of Populism", mimeo, LSE.

Morelli, Massimo, Antonio Nicoló and Paolo Roberti. 2021. "A Commitment Theory of Populism". *CEPR discussion paper series*, DP 16051.

Mudde, Cas. 2004. "The Populist Zeitgeist". *Government and opposition*, 39 (4), 541-563.

Norris, Pippa, and Ronald Inglehart. 2019. *Cultural Backlash: Trump, Brexit, and Authoritarian Populism*. Cambridge, *Cambridge University Press*.

Prato, Carlo, and Stephane Wolton. 2017. "Rational Ignorance, Populism, and Reform." *European Journal of Political Economy*, 55: 119-135.

Ricard, Julie, and Juliano Medeiros. 2020. "Using misinformation as a political weapon: COVID-19 and Bolsonaro in Brazil." *Harvard Kennedy School Misinformation Review*, 1(3).

Rodrik, Dani. 2018. "Populism and the Economics of Globalization." *Journal of International Business Policy*, 1 (1): 12-33.

Ross, Andrew S., and Damian J. Rivers. 2018. "Discursive Deflection: Accusation of 'Fake News' and the Spread of Mis- and Disinformation in the Tweets of President Trump." *Social Media*

+ Society.

Sasso, Gregory, and Massimo Morelli . 2021. "Bureaucrats under Populism." Forthcoming on the *Journal of Public Economics.*

Shayo, Moses. 2009. "A Model of Social Identity with an Application to Political Economy: Nation, Class, and Redistribution". *American Political Science Review* , 103 (2): 147-174.

Sonin, Konstantin, Ran Eilat, and Marina Agranov. 2021. "A Political Model of Trust." *CEPR discussion paper series*, DP 14672.

Stanley, Ben. 2008. "The thin ideology of populism." *Journal of political ideologies*, 13 (1): 95-110.

Van Kessel, Stijn, Javier Sajuria, and Steven M. Van Hauwaert. 2021. "Informed, uninformed or misinformed? A cross-national analysis of populist party supporters across European democracies." *West European Politics*, 44 (3): 585-610.

Van Weelden, Richard. 2013. "Candidates, Credibility, and Re-election Incentives." *Review of Economic Studies*, 80 (4): 1622-51.

Appendix

Proof of Lemma 1: Consider a lobby with preferred policy q^L that makes an offer m to the agent g. The agent receives the following utility, if she accepts the offer of the lobby: $m - \lambda^g \int_{q^* \in \mathbb{R}} \left(q^L - q^*\right)^2 dF\left(q^*\right)$, which includes m, the money from the lobby, and the expected utility loss from implementing the lobby's preferred policy. The agent receives utility $-\left(1 - t^g\right) \lambda^g \sigma^2$, if she refuses the lobby's offer. Indeed, with probability t^g she discovers the optimal policy and her utility from policy is 0. With the complementary probability, she does not discover the optimal policy, hence she implements $\bar{q}$ and her utility from policy, net of λ^g, is the expected distance from $\bar{q}$, i.e., the variance σ^2. The agent accepts the lobby's offer if

$$m - \lambda^g \int_{q^* \in \mathbb{R}} \left(q^L - q^*\right)^2 dF\left(q^*\right) \geq -\left(1 - t^g\right) \lambda^g \sigma^2.$$

Notice that

$$\int_{q^* \in \mathbb{R}} \left(q^L - q^*\right)^2 dF\left(q^*\right) = \int_{q^* \in \mathbb{R}} \left[\left(q^L - \bar{q}\right) - \left(q^* - \bar{q}\right)\right]^2 dF\left(q^*\right)$$
$$= \int_{q^* \in \mathbb{R}} \left[\left(q^L - \bar{q}\right)^2 + \left(q^* - \bar{q}\right)^2 - 2\left(q^L - \bar{q}\right)\left(q^* - \bar{q}\right)\right] dF\left(q^*\right) = \left(q^L - \bar{q}\right)^2 + \sigma^2. \tag{1}$$

By substituting the latter expression in the agent's incentive compatibility constraint, we obtain $m \geq \bar{m} := \lambda^g \left[t^g \sigma^2 + \left(q^L - \bar{q}\right)^2\right]$. The lobby chooses to bribe the agent if

$$-\bar{m} > -\lambda^L t^g \left[\left(q^L - \bar{q}\right)^2 + \sigma^2\right] - \lambda^L \left(1 - t^g\right) \left(q^L - \bar{q}\right)^2,$$

where we made use again of simplification (1). By substituting the expression for $\bar{m}$ in the lobby's incentive compatibility constraint, we obtain

$$-\lambda^g \left(q^L - \bar{q}\right)^2 - \lambda^g t^g \sigma^2 > -\lambda^L \left(q^L - \bar{q}\right)^2 - \lambda^L t^g \sigma^2.$$

From the latter inequality, we obtain that the lobby bribes the agent if $\lambda^L > \lambda^g$.

□

Proof of Proposition 1: Let us prove each point sequentially.

1. Consider the utility of the principal, when G selects C: $-\lambda^G\sigma^2 - \epsilon\mathbb{1}|_{t^g=h}$. If G chooses C, her utility from policy does not depend on the competence of her agent. Given that selecting a competent agent is costly, G selects an agent with low competence $t^g = l$. Suppose that G chooses NC. Let $s^* \in (0,1)$ denote the optimal amount of information acquired by G when choosing NC. We derive s^* and its properties in point 3 below. G's utility from NC is therefore $-\lambda^G(1-s^*)\left[(1-p)(1-t^g)\sigma^2 + p\bar{L}\right] - c\left(s^*\right) - \epsilon\mathbb{1}|_{t^g=h}$, which is increasing in t^g, because $(1-p)(1-t^g)\sigma^2$ decreases in t^g, and ϵ is small. We disregard the effect of t^g on s^*, because of the envelope theorem. Therefore, if G chooses NC, she selects a competent agent, for a positive but not too large ϵ.

2. Consider the incentive compatibility constraint, when G selects C, as opposed to selecting NC:

$$-\lambda^G\sigma^2 > -\lambda^G(1-s^*)\left[(1-p)(1-h)\sigma^2 + p\bar{L}\right] - c\left(s^*\right) - \epsilon. \tag{2}$$

   Analyzing inequality (2), when deriving the utility from choosing NC with respect to β and p, we disregard their effect on s^*, because of the envelope theorem. See point 3 below on the properties of s^*. The utility from choosing NC decreases with β, because $\frac{\partial\bar{L}}{\partial\beta} > 0$, as shown by Lemma 4, stated and proven at the end of this proof. The comparative statics of utility from choosing NC with respect to p have the same sign of $-(\bar{L} - (1-h)\sigma^2)$. The following lemma shows that the last expression is negative.

   Lemma 3. *If $\beta = 0$, $(1-h)\sigma^2 < \bar{L} = \sigma^2 + \tau^2$.*

Proof of Lemma 3. First of all notice that the following holds:

$$\bar{L} > (1-h)\int_{q^*\in\mathbb{R}}\int_{q^L\in\mathbb{R}}\left(q^* - q^L\right)^2 d\Phi\left(q^L\right) dF\left(q^*\right).$$

Moreover, for $\beta = 0$, the following holds:

$$\int_{q^*\in\mathbb{R}}\int_{q^L\in\mathbb{R}}\left(q^* - q^L\right)^2 d\Phi\left(q^L\right) dF\left(q^*\right) =$$
$$\int_{q^*\in\mathbb{R}}\int_{q^L\in\mathbb{R}}\left(q^* - \bar{q} - q^L + \bar{q}\right)^2 d\Phi\left(q^L\right) dF\left(q^*\right) =$$
$$\int_{q^*\in\mathbb{R}}\int_{q^L\in\mathbb{R}}[(q^* - \bar{q})^2 + \left(q^L - \bar{q}\right)^2 - 2\left(q^* - \bar{q}\right)\left(q^L - \bar{q}\right)]d\Phi\left(q^L\right) dF\left(q^*\right) = \sigma^2 + \tau^2.$$

Hence the lemma is proven. Finally $\bar{L}$ increases with β (see Lemma 4), hence $\bar{L} > (1-h)\sigma^2$ for any value of $\beta \geq 0$. The rest of the comparative statics follow from the analysis of the utility from choosing NC.

3. Conditional on choosing NC, the principal chooses the amount of information to be acquired maximizing

$$-\lambda^G(1-s)\left[(1-p)(1-h)\sigma^2 + p\bar{L}\right] - c\left(s\right) - \epsilon.$$

When s^* is interior, the FOC is:

$$\lambda^G\left[(1-p)(1-h)\sigma^2 + p\bar{L}\right] = c^{'}\left(s^*\right).$$

s^* is interior because of the INADA conditions on $c\left(\cdot\right) : c'(0) = 0, \lim_{s\to 1} c'(s) = +\infty$. The second order conditions and uniqueness of s^* are satisfied because $c\left(\cdot\right)$ is convex. Notice that λ^G increases the amount of information acquisition. In equilibrium, G selects C if

$$-\lambda^G\sigma^2 + \lambda^G\left[(1-s^*)\left((1-p)(1-h)\sigma^2 + p\bar{L}\right)\right] + c\left(s^*\right) + \epsilon \geq 0. \tag{3}$$

Note that, for $\lambda^G = 0$, G prefers commitment, because in this case G does not acquire any information when choosing NC, hence the inequality becomes $\epsilon \geq 0$. Moreover the derivative of the lhs of inequality (3) with respect to λ^G is equal to $-\sigma^2 + (1 - s^*)\left[(1-p)(1-h)\sigma^2 + p\bar{L}\right]$, where the derivatives of s^* with respect to λ^G are omitted. Indeed, by the envelope theorem, the change in s^* induced by an increase of λ^G does not affect the derivative of the utility from no commitment. The second derivative of the difference in utilities with respect to λ^G is negative, because $\frac{\partial s^*}{\partial \lambda^G} > 0$, hence the difference in utilities is concave in λ^G and it is ϵ at $\lambda^G = 0$. This proves the claim, because either the difference in utilities is positive for every $\lambda^G > 0$, or there exists a threshold for λ^G such that it is positive only below such threshold.

4. We now study the relationship between information acquisition, parameters p, β and h, and the utility from NC. In order to do so, we assume that the cost of information acquisition is parameterized by $y > 0$, so that the cost function is $\frac{1}{y}c(\cdot)$. In this way, a larger y increases information acquisition. The relationship between information acquisition, parameter p and the utility from NC is captured by the following cross-derivative:

$$\frac{\partial^2}{\partial y \partial p}\left\{-\lambda^G (1 - s^*)\left[(1-p)(1-h)\sigma^2 + p\bar{L}\right] - c(s^*)\right\} = \frac{\partial s^*}{\partial y}\lambda^G\left[\bar{L} - (1-h)\sigma^2\right],$$

which is positive (see Lemma 3), hence an increase in information acquisition reduces the negative effect of p on the utility from NC. The same holds for parameter β. Similarly an increase in information acquisition reduces the positive effect of h on the utility from NC.

Lemma 4. *$\bar{L}$ is increasing in β.*

Proof of Lemma 4: We compute the derivatives of the expected loss $\bar{L}$ with respect to β. We take

advantage of an integration by parts:

$$\int_{q^L\in\mathbb{R}}\int_{q^*\in\mathbb{R}}\left(q^*-q^L\right)^2 dF\left(q^*\right)\frac{\partial}{\partial\beta}\phi_\beta\left(q^L\right)dq^L=\int_{q^*\in\mathbb{R}}\left(q^*-q^L\right)^2\hat{\phi}(q^L+\beta)|_{-\infty}^{+\infty}dF\left(q^*\right)+$$
$$2\int_{q^L\in\mathbb{R}}\int_{q^*\in\mathbb{R}}\left(q^*-q^L+\beta\right)dF\left(q^*\right)\hat{\phi}\left(q^L\right)dq^L=2\int_{q^L\in\mathbb{R}}\int_{q^*\in\mathbb{R}}q^*dF\left(q^*\right)\hat{\phi}\left(q^L\right)dq^L-$$
$$2\int_{q^L\in\mathbb{R}}\int_{q^*\in\mathbb{R}}q^L dF\left(q^*\right)\hat{\phi}\left(q^L\right)dq^L+2\beta\int_{q^L\in\mathbb{R}}\int_{q^*\in\mathbb{R}}dF\left(q^*\right)\hat{\phi}\left(q^L\right)dq^L.$$

Recall that $\lim_{q^L\to\pm\infty}\left(q^*-q^L\right)^2\hat{\phi}(q^L+\beta)=0$. Moreover, the expected values of distributions f and $\hat{\phi}$ are equal. Hence the derivative of $\bar{L}$ with respect to β is equal to 2β, which is positive if $\beta>0$.

□

Proof of Lemma 2: The proof of the existence of an asymmetric equilibrium is provided below as a part of the proof of Proposition 2.

Proof of Proposition 2: Let us first define formally the extended model with voters. We will then briefly go through the subgame perfect equilibrium analysis by backward induction.

Citizens-Voters

There is a unitary mass of citizens V. Each citizen $j\in V$ belongs to one of two parties, A and B, with $A\cup B=V$. Party $G\in\{A,B\}$ has mass χ^G. For each party we assume that there exist at least two potential candidates, one with low ability l and one with high ability h. Thus candidate i is associated to ability (type) level t^i, with $t^i\in\{l,h\}$, $0\leq l\leq h\leq 1$. There is also a lobby that intervenes after elections, as in the baseline model. The first node where citizens play is the simultaneous primaries node. Each citizen j of party G observes t^i for each candidate i and the policy $P^i\in\{C,NC\}$ chosen by each candidate i . Denoting by E^G the set of candidates in party G primary, each citizen of party G decides:

1) whether to turnout ($\psi_j^{pr}=1$) or not ($\psi_j^{pr}=0$) in the primary, given her cost of voting $z^j\in[-\underline{z},\bar{z}]$ with $-\underline{z}<0<\bar{z}$, drawn from a uniform distribution;

2) conditional on turning out, which candidate $i \in E^G$ to vote for, given the pair t^i, P^i of each candidate. Selecting a high quality candidate implies a cost $\epsilon > 0$ for each citizen in party G, as in the baseline model.

Let us denote by g the candidate of party G selected in the primary. Second, in the general election node all citizens simultaneously decide whether to turn out or not ($\psi_j^{ge} \in \{0,1\}$), with cost of voting for voter j in G equal to $z^j + \gamma^G$, where γ^G is a party specific voting cost shock materializing for the general election (see below). Conditional on turning out, any citizen votes for her party candidate, like in all mobilization models.[26]

The third and last stage where voters move is after elections: upon observing who won, all citizens simultaneously decide how much information to acquire, given the specific individual realization of information acquisition cost, and given the relative benefit of information acquisition in terms of the possibility to discover what is the optimal policy that the policy-maker should pursue. Let us denote by s_G^j the amount of information acquired by citizen j of party G. The cost of information is a continuous increasing and convex function $k^j c\left(s_G^j\right)$ such that $c(0) = c'(0) = 0$, $\lim_{s\to 1} c'(s_G^j) = +\infty$, and k^j is a realization from a uniform distribution between $\underline{k} \geq 0$ and $\overline{k} > \underline{k}$. k^j realizes just after elections and before citizens get informed.

As in the baseline model, individual gathering of information has a public good component: it affects the probability that the optimal policy is discovered by some citizens, who can provide public and credible evidence about which policy should be chosen. When this occurs, the elected politician is forced to implement the optimal policy. In other words, with probability $S = \int_{j\in V} s_G^j dj$ the optimal policy is discovered by the society, and reelection concerns (unmodeled here) push an elected uncommitted politician to implement it. With probability $1 - S$ citizens do not know what is the optimal policy, and in this case the elected politician w observes the optimal policy with probability t^w, but it remains private information whether she observed it or not.

[26] We can easily show that in our model we do not need to constrain voting decisions to remain within the party by assumption, since before elections and conditional on voting, the expected utility of electing a given candidate is the same for each citizen in a party.

Citizen j of party G derives the following utility from policy q and realized optimal policy q^*:

$$U^{j,G}(w,q,q^*) = -\lambda^G (q-q^*)^2 + I^G \mathbb{1}_{w=g} - k^j c\left(s_G^j\right) - (z^j+\gamma^G)\mathbb{1}_{\psi_j^{ge}=1} - z^j \mathbb{1}_{\psi_j^{pr}=1} - \epsilon \mathbb{1}|_{t^g=h},$$

where function $\mathbb{1}_{x=1}$ takes value 1 if variable x is equal to 1. The weight $\lambda^G > 0$ captures the relative importance of the common policy w.r.t. ideology I^G for a citizen of party G, where we assume $\lambda^A \leq \lambda^B$.

Nature

Beside choosing the initial distribution of individual costs of voting, from a uniform distribution with support $[\underline{z}, \bar{z}]$, with $\underline{z} < 0 < \bar{z}$, Nature also chooses a party-specific voting cost shock $\gamma^G \in [-1/2, 1/2]$ common to all members of $G = A, B$ that modifies the distribution of voting costs right before the general elections. Each voter observes only their own party shock, not the one of the other party. γ^G is for simplicity distributed uniformly. Moreover, after elections Nature chooses a realization of information acquisition costs, the optimal policy $q^* \in \mathbb{R}$, and whether the optimal policy is revealed to all players (with probability S). Finally, the action set and payoffs for the elected politician and the lobby are equal to those of the baseline model.

Equilibrium concept with ethical voting

Now we formally introduce the equilibrium concept.

Definition 1. *A citizen $j \in G$ strategy is a triple $\left(\psi_j^{pr}(P^i, t^i, z^j)_{i\in E^G}, \psi_j^{ge}(P^a, P^b, z^j+\gamma^G), s_G^j(P^w, k^j)_{w\in\{a,b\}}\right)$ specifying whether citizen j votes at the primary and general election and how much information she acquires, conditional on the elected politician. A candidate $i \in E^G$ strategy is a policy proposal $\{P^i\}$ at the primary and a policy $q|_{P^w=NC}$ conditional on winning the election and being a trustee. The policy chosen by the trustee implies also the acceptance or refusal of the lobby's offer. The lobby's strategy is an offer function $m(q^L)$. The described strategy profile constitutes an equilibrium if and only if*

1. *each citizen j chooses a strategy that maximizes the aggregate expected utility $\int_{j\in G} U^{j,G}(w,q,q^*)$*

of her party supporters, as a best response to the strategies played by citizens of the other party;

2. *politicians and lobby follow the standard expected utility maximization criterion;*

3. *all players formulate consistent beliefs on the strategies of all other players.*

The form of behavioral rule implied at the voting stage by party-utility maximization is a threshold threshold voting cost below which the party members are required to turn out (in line with Coate and Conlin 2004 and Feddersen and Sandroni 2006a, 2006b). In particular, for the general election the voting rule for party G is a threshold z^G below which party members are supposed to turn out, and z^G is a function of the party shock to the citizen voting cost, and the platforms of the two candidates.

Equilibrium analysis

Before solving the model by backward induction, notice that electoral competition in the primaries ensures that the primaries always select the candidate that proposes the policy that maximizes the expected utility of citizens of a party. Moreover, when such policy is NC (C), the high (low) ability candidate will be nominated proposing NC (C). The proof is trivial hence omitted.

Let us solve the game by backward induction. The equilibrium analysis of the lobby is equal to the one in the previous model, hence omitted. Citizen $j \in G$ maximizes the following utility, when choosing the amount of information s_G^j:

$$-\lambda^G \chi^G \left(1 - \int_{j \in V} s_G^j dj\right) \left[(1-p)(1-h)\sigma^2 + p\bar{L}\right] - \int_{j \in G} k^j c\left(s_G^j\right) dj - \epsilon.$$

Similarly to the baseline model with information acquisition, the equation that determines the amount of information acquisition by citizen $j \in G$ is

$$\frac{\lambda^G \chi^G \left(p\bar{L} + (1-p)(1-h)\sigma^2\right)}{k^j} = c'\left(s_G^{j*}\right),$$

where we already substitute $t^w = h$, because, as in the baseline model, if citizens elect in the primaries a candidate who proposes NC, such candidate must have ability h given the equilibrium analysis of the primary. The comparative statics are the same as in the baseline model. Additionally, the amount of information acquisition increases with the population size χ^G, because each citizen in G is more motivated to get informed when its consequences fall on more of her party mates. We analyze now citizens' decision to turnout. With abuse of notation with respect to the baseline model, let us define the citizen's expected utility of having an elected politician of ability h who proposes a no-commitment policy:

$$v^G(NC) = -\lambda^G (1 - S^*) \left(p\bar{L} + (1-p)(1-h)\sigma^2\right) - \mathbb{E}_{k^j, j \in G} \left[k^j c\left(s_G^{j*}\right)\right] - \epsilon,$$

where $\mathbb{E}_{k^j, j \in G} \left[k^j c\left(s_G^{j*}\right)\right]$ denotes the expected cost of information of a citizen of party G and $S^* := \int_{j \in G, G \in \{A,B\}} s_G^{j*} dj$; citizens, in fact, do not know their cost of acquiring information before election.

The expected utility of having an elected politician who proposes an ex-ante optimal policy commitment is :

$$v^G(C) = -\lambda^G \sigma^2.$$

Notice that, while $v^G(\cdot)$ is the individual expected utility from having a given politician in office, when making choices of turnout citizens will consider the aggregate expected utility of their group, thus multiplying $v^G(\cdot)$ by the population size χ^G. Let us denote by $\Delta\mathbb{E}(P^G, P^{-G})$ the difference between the expected utility for voters of party G of electing a candidate who proposes the optimal policy for their own party and the expected utility of having in office the candidate who proposes the policy P^{-G} that is optimal for the other party. If the two candidates propose the same policy, $P^G = P^{-G}$, and they have the same ability, then $\Delta\mathbb{E}(P^G, P^{-G})$ is zero. The following lemma illustrates citizens' decision to turnout.

Lemma 5. *For each party G, there is a threshold $z^G = \frac{\left(\chi^G\right)^2}{\chi^{-G}}\left(\underline{z}+\bar{z}\right)\left(I^G + \Delta\mathbb{E}(P^G, P^{-G})\right) - \gamma^G$, such that only citizens in party G with a cost of voting below z^G vote. The probability of winning of the candidate of party G depends positively on $I^G + \Delta\mathbb{E}(P^G, P^{-G})$ and negatively on $I^{-G} + \Delta\mathbb{E}(P^{-G}, P^G)$.*

Proof of Lemma 5: Let us first define party A's candidate probability of victory, which is the probability that party A's voters are more than party B's voters: $\mathbb{P}\left(\chi^A\left(\frac{z^A+\underline{z}}{\underline{z}+\bar{z}}\right) > \chi^B\left(\frac{z^B+\underline{z}}{\underline{z}+\bar{z}}\right)\right)$. Notice that we can simplify the denominator $\underline{z}+\bar{z}$; z^A maximizes the following expected utility for citizens in party A:

$$\chi^A\mathbb{P}\left(\chi^A\left(z^A+\underline{z}\right) > \chi^B\left(z^B+\underline{z}\right)\right)\left(I^A + v^A\left(P^A\right)\right) +$$
$$\chi^A\left[1-\mathbb{P}\left(\chi^A\left(z^A+\underline{z}\right) > \chi^B\left(z^B+\underline{z}\right)\right)\right]v^A\left(P^B\right) - \int_{-\underline{z}}^{z^A}\frac{z^j+\gamma^A}{\underline{z}+\bar{z}}dz^j,$$

where P^A and P^B are the policies proposed respectively by candidates a and b (selected in the primaries); z^B maximizes the corresponding aggregate utility for a member of party B. It follows that the best response function of party A satisfies:

$$\chi^A\frac{\partial}{\partial z^A}\mathbb{P}\left(z^B < \frac{\chi^A}{\chi^B}z^A - \frac{\underline{z}(\chi^B-\chi^A)}{\chi^B}\right)\left(I^A + \Delta\mathbb{E}(P^A, P^B)\right) - \frac{1}{\underline{z}+\bar{z}}\left(z^A+\gamma^A\right) = 0.$$

Let us conjecture that $\frac{\partial}{\partial z^A}\mathbb{P}\left(\chi^A z^A > \chi^B z^B + \underline{z}(\chi^B-\chi^A)\right) = k^A$, a constant with respect to z^A. Similarly, we conjecture $\frac{\partial}{\partial z^B}\mathbb{P}\left(\chi^B z^B > \chi^A z^A + \underline{z}(\chi^A-\chi^B)\right) = k^B$. Hence, it follows that $z^A = k^A\chi^A\left(\underline{z}+\bar{z}\right)\left(I^A + \Delta\mathbb{E}(P^A, P^B)\right) - \gamma^A$. Analogously, from the best response function of party B, we get $z^B = k^B\chi^B\left(\underline{z}+\bar{z}\right)\left(I^B + \Delta\mathbb{E}(P^B, P^A)\right) - \gamma^B$. Given the best response function of party

B, the probability of winning of party A is

$$\mathbb{P}\left(z^B < \frac{\chi^A}{\chi^B} z^A - \frac{\underline{z}(\chi^B - \chi^A)}{\chi^B}\right) =$$
$$\mathbb{P}\left(k^B \chi^B \left(\underline{z} + \bar{z}\right)\left(I^B + \Delta\mathbb{E}(P^B, P^A)\right) - \gamma^B < \frac{\chi^A}{\chi^B} z^A - \frac{\underline{z}(\chi^B - \chi^A)}{\chi^B}\right) =$$
$$\mathbb{P}\left(k^B \chi^B \left(\underline{z} + \bar{z}\right)\left(I^B + \Delta\mathbb{E}(P^B, P^A)\right) - \frac{\chi^A}{\chi^B} z^A + \frac{\underline{z}(\chi^B - \chi^A)}{\chi^B} < \gamma^B\right) =$$
$$\frac{1}{2} - k^B \chi^B \left(\underline{z} + \bar{z}\right)\left(I^B + \Delta\mathbb{E}(P^B, P^A)\right) + \frac{\chi^A}{\chi^B} z^A - \frac{\underline{z}(\chi^B - \chi^A)}{\chi^B},$$

and therefore $k^A = \frac{\chi^A}{\chi^B}$, confirming our initial conjecture. Similarly it can be proven that $k^B = \frac{\chi^B}{\chi^A}$, which implies that the best responses are

$$z^A = \frac{\left(\chi^A\right)^2}{\chi^B}\left(\underline{z} + \bar{z}\right)\left(I^A + \Delta\mathbb{E}(P^A, P^B)\right) - \gamma^A,$$
$$z^B = \frac{\left(\chi^B\right)^2}{\chi^A}\left(\underline{z} + \bar{z}\right)\left(I^B + \Delta\mathbb{E}(P^B, P^A)\right) - \gamma^B.$$

Finally the probability of winning of the candidate of party A is

$$\frac{1}{2} - \frac{\chi^B}{\chi^A}\chi^B \left(\underline{z} + \bar{z}\right)\left(I^B + \Delta\mathbb{E}(P^B, P^A)\right) + \tag{4}$$
$$\frac{\chi^A}{\chi^B}\left(\frac{\left(\chi^A\right)^2}{\chi^B}\left(\underline{z} + \bar{z}\right)\left(I^A + \Delta\mathbb{E}(P^A, P^B)\right) - \gamma^A\right) - \frac{\underline{z}(\chi^B - \chi^A)}{\chi^B}.$$

The lemma is proven. Lemma 5 proves Proposition 2, with an additional comment on the relationship between the probabilities of winning and λ^A, λ^B. When party A supports C and party B supports NC, the difference in utilities of party B (hence its probability of winning) clearly increases with λ^B, because $v^B(C) < v^B(NC)$ and the difference in utilities $v^G(C) - v^G(NC)$ is concave in λ^G. Moreover the derivative of $v^A(C) - v^A(NC)$ with respect to λ^A and of $v^B(C) - v^B(NC)$ with respect to λ^B are both equal to $-\sigma^2 + (1 - S^*)\left[(1-p)(1-h)\sigma^2 + p\bar{L}\right]$, hence if $v^B(NC) - v^B(C)$ increases with λ^B, $v^A(C) - v^A(NC)$ decreases with λ^A.

□

Proof of Lemma 2: First of all, the impossibility of an equilibrium where $P^A = NC$ and $P^B = C$ follows the same reasoning of the proof of point (3) in Proposition 1, hence it is omitted. Let us show now that there are parameter values such that $P^A = C$ and $P^B = NC$. We need the following lemma to prove this point:

Lemma 6. *When $|\chi^A - \chi^B|$ is sufficiently low: $v^A(NC)/\lambda^A < v^B(NC)/\lambda^B$.*

Proof of Lemma 6. When $\chi^A = \chi^B$, the only heterogeneity between the two parties is given by a different λ^G. Notice that the following holds:

$$\frac{\partial}{\partial \lambda^G}\left\{\left[-\lambda^G \chi^G (1-S^*)\left(p\bar{L} + (1-p)(1-h)\sigma^2\right) - \mathbb{E}_{k^j, j\in G}\left[k^j c\left(s_G^{j*}\right)\right] - \epsilon\right]/\lambda^G\right\} = \frac{\mathbb{E}_{k^j, j\in G}\left[k^j c\left(s_G^{j*}\right)\right] - \epsilon}{\left(\lambda^G\right)^2} > 0,$$

where the derivatives of s_G^{j*} with respect to λ^G are omitted, because of the envelope theorem. Hence $v^A(NC)/\lambda^A < v^B(NC)/\lambda^B$ which, by continuity, must be true also for $|\chi^A - \chi^B|$ sufficiently low. The lemma is proven.

We now proceed to prove Lemma 2. Let us fix $I^A = I^B = 0$. Thus party supporters do not receive extra utility by electing their own candidate. Therefore, there is no strategic incentive of electing a party's own candidate per se: when the two candidates propose the same policy, the election of either candidate gives citizens in group G the same utility. Hence, the objective of party G supporters is to select the candidate who proposes a policy that, if elected, maximizes their expected utility.

Consider the conditions for an equilibrium where A selects C and B selects NC:

$$-\lambda^A \sigma^2 > -\lambda^A (1-S^*)\left[(1-p)(1-h)\sigma^2 + p\bar{L}\right] - \mathbb{E}_{k^j, j\in A}\left[k^j c\left(s_A^{j*}\right)\right] - \epsilon, \tag{5}$$

$$-\lambda^B \sigma^2 < -\lambda^B (1-S^*)\left[(1-p)(1-h)\sigma^2 + p\bar{L}\right] - \mathbb{E}_{k^j, j\in B}\left[k^j c\left(s_B^{j*}\right)\right] - \epsilon. \tag{6}$$

Let us focus on the conditions for B. When $\beta = 0$, and τ (the variance of the lobby's bliss point) are sufficiently low, the inequality is satisfied. Indeed, substituting $\bar{L} = \sigma^2 + \tau^2$ (see Lemma 3) and considering $s_G^{j*} = 0$ for all citizens $j \in G$ and for both parties G (which are all substitutions inducing a lower bound on the utility from NC), inequality (6) can be rewritten as follows: $\tau^2 < \frac{1-p}{p} h\sigma^2$, therefore for τ sufficiently low the inequality is satisfied.

When $\beta \to \infty$, inequality (6) is not satisfied, because $\frac{\partial \bar{L}}{\partial \beta} > 0$, with the derivative not converging to 0 for high levels of β. Therefore, there exists a threshold $\bar{\beta}$ such that inequality (6) is satisfied with equality. Let us consider $\beta = \bar{\beta}$. Inequalities (5) and (6) can be rewritten as follows:

$$-\sigma^2 > -(1-S^*)\left[(1-p)(1-h)\sigma^2 + p\bar{L}\right] - \frac{\mathbb{E}_{k^j, j\in A}\left[k^j c\left(s_A^{j*}\right)\right] - \epsilon}{\lambda_A},$$
$$-\sigma^2 < -(1-S^*)\left[(1-p)(1-h)\sigma^2 + p\bar{L}\right] - \frac{\mathbb{E}_{k^j, j\in B}\left[k^j c\left(s_B^{j*}\right)\right] - \epsilon}{\lambda_B}.$$

Notice that, by Lemma 6, the following holds:

$$-\sigma^2 = -(1-S^*)\left[(1-p)(1-h)\sigma^2 + p\bar{L}\right] - \frac{\mathbb{E}_{k^j, j\in B}\left[k^j c\left(s_B^{j*}\right)\right] - \epsilon}{\lambda_B} >$$
$$-(1-S^*)\left[(1-p)(1-h)\sigma^2 + p\bar{L}\right] - \frac{\mathbb{E}_{k^j, j\in A}\left[k^j c\left(s_A^{j*}\right)\right] - \epsilon}{\lambda_A}.$$

Hence, inequality (5) is satisfied. For β marginally larger than $\bar{\beta}$, inequality (6) is satisfied with strict inequality and inequality (5) holds, if λ^A is not marginally lower than λ^B.

□

Proof of Proposition 3

First of all, notice that the existence of an asymmetric equilibrium where party A supports a candidate who runs on C and party B a candidate who runs on NC, is guaranteed by Lemma 2, with $I^A = I^B = 0$. Next we investigate what happens when ideologies are positive, in particular when I^A grows large. We make use of two lemmas that prove point (iii) of Proposition 2.

Lemma 7. *Suppose λ^A is sufficiently low or λ^B is sufficiently large. An increase in party A's ideology I^A shrinks the set of parameters in which the equilibrium is asymmetric.*

Proof of Lemma 7: Let us define for simplicity $\mathbb{P}^G(C, NC)$ the probability that the candidate of party G wins when the candidate of party A proposes commmitment, and the candidate of party B proposes no commitment. Similarly we can define the probabilities referring to all the other possible policy proposals by candidates. Moreover we define the (subgame) equilibrium threshold for turnout of party G as a function of the policy proposals of the two candidates in the general election: $z^G(C, NC)$. Consider the conditions of party A, for an equilibrium where party A selects a candidate that proposes commitment and party B selects a candidate that proposes no commitment. In the following analysis, $\mathbb{E}_{\gamma^G}\{\cdots\}$ denotes the difference in expected utility for a member of party G between having a committed candidate and an uncommitted one, at the time of the primary election, where the expectation is taken with respect to the party specific voting cost shock that realizes at the general election:

$$\mathbb{E}_{\gamma^A}\left\{\chi^A\mathbb{P}^A(C, NC)\left(I^A + v^A(C)\right) + \chi^A\left(1 - \mathbb{P}^A(C, NC)\right)v^A(NC) - \int_{-\underline{z}}^{z^A(C,NC)} \frac{z + \gamma^A}{\underline{z} + \overline{z}}dz - \right.$$
$$\left. \chi^A\mathbb{P}^A(NC, NC)\, I^A - \chi^A v^A(NC) + \int_{-\underline{z}}^{z^A(NC,NC)} \frac{z + \gamma^A}{\underline{z} + \overline{z}}dz\right\} \geq 0. \quad (7)$$

If we derive the lhs of this inequality with respect to I^A we have the following expression:

$$\frac{\left(\chi^A\right)^4}{\left(\chi^B\right)^2}(\underline{z} + \overline{z})\left(v^A(C) - v^A(NC)\right) - \left(\chi^B\right)^2(\underline{z} + \overline{z})\left(v^B(NC) - v^B(C)\right), \quad (8)$$

which does not depend on I^A. By the envelope theorem we did not include the derivatives of the equilibrium threshold z^A with respect to I^A. Recall that we assume $v^B(NC) - v^B(C) > 0$ and $v^A(C) - v^A(NC) > 0$. When either λ^A is sufficiently low (which implies that $v^A(C) - v^A(NC)$ becomes arbitrary small), or λ^B is sufficiently large ($v^B(NC) - v^B(C)$ becomes arbitrary large), expression (8) is negative, hence the probability that A deviates to no commitment increases with I^A.

Consider the conditions of party B, for an equilibrium where party A selects a candidate that proposes commitment and party B selects a candidate that proposes no commitment:

$$\mathbb{E}_{\gamma^B}\left\{\chi^B\mathbb{P}^B\left(C,NC\right)\left(I^B+v^B(NC)\right)+\chi^B\left(1-\mathbb{P}^B\left(C,NC\right)\right)v^B(C)-\int_{-\underline{z}}^{z^B(C,NC)}\frac{z+\gamma^B}{\underline{z}+\overline{z}}dz-\right.$$
$$\left.\chi^B\mathbb{P}^B\left(C,C\right)I^B-\chi^Bv^B(C)+\int_{-\underline{z}}^{z^B(C,C)}\frac{z+\gamma^B}{\underline{z}+\overline{z}}dz\right\}\geq 0. \quad (9)$$

If we derive the lhs of this inequality with respect to I^A we have the following expression:

$$-\frac{\left(\chi^A\right)^3}{\chi^B}\left(\underline{z}+\overline{z}\right)\left(v^B(NC)-v^B(C)\right).$$

This expression is negative. Hence the incentives of party B to propose commitment increase.

□

Lemma 8. *If λ^B and I^A are sufficiently large, then in equilibrium both candidates propose the no-commitment policy.*

Proof of Lemma 8: Consider the conditions (7) of party A, for an equilibrium where party A selects a candidate that proposes commitment and party B selects a candidate that proposes no commitment. By taking the derivative of the lhs of that inequality with respect to I^A, it can be noticed that, for a sufficiently large λ^B, the derivative becomes arbitrarily negative. Moreover notice that the derivative of the lhs of inequality (7) with respect to λ^B is

$$-\left(I^A+v^A(C)-v^A(NC)\right)\left(\chi^B\right)^2\left(\underline{z}+\overline{z}\right)\frac{\partial[v^B(NC)-v^B(C)]}{\partial\lambda^B},$$

which is negative, because $\frac{\partial[v^B(NC)-v^B(C)]}{\partial\lambda^B}>0$. So the arbitrarily large decrease in the lhs of inequality (7), given by a large I^A, is not compensated by an increase in the lhs when λ^B is large. Hence, for a sufficiently large λ^B, there exists a threshold for I^A such that, when I^A is larger than this threshold, the lhs of inequality (7) is negative: the candidate of party A proposes no

commitment. Notice also that for a large I^A, it is always possible to find a sufficiently large λ^B such that the probability of winning of a party is strictly larger than 0 and strictly lower than 1, so the derivative of inequality (7) with respect to I^A takes expression (8). Now, consider the conditions of party B, for an equilibrium where party A selects a candidate that proposes commitment and party B selects a candidate that proposes no commitment (condition (9)). Given that $v^B(NC) > v^B(C)$, for $I^B = 0$ party B's candidate proposes no commitment, because it is a dominant strategy and does not depend on the choice of party A's candidate and on the level of I^A. The derivative of the lhs of inequality (9) with respect to I^B is

$$\frac{\left(\chi^B\right)^4}{\left(\chi^A\right)^2}\left(\underline{z}+\overline{z}\right)\left(v^B(NC)-v^B(C)\right)-\left(\chi^A\right)^2\left(\underline{z}+\overline{z}\right)\left(v^A(C)-v^A(NC)\right). \tag{10}$$

Therefore, when λ^B is sufficiently large, expression (10) is positive, hence the probability that B deviates to commitment decreases with I^B. The derivative of the lhs of inequality (9) with respect to λ^B is

$$\mathbb{P}^B\left(C, NC\right)\frac{\partial[v^B(NC)-v^B(C)]}{\partial\lambda^B},$$

which is positive. So the decrease in the probability that B's candidate deviates to commitment is not compensated by an increase in this probability when λ^B increases. Given that, for any level of I^A, and for $I^B = 0$, the dominant strategy by party B's candidate is to propose no commitment and given the sign of expression (10), this must be true also for any level of I^A and I^B and for the set of parameters such that party A's candidate proposes commitment. The same reasoning can be applied to the case in which party A's candidate proposes no commitment. Indeed, for $I^B = 0$, it is a dominant strategy for party B's candidate to propose no commitment. Consider the conditions of party B, for an equilibrium where party A selects a candidate that proposes no

commitment and party B selects a candidate that proposes no commitment:

$$\mathbb{E}_{\gamma^B}\left\{\chi^B\mathbb{P}^B\left(NC,NC\right)I^B+\chi^B v^B(NC)-\int_{-\underline{z}}^{z^A(NC,NC)}\frac{z+\gamma^B}{\underline{z}+\overline{z}}dz-\right.$$
$$\left.\chi^B\mathbb{P}^B\left(NC,C\right)\left(I^B+v^B(C)\right)-\chi^B\left(1-\mathbb{P}^B\left(NC,C\right)\right)v^B(NC)+\int_{-\underline{z}}^{z^A(NC,C)}\frac{z+\gamma^B}{\underline{z}+\overline{z}}dz-\right\}\geq 0 \quad (11)$$

If we derive the lhs of this inequality with respect to I^B we have the following expression:

$$\frac{\left(\chi^B\right)^4}{\left(\chi^A\right)^2}\left(\underline{z}+\overline{z}\right)\left(v^B(NC)-v^B(C)\right)-\left(\chi^A\right)^2\left(\underline{z}+\overline{z}\right)\left(v^A(C)-v^A(NC)\right).$$

Hence, when λ^B is sufficiently large, this expression is positive and the same reasoning done for the equilibrium with A' candidate proposing commitment can be done for this equilibrium. Notice also that the derivative of the lhs of inequality (11) with respect to λ^B is positive.

□

Proof of Proposition 4: We focus on the asymmetric equilibrium where candidate a is committed and candidate b is not. The utility of candidate a if b is elected, is $-(1-S^*)\lambda^w\left[(1-p)(1-h)\sigma^2+p\bar{L}\right]$. Candidate a maximizes

$$\mathbb{P}^A(C,NC)\left[-\lambda^w\sigma^2+R\right]-\left(1-\mathbb{P}^A(C,NC)\right)(1-S^*)\lambda^w\left[(1-p)(1-h)\sigma^2+p\bar{L}\right]-\Psi\left(n_a\right),$$

with respect to n_a.

The utility from having the NC policy for candidate b, if she is elected and the optimal policy is not revealed through information acquisition, is

$$-(1-p)(1-h)\lambda^w\sigma^2-p\lambda^w\bar{L}.$$

The expected utility from being bribed for candidate b, if she is elected and the optimal policy is

not revealed through information acquisition, is $p\mathbb{E}_{q^L\in\mathbb{R}}[\bar{m}]$. Candidate b therefore maximizes

$$\left(1-\mathbb{P}^A(C,NC)\right)\left[(1-S^*)\left(-\lambda^w(1-p)(1-h)\sigma^2-p\lambda^w\bar{L}+p\mathbb{E}_{q^L\in\mathbb{R}}[\bar{m}]\right)+R\right]-\mathbb{P}^A(C,NC)\lambda^w\sigma^2-\Psi\left(n_b\right),$$

with respect to n_b. Notice that, by increasing the cost of information acquisition, candidate b increases the probability that a is elected, because the probability $\mathbb{P}^A(C,NC)$ depends negatively on the utilities that every citizen gets from having an uncommitted candidate in office.

The optimal amount of effort exerted by candidate a solves the following first order condition, where we used expression (4) for $\mathbb{P}^A(C,NC)$:

$$\begin{aligned}&\left(-\lambda^w\sigma^2+R+\lambda^w(1-p)(1-h)\sigma^2+\lambda^wp\bar{L}\right)\cdot\\ \cdot&\left(\frac{\left(\chi^B\right)^2}{\chi^A}\mathbb{E}_{k^j,j\in B}\left[c\left(s_B^{j*}\right)\right]+\frac{\left(\chi^A\right)^3}{\left(\chi^B\right)^2}\mathbb{E}_{k^j,j\in A}\left[c\left(s_A^{j*}\right)\right]\right)(\underline{z}+\overline{z})+\\ &\left(1-\mathbb{P}^A(C,NC)\right)\lambda^w\left[(1-p)(1-h)\sigma^2+p\bar{L}\right]\frac{\partial S^*}{\partial n_a}=\Psi'\left(n_a\right).\end{aligned}$$

By the envelope theorem, derivatives with respect to s^j in the utilities of voters (present in the probability of winning) are not included. Expression $-\lambda^w\sigma^2+R+\lambda^w(1-p)(1-h)\sigma^2+\lambda^wp\bar{L}$ is the difference in utility for candidate a when she is elected with respect to her opponent. This difference is positive when R is sufficiently large. Moreover, the following holds: $\frac{\partial S^*}{\partial n_g}<0$, because a larger cost of information acquisition reduces all citizens' information. Hence candidate a benefits from fake news because, increasing the cost of monitoring, she reduces the expected benefit of the voter of electing an uncommitted candidate and therefore increases her chances to be elected. However, if she is not elected, a higher cost of information acquisition implies less monitoring for her opponent and therefore a has a lower expected utility from policy.

The optimal amount of effort exerted by candidate b solves the following first order condition:

$$
\begin{aligned}
&-\Big((1-S^*)\left(-\lambda^w(1-p)(1-h)\sigma^2-p\lambda^w\bar{L}+p\mathbb{E}_{q^L\in\mathbb{R}}[\bar{m}]\right)+R+\lambda^w\sigma^2\Big)\cdot \qquad (12)\\
&\cdot\left(\frac{\left(\chi^B\right)^2}{\chi^A}\mathbb{E}_{k^j,j\in B}\left[c\left(s_B^{j*}\right)\right]+\frac{\left(\chi^A\right)^3}{\left(\chi^B\right)^2}\mathbb{E}_{k^j,j\in A}\left[c\left(s_A^{j*}\right)\right]\right)(\underline{z}+\overline{z})\\
&-\Big(1-\mathbb{P}^A(C,NC)\Big)\left(-\lambda^w(1-p)(1-h)\sigma^2-p\lambda^w\bar{L}+p\mathbb{E}_{q^L\in\mathbb{R}}[\bar{m}]\right)\frac{\partial S^*}{\partial n_b}=\Psi'\left(n_b\right),
\end{aligned}
$$

where again, by the envelope theorem, derivatives with respect to s_G^{j*} in the utilities of voters are not included. Expression $(1-S^*)\left(-\lambda^w(1-p)(1-h)\sigma^2-p\lambda^w\bar{L}+p\mathbb{E}_{q^L\in\mathbb{R}}[\bar{m}]\right)+R+\lambda^w\sigma^2$ is the difference in utility for candidate b when she is elected with respect to her opponent. Again, this difference is positive, when R is large. Increasing the circulation of fake news lowers the probability that candidate b is elected, because it increases the cost of information, and therefore of monitoring a non-committed candidate. Notice that the following holds: $-p\lambda^w\bar{L}+p\mathbb{E}_{q^L\in\mathbb{R}}[\bar{m}]=-\lambda^w p(1-h)\sigma^2$, because the elected politician is kept indifferent between accepting and rejecting the lobby's contribution (see the proof of Lemma 1). Hence we can simplify the following expression: $-\lambda^w(1-p)(1-h)\sigma^2-p\lambda^w\bar{L}+p\mathbb{E}_{q^L\in\mathbb{R}}[\bar{m}]=-\lambda^w(1-h)\sigma^2$, which is negative. Therefore, conditional on being elected, the non-committed politician in expectation is worse off when citizens do not discover the optimal policy. Given that all expressions on the left hand side of equation (12) are negative, there is no benefit for candidate b in circulating fake news.

A note on the generalization of this finding: This result derives from a feature of the lobbying subgame, in which the politician does not have any bargaining power with the lobby. In a model where the politician had bargaining power, she would receive a larger offer and in this case candidate b could benefit from increasing the cost of information acquisition because it would reduce voter's monitoring ex-post and consequently increase the possibility of being bribed. However, increasing the cost of information acquisition would still imply a reduction of the citizen's utility of voting for the uncommitted candidate, hence the probability that her opponent a wins increases. When candidates have strong incentives to win elections, that is R sufficiently large, the committed

candidate would still produce more fake news than non-committed candidate.

□