# Microtargeting, Voters' Unawareness, and Democracy[*†]

Freek van Gils[‡] Wieland Müller[§] Jens Prüfer[¶]

November 21, 2023

**Abstract**

Recent technological developments have raised concerns about threats to democracy because of their potential to distort election outcomes: (a) data-driven voter research enabling political microtargeting, and (b) growing news consumption via social media and news aggregators that obfuscate the origin of news items, leading to voters' unawareness about a news sender's identity. We provide a theoretical framework in which we can analyze the effects that microtargeting by political interest groups and unawareness have on election outcomes in comparison to "conventional" news reporting. We show which voter groups suffer from which technological development, (a) or (b). While both microtargeting and unawareness have negative effects on voter welfare, we show that only unawareness can flip an election. Our model framework allows the theory-based discussion of policy proposals, such as to ban microtargeting or to require news platforms to signal the political orientation of a news item's originator.

**Keywords:** disinformation, interest groups, news platforms, microtargeting, voter awareness.

[*]This paper previously circulated under the title "Big Data and Democracy."

[†]We thank Stefan Behringer, Dan Bogart, Giacomo Calzolari, Rahul Deb, Mikhail Drugov, Armando Garcia Pires, Simon Hartmann, Maarten Janssen, Amedeo Piolatto, Molly Roberts, Tanya Rosenblat, Yossi Spiegel, Juha Tolvanen, seminar participants in Tilburg, Liège and St. Gallen, and conference participants at the Media Economics Workshop (Salento, 2018), the Media and Digital Economy Seminar (EUI Florence, 2019), the Workshop on Governance of Big Data and AI (Tilburg, 2019), the SIOE conference (Stockholm, 2019), the EEA Annual Congress (Manchester, 2019) and the Digital Economics Conference (Toulouse, 2020) for helpful comments on earlier drafts. We have no conflicts of interest but acknowledge funding from the Data Science research fund of the Tilburg School of Economics and Management.

[‡]Netherlands Authority for Consumers and Markets (ACM), PO Box 16326, 2500 BH The Hague, the Netherlands; freekvangils@outlook.com. Address when this research was carried out: Department of Economics, CentER, TILEC, Tilburg University, PO Box 90153, 5000 LE Tilburg, the Netherlands.

[§]Department of Economics, VCEE, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria and Department of Economics, CentER, TILEC, Tilburg University, PO Box 90153, 5000 LE Tilburg, The Netherlands; wieland.mueller@univie.ac.at.

[¶]Department of Economics and TILEC, Tilburg University, PO Box 90153, 5000 LE Tilburg, the Netherlands and School of Economics and Centre for Competition Policy, University of East Anglia, Norwich, United Kingdom; j.prufer@tilburguniversity.edu.

# 1 Introduction

Democracy comes with many virtues. To hold governments accountable in representative democracies, voters depend on political information provision. The level and credibility of voters' information affects their trust in political leaders, institutions, election outcomes, and, hence, in the functioning of democracy itself (Van der Meer (2017)). As a principal source of political information, the media is of critical importance to democratic societies.

However, recent technological developments affecting information available *about* voters, means to provide information *to* voters, and the nature of information acquisition *by* voters have raised concerns about threats to democracy because of their alleged potential to distort election outcomes (Kavanagh et al., 2019). These developments concern (a) data-driven voter research and the possibility of political microtargeting, and (b) news consumption of growing numbers of people using social media and news aggregators that obfuscate the origin of news, leading to voter unawareness about the news sender's identity.

Platforms collect vast amounts of data on users' preferences and characteristics by tracking them on and outside of the platform and by acquiring third-party data. Platforms can infer a range of attributes from these data, most notably users' political views (Kosinski, Stillwell, and Graepel, 2013). Some platforms also offer microtargeted advertising services, which can be used by political interest groups to tailor news to preferences and characteristics of individual voters. Microtargeting allows interest groups (or advertisers) to differentiate their news reports, which may contain disinformation, to influence voters' beliefs in their favor in each subgroup of the electorate.[1]

Today more than half of digital news consumers use an algorithm-driven platform such as social media, search engines, and news aggregators as their main way to obtain news (Newman et al., 2020). Users may find it difficult to distinguish among the multitude of news senders, which, arguably, leads to voters' unawareness of a sender's identity. Indeed, platform users demonstrate a lack of recognition of outlet identity and are less able to attribute news items to the outlets that reported them if they saw the news on a platform than if they accessed it directly.[2]

[1]Facebook's Custom Audience is a prominent example of a microtargeted advertising service. According to investigative journalism outlet ProPublica, Facebook offers a list of 29,000 user categories that ad buyers can use to determine their target audience (`https://www.propublica.org/article/facebook-doesnt-tell-users-everything-it-really-knows-about-them`).

[2]Kalogeropoulos and Newman (2017) report that, when a news item was accessed directly on the original website, users' recall rate of the originator's identity in their study was 81%, as compared to 47% if the news item was accessed via social media and to 37% if accessed via a search engine. Kang et al. (2011) find that a news portal website user's assessment of the credibility of a news story tends to be primarily influenced by the identity of the portal and less by the original source of the news if the user reports to be not too 'involved' with the news story. Even if a user notices the source of a political message,

A variety of actors spread untrustworthy content on platforms with the aim of promoting their own political goals (Tucker et al., 2018). Disinformation produced by highly partisan websites (Faris et al., 2017), false news websites (Allcott and Gentzkow, 2017), and foreign governments (Maréchal, 2017) has been disseminated on news platforms, which are also reported to advance blurring of the line between fact and opinion (Kavanagh and Rich, 2018). If the electorate is led astray by disinformation, it is not clear that election outcomes reflect voters' true preferences.[3] This is already problematic *per se*. Even more concerning, it calls into question the legitimacy of elections and, hence, may undermine citizens' trust in democracy. [4] Our paper suggests explanations on how this is possible and studies who is most affected.

Despite the widespread attention to disinformation, microtargeting, and potential user unawareness on news platforms in public debate, the academic literature is arguably lagging behind. There is only one empirical research paper that shows a causal link from politically-motivated social media use in an election (the 2016 U.S. Presidential election) to voting behavior (Liberini et al., 2020). There is no political theory about the influence of political interest groups on voting behavior that captures the specificities of news platforms.[5]

We address this gap in the literature and ask what role microtargeting technologies and voter unawareness about the political position of a news sending interest group play for the potential to manipulate elections. Can rational voters' ex-post beliefs be affected by news they receive via news platforms from ideological interest groups, such that voters make "wrong" voting decisions in equilibrium? If so, what is the bigger problem, microtargeting or voters' unawareness about interest groups' political positions, and why?

---

it might be hard to find out the ideological leaning of the sender. Despite implemented transparency initiatives, political campaign groups with undisclosed funding entities ran more political ads on Facebook than any of the registered parties in the two months before the 2019 United Kingdom general election (https://www.ft.com/content/f42f9aa2-16ba-11ea-8d73-6303645ac406).

[3]We follow the definition of Tucker et al. (2018): "Disinformation [...] is intended to be a broad category describing the types of information that one could encounter online that could possibly lead to misperceptions about the actual state of the world." For instance, by selectively reporting one-sided information (truthfully) an interest group produces disinformation but not fake news.

[4]64 percent of U.S. adults say fabricated news stories cause a great deal of confusion about the basic facts of current issues and events (https://www.pewresearch.org/fact-tank/2017/10/04/key-trends-in-social-and-digital-news-media/). 68 percent of U.S. adults say false news undermine their confidence in the government (https://www.journalism.org/2019/06/05/many-americans-say-made-up-news-is-a-critical-problem-that-needs-to-be-fixed/).

[5]We use the term *interest group* as reference to all ideologically motivated suppliers of political content, that is, it includes traditional newspapers or TV channels but also the websites and social media accounts of parties, political organizations, and individuals. Whereas many of these organizations also have other communication channels, where voters are aware of a sender's identity, we focus on communication via platforms, which disguises the original sender. *News platforms* include but are not limited to social media, news aggregators, and search engines.

## Overview of Our Theory

We incorporate the two technological key features of today's news platforms into a game-theoretical model: platforms may (*i*) enable *microtargeted* matching of news to users based on users' preferences and characteristics and (*ii*) *impede users' awareness* of the original interest group that reports the news. Our model comprises two kinds of active players, an interest group and voters. The interest group reports political news that is disseminated via a news platform and voters consume news and can elect political parties. A binary state of the world, which is drawn from a commonly known probability distribution, objectively favors either a left-wing or right-wing policy. The interest group knows the state of the world but voters do not. On a classical left-right political spectrum, nature determines the positions of the interest group and two political parties (who are committed to implement commonly known policies if elected). Voters are uniformly distributed over the political spectrum and have a privately known cost of voting.

The timing of actions is that, first, the interest group sends a message to voters about whether the state of the world is either favoring a left- or a right-leaning policy. Voters receive the message, update their beliefs about the state of the world, and then cast their vote for one of the parties, or abstain. While voters maximize expressive utility from voting, the interest group minimizes the weighted mismatch costs between its own preferred policy position and the ones of the two parties, depending on the realized state of the world.

We consider four different games, which are determined by varying the two essential features discussed above. First, the interest group must either send the same message to all voters (*public*) or can let the message depend on individual voters' ideological position (*microtargeting*). Second, when updating beliefs, voters can either be *aware* or *unaware* of the political position of the interest group.

For each of the four games, we characterize the Perfect Bayesian Equilibrium with the highest voter welfare. We show that in all games some voters and interest group types always prefer the same party over the other one independent of the state of the world (which we call "radicals"), whereas others change their party preferences in line with the state of the world (called "moderates"). Because radical interest groups ignore their information about the state of the world and always try to send messages that support their preferred party, in games with awareness about the sender's position all voters ignore messages from radical groups. By contrast, we show that the messages sent by moderate interest groups can be truthful in equilibrium: those groups have an incentive to inform moderate voters truthfully as their goals are aligned. If moderate groups are constrained by public news dissemination, they inevitably also inform radical voters about the truth. However, with microtargeting that *disciplining effect* disappears and radical voters do not receive valuable information

about the state of the world anymore because all types of interest groups have an incentive to manipulate their beliefs. Hence, radical voters, who have very strong party preferences but still benefit from truthful news, now rationally ignore all news and suffer most from microtargeting.

By contrast, the switch from awareness to unawareness hurts in particular "moderate" voters. In equilibrium these voters know that a moderate interest group would inform them correctly but a radical interest group would always send them the same uninformative message. Thus, with awareness they can either completely rely on the message received or completely discard it. Without awareness, moderate voters have to guess, which means that they will sometimes discard a truthful message from a moderate interest group and sometimes believe an uninformative message sent by a radical interest group. Both changes hurt the payoffs of moderate voters.

A voter-welfare ranking among our games produces policy-relevant results: the *public game with voter awareness* is ranked highest, whereas the *microtargeting game with voter unawareness* is ranked lowest. The other two games occupy intermediate ranks, depending on parameter values. Additionally, we show that voter unawareness is a necessary condition for *election flipping* (to change the election winner). Microtargeting alone cannot distort election outcomes qualitatively. Studying competition among interest groups, we show that voter welfare increases with increasing competition.

These results allow the theory-based discussion of policy proposals. One proposal is that news platforms could be compelled to *implement technologies by which users can identify a message's original sender*. This should help users to also infer the sender's political position. In our model's language, this provision would help to establish *awareness* among voters and thereby decrease the risk of flipping election outcomes even in the presence of *microtargeting*.[6]

## 2 Literature Review

This is the first paper analyzing the effects of interest groups' information provision via news platforms on voting behavior. It contributes to the literature on interest groups' influence on policy outcomes through information transmission to voters.[7] Yu (2005) models how two competing interest groups influence both an incumbent government and homogeneous voters,

[6]Facebook recently implemented a requirement in its Custom Audience service that "In the drop-down menu of each ad, the "Why am I seeing this?" section will show people the *source of the information (advertiser or partner)* [...]" (our emphasize). See https://www.facebook.com/business/news/introducing-new-requirements-for-custom-audience-targeting. However, it remains unclear how much this feature is used and to what extent it diminishes users' unawareness.

[7]Most of this literature studies direct lobbying of politicians by interest groups, which we do not study. See Grossman and Helpman (2001) and Van Winden (2004) for broad discussions.

who are exposed to the same message. Voters' posterior beliefs are an exogenous function of a prior belief and the number of messages received from both interest groups. In contrast, we consider heterogenous voters, study the possibility of microtargeted communication and model voters as Bayesian updaters, taking into account all aspects of the game.

Like our paper, Shapiro (2016) studies the effects of false claims made by interest groups on voting behavior. Whereas we study direct communication from an interest group to voters, in Shapiro (2016) an interest group can only reach voters through a journalist's news coverage.[8] In Shapiro (2016), disclosure of the information sender's political position helps voters because it takes away the journalist's reputational incentive to report ambiguous news when facts are not in line with his predisposition. In our framework, voters benefit from disclosure even without reputational concerns.

Our paper also adds to theoretical research on supply-driven media bias and political outcomes. There, media organizations may manipulate news content to advance the ideological agenda of journalists (Baron, 2006) or editors (Sobbrio, 2014) or yield to pressure from governments (Besley and Prat, 2006). A typical supply-driven media bias model contains a media outlet that commits to a (potentially biased) reporting strategy to maximize a payoff function that captures both a profit motive and a political or commercial motive.[9] By contrast, our model abstracts from profit motives and the reporting strategy is unobservable to voters.

Alonso and Padró I Miquel (2023) consider a model in which two special interest groups with diametrically opposed interests can spend resources to capture media sources. Once captured, media sources can disseminate any message (from a continuous message space) independent of the underlying binary state of the world. In equilibrium, the levels of capture and lying by interest groups lead to polarization regarding news as more extreme messages are sent more often, but rational news consumers (i.e., citizens) cannot be deceived completely as they become skeptical. However, in equilibrium social learning is weakened as informative messages are jammed, so that overall informativeness is reduced. While we mainly analyze a model featuring just one interest group, in an extension we show that allowing for competition among interest groups increases voter welfare in all of our four games. Note also that, in contrast to our model, there is no voting in Alonso and Padró I Miquel (2023).

We contribute to an emerging, and mostly empirical, literature on the political effects of social media, reviewed by Zhuravskaya, Petrova, and Enikolopov (2020) and Tucker et al. (2018). [10] Our theory complements Liberini et al. (2020), who study Facebook advertisement

[8] Related, Sobbrio (2011) considers policy-motivated media outlets; Petrova (2012) analyzes the link between advertising profitability and media bias.

[9] Gentzkow, Shapiro, and Stone (2015) survey this literature.

[10] Germano and Sobbrio (2020) theoretically study opinion formation through the usage of algorithm-

price variation for different audiences and ask to what extent political campaigns during the 2016 U.S. Presidential elections used social media to microtarget voters. They find that the Republican campaign used extensive Facebook ads and microtargeting and conclude that microtargeted ad campaigns had significant effects on voting behavior.

In an early contribution to the empirical literature on political microtargeting, Hillygus and Shields (2014) use a large variety of data sources to argue that political candidates exploit data-mining technologies and enormous voter databases to identify and target voters by raising "wedge issues," on which voters share a candidate's opinion. Contrasting this view, Hersh (2015) claims that even highly advanced political campaigns often lack accurate information about voter preferences and shows that the political campaigns in the U.S. in the period 2008-2012 mostly relied on a limited set of public voter-data records, even when alternative data sources were available.

Theoretical work on microtargeted advertising in political campaigns includes Schipper and Woo (2019) and Hoffmann, Inderst, and Ottaviani (2020). Schipper and Woo (2019) develop a model in which two political candidates with fixed policy positions on multiple political issues communicate (some) information about their political stance to voters, who vote for the candidate that they perceive as closest to their preferred policy position. The authors show that, with sophisticated voters, microtargeting abilities of candidates are sufficient to get an election outcome that is equivalent to the outcome under complete information. A voter realizes that a political candidate only shares information with her when it makes him look more attractive to her, which gives rise to an unravelling of all relevant information. Without microtargeting abilities, a voter might think that information is withheld to her because it would make the candidate look bad to other voters (but not to her), which implies less information revelation. Within our framework, microtargeting abilities can only hurt and never help information transmission to voters. This stark difference in findings results from the different ways we model communication. Schipper and Woo (2019) study truthful but potentially imprecise communication, whereas communication is cheap talk in our model. Hoffmann, Inderst, and Ottaviani (2020) apply their selective disclosure model to study microtargeted political campaigning and find that both campaigners and voters can benefit from selective disclosure based on voter data. Unlike in our model, information disclosure is necessarily truthful and the incentives of the campaigners are independent of the state of the world.

Levy and Razin (2019) study (online) echo chambers, which result from the choice of news consumers to cluster with like-minded others in combination with a number of

driven platforms such as search engines and highlight the platforms' role in the diffusion of misinformation but do not address the implications for political outcomes.

behavioral biases. In our model, voters do not choose their news source (the interest group). Instead, we assume that voters are randomly matched to an interest group to capture the gatekeeper role of the platform's algorithm.

Lastly, our model builds on cheap talk models, in which a sender observes a payoff-relevant state and sends a costless and non-verifiable message to an uninformed receiver, who then chooses an action that determines the payoffs of both sender and receiver together.[11] Li and Madarász (2008) exclusively consider sender types with state-dependent preferences, whereas we also consider sender types with state-independent preferences (radical interest groups) over voters' actions. Disclosure of interests may harm receivers in Li and Madarász (2008) but unambiguously benefits them in our model.

Farrell and Gibbons (1989) and Goltsman and Pavlov (2011) study cheap talk communication with multiple audiences, which resembles our microtargeting games. They show (in a model without uncertainty about the sender's type) that private communication is beneficial for receivers under some bias configurations and hurts them under other bias configurations.[12] Like them, we find that a restriction to send the same message to multiple audiences can discipline the sender to transmit information to audiences that otherwise would receive uninformative communications.

# 3 A Formal Model of Political News Dissemination

We first describe the model and discuss its key assumptions thereafter. The game features two kinds of players, an interest group and voters, who act in a world in which a news platform disseminates political news from interest groups to voters and voters can elect political parties. There are two political parties, indexed by $j \in \{L, R\}$.

**Voters' preferences and actions:** It is common knowledge that party $L$ and party $R$ are committed to implement policies $x_L$ and $x_R$, respectively, if elected. These policies are elements of a left-right political spectrum (more on this below) and are exogenously given. Voters have preferences over policies and receive expressive utility from voting. Voters are characterized by their position $y$ on the uniformly distributed left-right spectrum $[-b, b]$: voter $-b$ has the most left-wing and voter $b$ has the most right-wing ideological position.[13]

[11] There is a long tradition of applying cheap talk models to explain political phenomena in economics and political science. See for instance Austen-Smith and Banks (2002).

[12] Galeotti, Ghiglino, and Squintani (2013) study public and private cheap talk communication in an environment with multiple players who are both senders and receivers of messages. Our model is closer to Farrell and Gibbons (1989) and Goltsman and Pavlov (2011).

[13] Our results are robust to other distributions of voter ideologies, as long as they have strictly positive density on the interval $[-b, b]$.

A voter's ideal policy $x_v$ depends on her ideological position $y$ and the state of the world $\theta \in \{-1, 1\}$, which, however, voters cannot observe:

$$x_v(y, \theta) = y + \theta. \tag{1}$$

Hence, all voters prefer a relatively more left-wing policy if $\theta = -1$ and a more right-wing policy if $\theta = 1$. Each voter chooses a voting action $a \in \{L, R, 0\}$, where $a = L$ if she votes for party L, $a = R$ if she votes for party R, and $a = 0$ if she abstains from voting. A voter's utility from voting, in the absence of voting costs, is specified as follows:

$$U(a, y, \theta) = \begin{cases} g - t\left(x_a - x_v(y, \theta)\right)^2 & \text{if } a \in \{L, R\} \\ 0 & \text{if } a = 0. \end{cases} \tag{2}$$

The parameter $g > 0$ represents the psychological gain from voting and is needed to endogenize abstentions. The utility received from voting for party $j$ is decreasing in the mismatch cost $t\left(x_j - x_v\right)^2$ that the voter incurs if her ideal policy differs from party $j$'s policy position. Abstaining yields 0 utility. Each voter $y$ incurs a cost of voting, $c_y \in [0, \bar{c}]$, which is independent of $y$ and $\theta$ and is an *i.i.d.* draw from a uniform distribution over $[0, \bar{c}]$. Net voter utility is equal to $U(a, y, \theta) - c_y$ if $a \in \{L, R\}$ and 0 otherwise. Voter welfare is defined as:

$$W(a, c, \theta) = \int_{-b}^{b} U(a, y, \theta)\, dy - \int_{-b}^{b} \mathbb{1}_{\{a_y \in \{L,R\}\}} c_y dy. \tag{3}$$

**Voters' beliefs:** Voters have common prior beliefs $p = Pr(\theta = -1)$ about the state of the world, where $0 < p < 1$. Before voting, voters receive via a news platform a single cheap-talk news item $m \in \{-1, 1\}$ concerning the state of the world. Denote by $\mu(m) = \mu(\theta = -1|m)$ the probability (*posterior belief*) that a voter assigns to the event $\theta = -1$ after observing news item $m$. Denote by $\mathbb{E}\left[U(a, y, \theta)\,|\mu(m)\right]$ a voter's expected utility from voting if she has ideological position $y$ and posterior belief $\mu(m)$.

**Interest Group:** There is a single ideologically motivated interest group that accurately observes the state of the world $\theta$ and sends a news item $m$ about $\theta$.[14] The interest group has ideological position $z$, which is drawn from a uniform distribution over $[-h, h]$ and is unobserved by voters. Corresponding to voters' preferences, the ideal policy position of the

[14] We study competition among interest groups below.

interest group, $x_n$, is determined by its ideological position $z$ and the state of the world $\theta$:

$$x_n(z,\theta) = z + \theta. \tag{4}$$

The interest group reports cheap talk news item $m \in \{-1,1\}$ and earns the payoff

$$\Pi(a,z,\theta) = -\nu_L(a)(x_L - x_n(z,\theta))^2 - (1-\nu_L(a))(x_R - x_n(z,\theta))^2, \tag{5}$$

where

$$\nu_L(a) = \frac{\int_{-b}^{b} \mathbb{1}_{\{a_y=L\}} dy}{\int_{-b}^{b} \mathbb{1}_{\{a_y=L\}} dy + \int_{-b}^{b} \mathbb{1}_{\{a_y=R\}} dy} \tag{6}$$

denotes party $L$'s vote share, and $(x_L - x_n(z,\theta))^2$ is the interest group's mismatch cost. Analogous for party R.

**Four Games:** We study the interaction of two news dissemination technologies and two awareness states of the voters about the ideological position of the interest group $z$, resulting in four different games. In the two *public games*, the interest group is restricted to producing a single news item $m \in \{-1,1\}$ for the entire electorate. In the two *microtargeting games*, the interest group reports a news item $m_y \in \{-1,1\}$ for each voter $y$, unobserved by others. In all games, the interest group may misrepresent the true state of the world. Voters have no way to learn about $\theta$ apart from observing $m$.

All aspects of the game, including the distribution of interest group ideologies and the news dissemination technology but not the realizations of $\theta, z$, and $c_y$, are common knowledge.

**Timing:** The timing of each game is as follows:

*Stage 0:* Nature determines $\theta$ according to $p$, draws $c_y \sim U[0,\bar{c}]$ for each voter $y$, and draws $z \sim U[-h,h]$ for the interest group. Each voter $y$ privately learns $c_y$ and in games with voter awareness also $z$. The interest group observes $z$ and $\theta$.

*Stage 1:* The **interest group** chooses $\mathbf{m} \in \{-\mathbf{1},\mathbf{1}\}$ in the **public games** and $\mathbf{m_y} \in \{-\mathbf{1},\mathbf{1}\}$ for each voter $y$ in the **microtargeting games**.

*Stage 2:* **Each voter y** observes news item $m$ if a public news dissemination technology is in place and privately observes $m_y$ in games with microtargeting. Each voter **updates belief $\boldsymbol{\mu(m)}$**, and **chooses voting action a** $\in \{\mathbf{L},\mathbf{R},\mathbf{0}\}$. All payoffs are realized.

**Equilibrium Concept:** Our solution concept is Perfect Bayesian Nash Equilibrium (PBE). A PBE of the game consists of a reporting strategy $m^*$ of an interest group and a voting

strategy $a^*$ and a belief $\mu^*$ of a voter, which maximize a player's expected payoff, given her beliefs about other players. The games we analyze have multiple PBEs. Therefore, we focus on the *Voter Welfare-Maximizing Perfect Bayesian Equilibrium (VWMPBE)*, *i.e.*, the equilibrium with the highest voter welfare. As we will show for all games, this equilibrium coincides with the equilibrium in which most voters receive news that is informative to them. A formal definition of the equilibrium concept and a set of technical assumptions that make sure the solutions are well behaved are in the Appendix.

## Model Discussion

**Voters' utility:** The probability that a single vote is decisive is low in large elections. Hence, rational voters are unlikely to turn out to vote if they are solely interested in the election outcome (Downs, 1957). Turning out to vote is not paradoxical if voters derive direct *expressive utility from voting.*[15] Following Chan and Suen (2009), the utility that a voter derives from voting for a political party depends on the party's policy platform, the voter's individual-specific taste ("ideology") and an unobserved state of the world. A mathematical equivalent to expressive voting is modeling one representative voter, who is pivotal for election outcomes, by definition. Close to our formulation of voters' utility, in Binswanger and Prüfer (2012) the voter's optimal voting action also depends on a politician's policy platform and an unknown state of the world.

**Interest group's payoffs and information:** Whereas *individual* voters have a negligible impact on the election outcome, an interest group could be decisive for the election outcome by influencing the voting behavior of *multiple* voters. For this reason, we let the payoffs of an interest group depend on the realized election outcome. This modelling decision does not drive our results. Alternatively, we could have assumed that an interest group enjoys expressive utility from a vote for its favored party and expressive disutility from a vote for the opposing party. This would not have altered our findings. As we assume the same structure of the ideal policy position of voters (1) and interest groups (4), the latter can be thought of as being managed by members of the electorate. The weighted mismatch cost of interest groups (5) captures that parties' political influence, e.g. the number of seats held in parliament or the amount of campaign contributions received, usually depends on their vote shares. We also assume that interest groups are only motivated by political interests and not by a preference for truth telling. We will nevertheless show that, depending on the game,

[15] Expressive voters are frequently modeled (*e.g.*, Schuessler (2000) and Glaeser, Ponzetto, and Shapiro (2005)). There is ample empirical support for the expressive voting theory (*e.g.*, Pons and Tricaud (2018)). See Tyran and Wagner (2019) for a survey on expressive voting experiments in the laboratory.

it is a result, not an assumption, that moderate interest groups report truthfully in specific, clearly delineated cases. Shapiro (2016) and Kartal and Tremewan (2018) offer discussions and justification for the assumption that interest groups, having access to expert knowledge and resources, are perfectly informed about the true state of the world.

**A single interest group:** With one interest group, the models' mechanisms can be easier understood. Results for the case of competing messages received by a voter are shown after the baseline model.

**Perfect rationality:** We model perfectly rational voters, who understand the incentives of interest groups to misreport news. This is a strong assumption as many voters have cognitive limitations and imperfect foresight. However, the significance of our results is only strengthened if we can show that and how even rational voters can be manipulated in equilibrium and make voting decisions that are against their own interests. Then, voters with naive beliefs about political messages could, arguably, be manipulated even easier by demagogues. Complementarily, if only a share $\alpha \in (0,1)$ of voters are rational (Bayesian updaters) and $1-\alpha$ voters do not update their beliefs (and if $\alpha$ is not correlated with $y$), our results hold for the rational voters and, hence, qualitatively for the entire electorate.

**Common beliefs:** We assume common prior beliefs for tractability. Replacing the common prior belief $p$ by an individual belief $p_y$ for each voter $y$, would not affect our results qualitatively. Posterior beliefs $\mu(m)$ can differ across voters, which is a crucial feature in our microtargeting games.

**Exogenous awareness:** In practice, interest groups have various communication channels to send messages to voters. In this paper, however, we take the overarching empirical relevance of news platforms for political communication as given and study the consequences of this news consumption pattern for voters' political beliefs, where awareness about a sender's identity is decreased exogenously (Kalogeropoulos and Newman, 2017). The case of endogenous awareness is briefly discussed in footnote 21.

# 4 Equilibrium Analysis

In all four games, the critical task of each voter $y$ is to update her prior belief $p$ about the state of the world $\theta$ to posterior belief $\mu(m)$ after receiving message $m$. Then casting a vote for the party that is perceived to be located closest to $y$'s ideal policy $x_v(y,\theta)$ is straightforward.

The difficulty of voter $y$ is that $m$ is potentially valuable because the interest group has perfect knowledge about $\theta$ but the interest group's payoff increases if it can make the voter vote for *its* preferred party. Hence, the voter should not trust the message blindly.

## Public Game with Voter Awareness

Here the voter knows the political position $z$ and, hence, the objectives of the interest group. However, the voter does not know $\theta$. The interest group, in turn, knows all aspects of the game apart from the realization of an individual voter's voting cost $c_y$. This ignorance does not affect its decision, though.

Denote the ideology of a voter with belief $\mu'$, who is *indifferent* between voting for parties $L$ and $R$, by $\hat{y}_{\{\mu'\}}$. Similarly, denote the ideology of an indifferent interest group by $\hat{z}(\cdot)$. We show in the Appendix that the following proposition holds.

**Proposition 1.** *The following strategy profiles and beliefs constitute the voter welfare-maximizing Perfect Bayesian Equilibrium of the public game with voter awareness:*

$$a^*(y, c_y, m, z) = \begin{cases} L & \text{if } y < \hat{y}_{\{\mu^*(m,z)\}} \text{ and } c_y < \mathbb{E}\left[U\left(a = L, y, \theta\right) | \mu^*\left(m, z\right)\right] \\ R & \text{if } y > \hat{y}_{\{\mu^*(m,z)\}} \text{ and } c_y < \mathbb{E}\left[U\left(a = R, y, \theta\right) | \mu^*\left(m, z\right)\right] \\ 0 & \text{otherwise,} \end{cases} \tag{7}$$

$$m^*(z, \theta) = \begin{cases} -1 & \text{if } z \leq \hat{z}\left(\theta\right) \\ 1 & \text{if } z > \hat{z}\left(\theta\right), \end{cases} \tag{8}$$

$$\mu^*\left(m = -1, z\right) = \begin{cases} p & \text{if } z \leq \hat{z}\left(\theta = 1\right) \text{ or } z > \hat{z}\left(\theta = -1\right) \\ 1 & \text{if } \hat{z}\left(\theta = 1\right) < z \leq \hat{z}\left(\theta = -1\right), \end{cases}$$

$$\mu^*\left(m = 1, z\right) = \begin{cases} p & \text{if } z \leq \hat{z}\left(\theta = 1\right) \text{ or } z > \hat{z}\left(\theta = -1\right) \\ 0 & \text{if } \hat{z}\left(\theta = 1\right) < z \leq \hat{z}\left(\theta = -1\right). \end{cases} \tag{9}$$

Intuitively, the decision rule (7) shows that a voter only votes for her preferred party if the expected utility from voting exceeds her voting cost. An interest group with rather left-wing (right-wing) ideology prefers voters to cast their vote for the left (right) party. However, "rather left-wing" depends on the state of the world, as depicted in Figure 1. If interest group ideology $z$ is "moderate," *i.e.* if $\hat{z}\left(\theta = 1\right) < z \leq \hat{z}\left(\theta = -1\right)$, then it depends on the state of the world which party the interest group favors. If interest group ideology is "radical," *i.e.* if $z \leq \hat{z}\left(\theta = 1\right)$ or $z > \hat{z}\left(\theta = -1\right)$, then the interest group's favorite party is state-independent.

Equation (8) states that the interest group reports $m = -1$ if it favors party $L$ and $m = 1$ if it favors party $R$. Hence, (endogenously emerging) moderate interest groups truthfully report about the state of the world, whereas radical interest groups always report $m = -1$ (if left-wing radical) or $m = 1$ (right-wing radical), whether it is truthful, or not.

The rational voters in this game understand the interest group's strategy. Therefore, if the interest group's ideology is moderate, voters assign a higher probability to $\theta = -1$ if they receive the message $m = -1$ rather than $m = 1$. Therefore, the indifferent voter is located further to the right of the political spectrum for $m = -1$ than for $m = 1$. This is visualized in Figure 2, which also shows that, just as interest groups, voters endogenously come in two variants: "moderate" voters with state-dependent party preferences (with $\hat{y}_{\{\mu^*(m=1)\}} < y < \hat{y}_{\{\mu^*(m=-1)\}}$) and "radical" voters (with $y$ outside of these bounds) whose party preferences do not depend on their beliefs about $\theta$.

As voters can observe $z$ perfectly in this game, they can update their belief about the state of the world conditional on $z$. Equation (9) states that voters trust the messages of moderate interest groups fully, whereas they refrain from updating their prior beliefs after receiving news from a radical interest group.

## Public Game with Voter Unawareness

Now imagine that voters are unaware of the political position $z$ of the message sender.

**Proposition 2.** *The following strategy profiles and beliefs constitute the voter welfare-maximizing Perfect Bayesian Equilibrium of the public game with voter unawareness:*

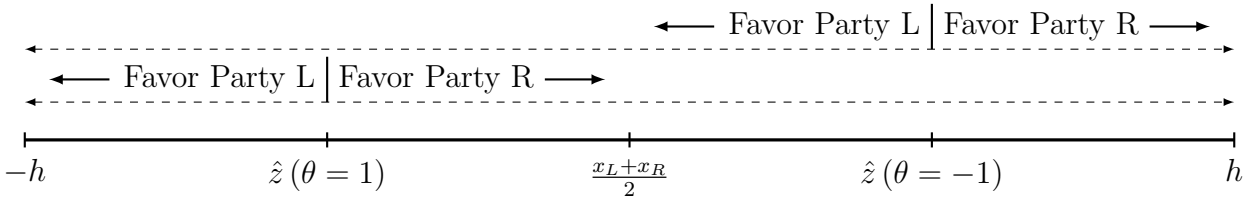



Figure 1: Location of the indifferent *interest group* conditional on the state of the world.

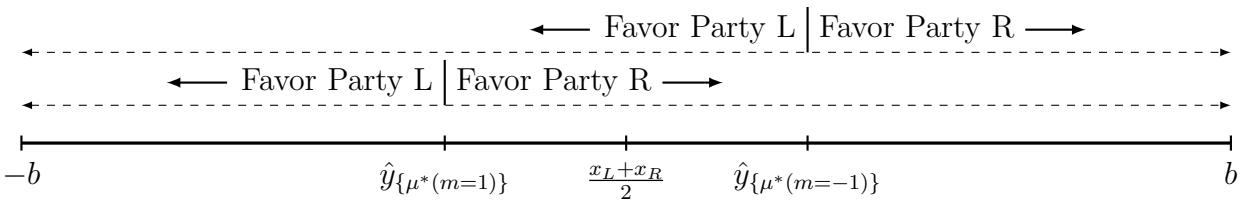



Figure 2: Location of the indifferent *voter* conditional on the public news item received.

$$a^*(y, c_y, m) = \begin{cases} L & \text{if } y < \hat{y}_{\{\mu^*(m)\}} \text{ and } c_y < \mathbb{E}\left[U\left(a = L, y, \theta\right) | \mu^*\left(m\right)\right] \\ R & \text{if } y > \hat{y}_{\{\mu^*(m)\}} \text{ and } c_y < \mathbb{E}\left[U\left(a = R, y, \theta\right) | \mu^*\left(m\right)\right] \\ 0 & \text{otherwise,} \end{cases} \tag{10}$$

$$m^*(z, \theta) = \begin{cases} -1 & \text{if } z \leq \hat{z}\left(\theta\right) \\ 1 & \text{if } z > \hat{z}\left(\theta\right), \end{cases} \tag{11}$$

$$\mu^*\left(m = -1\right) = \begin{cases} \dfrac{p\left(2\left(h+1\right) + x_L + x_R\right)}{2\left(h + 2p - 1\right) + x_L + x_R}, \\ \mu^*\left(m = 1\right) = \begin{cases} \dfrac{p\left(2\left(h-1\right) - x_L - x_R\right)}{2\left(h - 2p + 1\right) - x_L - x_R}. \end{cases} \end{cases} \tag{12}$$

Proposition 2 shows that the incentives of interest groups (11) are similar to the public game with voter awareness: moderate interest groups report truthfully, whereas radical groups report their preferred message state independent. Voters' best response to this reporting strategy is therefore also unchanged (10). See Figures 1 and 2.

What differs now is that voters do not know whether the sender of the message is a moderate or radical interest group. Therefore, voters have to form a belief, which depends on the relative shares of left-wing radical, right-wing radical, and moderate interest groups (12). Consequently, they trust all news a bit—and hence news reporting is payoff-relevant for all voters. This implies that, as compared to the public game with awareness, in equilibrium voters may trust the disinformation of a radical interest group or discount the truthful information from a moderate interest group.

## Microtargeting with Voter Unawareness

Now the platform still distributes news to voters who are unaware of the news source's political position. However, the platform knows a lot about each voter and enables the interest group to microtarget its message to every voter's individual characteristic (*i.e.*, to her political position $y$).

**Proposition 3.** *The following strategy profiles and beliefs constitute the voter welfare-maximizing Perfect Bayesian Equilibrium of the microtargeting game with voter unawareness:*

$$a^*(y, c_y, m_y) = \begin{cases} L & \text{if } y < \hat{y}_{\{\mu^*(m_y)\}} \text{ and } c_y < \mathbb{E}\left[U\left(a = L, y, \theta\right) | \mu^*\left(m_y\right)\right] \\ R & \text{if } y > \hat{y}_{\{\mu^*(m_y)\}} \text{ and } c_y < \mathbb{E}\left[U\left(a = R, y, \theta\right) | \mu^*\left(m_y\right)\right] \\ 0 & \text{otherwise,} \end{cases} \tag{13}$$

$m^*(z, \theta) =$

$$\begin{cases} m_y = -1 & \text{for } Y_1^U < y \leq Y_2^U \\ m_y = -1 \text{ (or 1) with prob. } p \text{ (or } 1-p) & \text{for } y \leq Y_1^U \text{ and } y \geq Y_2^U \end{cases} \text{if } z \leq \hat{z}(\theta)$$

$$\begin{cases} m_y = \ \ 1 & \text{for } Y_1^U < y \leq Y_2^U \\ m_y = -1 \text{ (or 1) with prob. } p \text{ (or } 1-p) & \text{for } y \leq Y_1^U \text{ and } y > Y_2^U \end{cases} \text{if } z > \hat{z}(\theta), \tag{14}$$

*where* $Y_1^U = min\{x_L, \hat{y}_{\{\mu_1\}}\}$ *and* $Y_2^U = max\{x_R, \hat{y}_{\{\mu_2\}}\}$,

$$\mu^*\left(m_y = -1\right) = \begin{cases} \mu_2 = \dfrac{p\left(2\left(h+1\right) + x_L + x_R\right)}{2\left(h + 2p - 1\right) + x_L + x_R} & \text{for } Y_1^U < y \leq Y_2^U \\ p & \text{for } y \leq Y_1^U \text{ and } y > Y_2^U, \end{cases}$$

$$\mu^*\left(m_y = 1\right) = \begin{cases} \mu_1 = \dfrac{p\left(2\left(h-1\right) - x_L - x_R\right)}{2\left(h - 2p + 1\right) - x_L - x_R} & \text{for } Y_1^U < y \leq Y_2^U \\ p & \text{for } y \leq Y_1^U \text{ and } y > Y_2^U. \end{cases} \tag{15}$$

Proposition 3 shows that voters' best voting action (13) has the same structure as in the public games (see (7) and (10)). However, the structure of the interest group's reporting strategy differs: while in the public games messages depend only on a group's own political position $z$ (see (8) and (11)), now they also depend on voter characteristics $y$ (14).

Nevertheless, the interest group is still more likely to favor party $L$ if $\theta = -1$ than if $\theta = 1$, as depicted in Figure 1. Therefore, moderate voters assign a higher probability to $\theta = -1$ if they receive the message $m_y = -1$ than if they receive $m_y = 1$ (15). Notably, which voters are "moderate" or "radical" slightly differs from the public games. Figure 3.(a) captures the case where $x_L > \hat{y}_{\{\mu_1\}}$ and $x_R < \hat{y}_{\{\mu_2\}}$. Here, voters with positions $y \in [\hat{y}_{\{\mu_1\}}, \hat{y}_{\{\mu_2\}}]$ are moderate, *i.e.* they react to the message received by updating beliefs. By contrast, Figure 3.(b) captures the case where $x_L \leq \hat{y}_{\{\mu_1\}}$ and $x_R \geq \hat{y}_{\{\mu_2\}}$. Here, voters

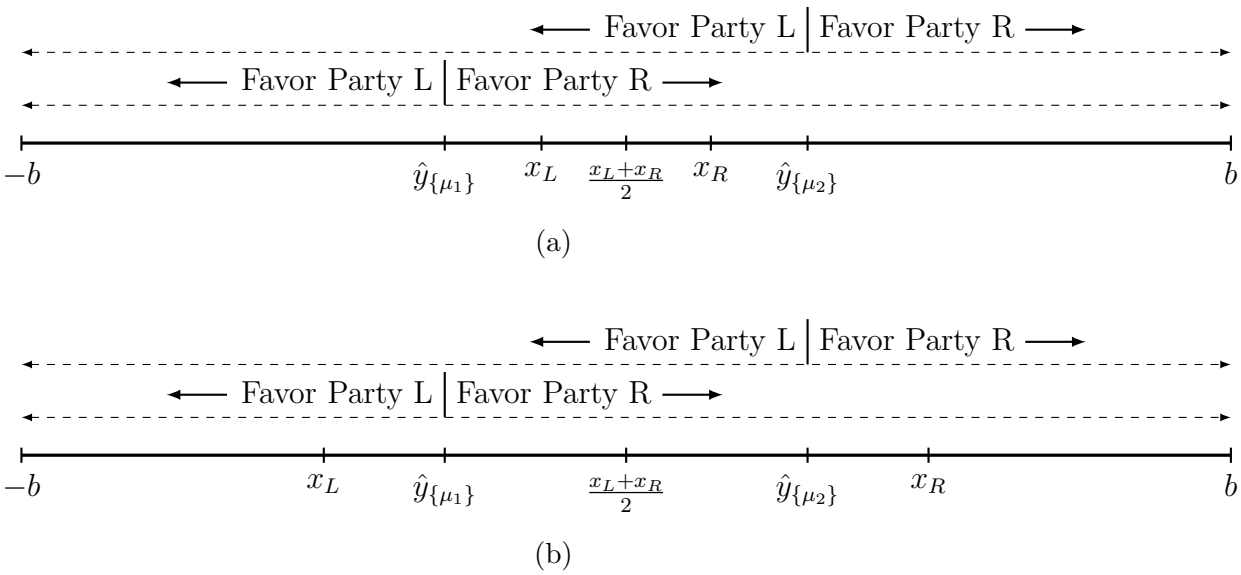



Figure 3: Location of the indifferent voter conditional on the microtargeted news item received. (a) $x_L > \hat{y}_{\{\mu_1\}}$ and $x_R < \hat{y}_{\{\mu_2\}}$. (b) $x_L \leq \hat{y}_{\{\mu_1\}}$ and $x_R \geq \hat{y}_{\{\mu_2\}}$.

with positions $y \in [x_L, x_R]$ are "moderate."[16]

The news strategy (14) shows that moderate voters get truthful news from moderate interest groups. Their problem is that they cannot identify a news sender's type. Therefore, just as in the public game with unawareness, moderate voters react to news a bit (where the updating probability depends on the share of moderate vs. radical left and radical right interest groups; see $\mu_1$ and $\mu_2$ in (15)). However, "radical" voters realize that interest groups have an overwhelming incentive to disinform them in order to make them vote for the interest group's preferred party—and hence rationally ignore the content of all news.

Summarizing, microtargeting makes radical voters dismiss all incoming news. Moderate voters do take news into account to some extent, which makes news payoff-relevant to them.

## Microtargeting with Voter Awareness

Finally, assume the interest group can microtarget voters but voters are aware of the political position $z$ of the interest group.

**Proposition 4.** *The following strategy profiles and belief constitute the voter welfare-maximizing Perfect Bayesian Equilibrium of the microtargeting game with voter awareness:*

[16] In both cases, moderate voters are those voters who become more likely to vote for Party L (Party R) or less likely to vote for Party R (Party L) if they receive the message $m = -1$ ($m = 1$).

$$a^*(y, c_y, m_y, z) = \begin{cases} L & \text{if } y < \hat{y}_{\{\mu^*(m_y)\}} \text{ and } c_y < \mathbb{E}\left[U\left(a = L, y, \theta\right) | \mu^*\left(m_y\right)\right] \\ R & \text{if } y > \hat{y}_{\{\mu^*(m_y)\}} \text{ and } c_y < \mathbb{E}\left[U\left(a = R, y, \theta\right) | \mu^*\left(m_y\right)\right] \\ 0 & \text{otherwise,} \end{cases} \tag{16}$$

$$m^*(z, \theta) = \begin{cases} m_y = -1 & \text{for } Y_1^A < y \leq Y_2^A \\ m_y = -1 \text{ (or } 1) \text{ (with prob. } p \text{ (or } 1-p)) & \text{for } y \leq Y_1^A \text{ and } y > Y_2^A \end{cases} \text{if } z \leq \hat{z}(\theta)$$

$$\begin{cases} m_y = 1 & \text{for } Y_1^A < y \leq Y_2^A \\ m_y = -1 \text{ (or } 1) \text{ (with prob. } p \text{ (or } 1-p)) & \text{for } y \leq Y_1^A \text{ and } y > Y_2^A \end{cases} \text{if } z > \hat{z}(\theta), \tag{17}$$

*where* $Y_1^A = min\{x_L, \hat{y}_{\{\mu=0\}}\}$ *and* $Y_2^A = max\{x_R, \hat{y}_{\{\mu=1\}}\}$,

$$\mu^*\left(m_y = -1, z\right) = \begin{cases} 1 & \text{if } \hat{z}\left(\theta = 1\right) < z \leq \hat{z}\left(\theta = -1\right) \text{ and } Y_1^A < y \leq Y_2^A \\ p & \text{otherwise,} \end{cases}$$

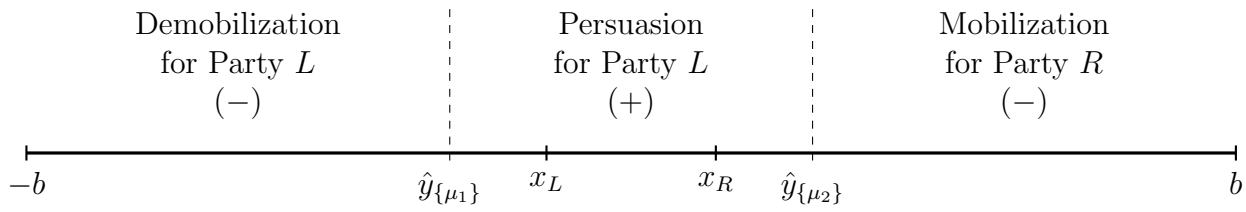
$$\mu^*\left(m_y = 1, z\right) = \begin{cases} 0 & \text{if } \hat{z}\left(\theta = 1\right) < z \leq \hat{z}\left(\theta = -1\right) \text{ and } Y_1^A < y \leq Y_2^A \\ p & \text{otherwise.} \end{cases} \tag{18}$$

Proposition 4 shows that the structure of the interest group's reporting is similar to Proposition 3; see (17): moderate interest group's message contains correct information for moderate voters and uninformative news for radical voters; a message from a radical interest group is always uninformative. As opposed to the microtargeting game with voter unawareness, voters know who is sending the news they receive. Therefore, moderate voters rationally ignore messages by radical groups and fully trust messages from moderate groups. Radical voters still do not trust any messages.

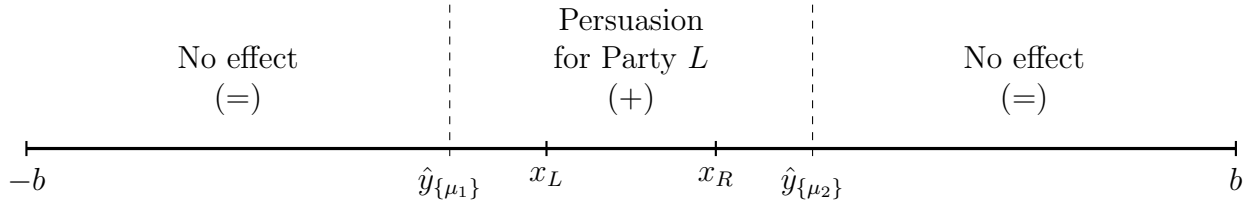
## The Persuasion, (De-)Mobilization, and Disciplining Effects

Based on Propositions 1 to 4, with voters' unawareness the interest group can have a *persuasion effect* on moderate voters and a *(de)mobilization effect* on radical voters. The first channel affects the *incentive-compatibility constraint* of voters, persuading them to vote for the interest group's preferred party instead of the opponent. The second channel affects the *participation constraint* of voters, either mobilizing radicals favoring the interest group's preferred party to participate in the election or demobilizing radicals supporting the opponent by convincing them to abstain.[17]

[17]Empirically, Liberini et al. (2020) show that the advertisement strategies employed in the 2016 U.S. Presidential elections to (a) persuade swing voters to vote for candidate Trump and (b) to make Republican




(a) Public game.




(b) Microtargeting game.

Figure 4: Influence on voting behavior by an interest group that favors party $L$.

Both effects are illustrated in Figure 4 for an interest group supporting party $L$ in a game in which $\hat{y}_{\{\mu_1\}} < x_L < x_R < \hat{y}_{\{\mu_2\}}$: "+" indicates a wanted effect for the interest group; "−" indicates an unwanted side-effect of sending $m = -1$. The interest group can persuade moderate voters ($\hat{y}_{\{\mu_1\}} < y < \hat{y}_{\{\mu_2\}}$) to favor party $L$ over party $R$ in both the public game (panel 4a) and the microtargeting game (panel 4b). However, in the public game sending $m = -1$ has an unwanted side-effect on radical voters,[18] which disappears if messages can be personalized to voter types.[19]

Moreover, Propositions 1 to 4 show that a moderate interest group has an incentive to communicate truthfully to moderate voters because their voting preferences are aligned. This is different w.r.t. radical voters. However, if the interest group can only send one uniform message to all voters, there is a *disciplining effect of the public news dissemination technology*: moderate interest groups report truthfully to *all* voters. Therefore, voters fully trust the message of a moderate interest group. By contrast, a radical interest group always reports the same news. Consequently, if voters can identify a radical interest group, they ignore its message and do not update $p$.

On the voters' side, there are three groups that are affected differently by a change in

voters vote, differed substantially. In our theory, these two different effects arise endogenously despite the simple uni-dimensional message space, $\{-1, 1\}$.

[18]Sending $m = -1$ demobilizes radical left voters, who think party $L$ is not left enough, and mobilizes radical right voters for party $R$, who now believe that $R$ is acceptable and do not abstain.

[19]In the example, the left-favoring interest group would send $m = -1$ to moderate voters and would like to send $m = 1$ to all radicals. The latter, however, is no part of an equilibrium because voters could then infer $\theta$ from $m_y$. Only a mixed strategy solves this dilemma; see (14) and (17).

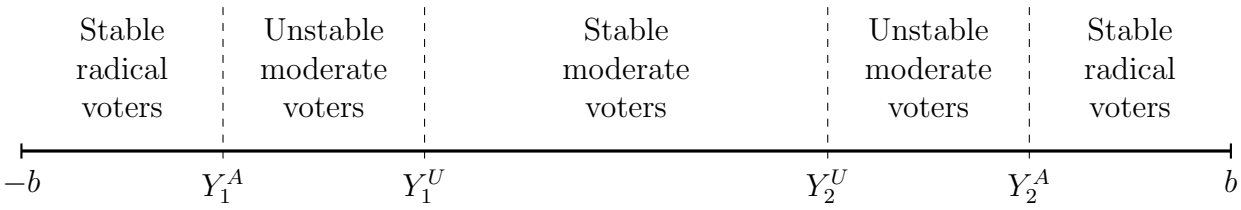



Figure 5: Overview Voter Groups.

their awareness of the interest group's ideological position (see Figure 5): *stable moderate voters* (with ideological position $Y_1^U < y \leq Y_2^U$) and *stable radical voters* ($y \leq Y_1^A$ and $y > Y_2^A$) are moderate or radical, respectively, in all games. However, *unstable moderate voters* ($Y_1^A < y \leq Y_1^U$ and $Y_2^U < y \leq Y_2^A$) are moderate in games with awareness but radical in games with unawareness.[20] Then they always prefer their favorite party over the opponent. With awareness, by contrast, news from an identified, moderate interest group can persuade them because it is truthful, without a shadow of doubt.

# 5 Welfare, Election Flipping and Competition

A news report that is (partly) trusted serves two functions for voters. First, it informs them whether it is worthwhile to turn out to vote, or not. Second, it helps moderate voters to find out which party's policy position is closest to their ideal position. A higher share of trustworthy news and a higher number of voters receiving such news increase total voter welfare.

**Proposition 5.** ***(Voter Welfare)*** *In the voter welfare-maximizing equilibria of the four games analyzed, voter welfare compares as follows:*

*(1) Total voter welfare is strictly higher in public games than in microtargeting games.*

*(2) Total voter welfare is strictly higher in games with awareness than in games with unawareness.*

Proposition 5 yields the following Corollary:

**Corollary 1.** ***(Voters' Ranking of Games)***
*Total voter welfare across the four games ranks as follows:*

1. *Public news with Awareness*

[20] Unstable moderate voters only exist if $x_L > y_{\{\mu=0\}}$ or $x_R < y_{\{\mu=1\}}$.

*2./3. Microtargeting with Awareness*

*2./3. Public news with Unawarenes*

*4. Microtargeting with Unawareness*

The intuition of Corollary 1, which is a central result of this paper, is straightforward. Public news dissemination and awareness of the interest group's political position maximize voter welfare. Public news helps *radical* voters benefit from the discipline effect, whereas awareness enables *all* voters to recognize if an interest group is *moderate* (and hence trustworthy) or *radical* (and hence untrustworthy). On the flip-side, voters fare worst under microtargeting and unawareness, where both of these effects do not exist. The ranking of the two intermediate regimes depends on parameter realizations.

## Flipping an Election with Disinformation

While we have shown that manipulating voters is possible in equilibrium, a critical question remains whether this effect could be large enough to flip an election through disinformation if the winning party would have lost the election ($i$) in the absence of news and ($ii$) if voters had complete information about the state of the world.

The degree to which election flipping is possible depends on the probability that voters are exposed to a malevolent interest group with an incentive to flip an election and on the ability of this interest group to actually flip an election. While the probability of exposure to a malevolent interest group is the same in all games, the ability of an interest group to flip an election differs across games. We construct a single measure, the *election flipping potential*, to compare an interest group's ability to influence the outcome of an election in its favor across the games.

**Definition 1** (**Election Flipping Potential**). *The election flipping potential is defined as the maximum (minimum) prior belief $\hat{p}$ up to which a right-wing (left-wing) interest group can get party R (party L) elected.*

**Proposition 6.** ***(Election Flipping)***

*(1) With voter unawareness, an interest group's election flipping potential is larger in the microtargeting game than in the public game.*

*(2) Voter unawareness is a necessary condition for election flipping.*

Proof: see Appendix. The key insight of Proposition 6 is part (2): as long as voters are aware of an interest group's political position, election flipping is impossible. Without such awareness, elections can be flipped, in principle.[21]

## Interest Group Competition

Assume there are $K \geq 1$ interest groups, each sending one message to voters. Each interest group $k \in \{1, ..., K\}$ has an ideological position, $z_k$, which is *i.i.d.* from a uniform distribution over $[-h, h]$. Denote by $q$ the number of messages $m = -1$ and by $s$ the number of messages $m = 1$ that a voter receives. Hence, $K = q + s$. The remaining elements of the model are unchanged. We prove in the Appendix:

**Proposition 7.** ***(Interest Group Competition)*** *Interest group competition increases voter welfare in all games.*

In games with voter unawareness, posterior beliefs of moderate voters are increasing in $q$, decreasing in $s$ and concave in both $q$ and $s$. Hence, a marginal news report with the same message still affects beliefs but does so at a decreasing rate. Crucially, conflicting messages do not cancel each other out (meaning that voters' posterior beliefs are not equal to their prior beliefs) unless party positions are exactly symmetric. For instance, suppose that $x_L + x_R < 0$, implying that there are more radical right interest groups than radical left interest groups. Receiving a message $m = -1$ (which favors party $L$) is now more informative about the true state than $m = 1$, which implies that a (moderate) voter weighs the former messages more strongly than the latter. Moreover, moderate voters' beliefs converge to the truth if $K$ increases.[22] Finally, the ratio of radical to moderate voters weakly decreases if $K$ increases. The intuition is that one news item might not be enough to persuade voters to change their party preference but multiple news items with the same message can.

[21] Our findings imply that a radical interest group is best off in an environment with voter unawareness. With unawareness, moderate voters believe that every message is somewhat credible, even if it is sent by an unreliable radical interest group. Hence, the radical interest group can deceive some moderate voters to vote for its preferred party. The moderate interest group type, however, is better off with voter awareness, which prevents moderate voters from discounting its message. In an environment in which voter (un)awareness could be endogenously determined by the interest groups (e.g., by disclosure of their types), moderate interest groups would like to distinguish themselves from radical interest groups. If a moderate interest group can perfectly disclose its type, a radical interest group would no longer be able to influence voting and election flipping would not occur anymore. If disclosure is imperfect (i.e., if there is a positive probability that voters do not observe the interest group type), voters still (somewhat) believe a message from a radical interest group but less so than without any disclosure. Hence, election flipping would still be possible but less likely than without any disclosure.

[22] For instance, consider a game with voter unawareness and parameter values $p = 0.5$, $x_L = -1$, $x_R = 1$ and $h = 3$. Suppose $\theta = -1$. In expectation, a moderate voter holds belief $\mathbb{E}[\mu(\cdot)|K = 1] \approx 0.56$ if she receives one message, $\mathbb{E}[\mu(\cdot)|K = 5] \approx 0.76$ if she receives five messages, and $\mathbb{E}[\mu(\cdot)|K = 25] \approx 1$ if she receives 25 messages.

For games with voter awareness, equations (9) and (18) show that a single news item from a moderate interest group is sufficient to resolve all uncertainty about the state of the world. Hence, receiving an additional news item from a moderate interest group has no added value (and radical interest groups' messages are ignored, anyways). However, the *ex ante* probability that a voter is exposed to news from a moderate interest group increases in $K$. Therefore, voters also benefit from interest group competition in games with voter awareness.

# 6 Discussion and Conclusion

Our analysis has shown that microtargeting hurts especially radical voters (because the discipline effect of public news dissemination falls away) and that unawareness about a sender's political position prevents all voters from recognizing whether an interest group is *moderate* (and hence trustworthy) or *radical* (and hence untrustworthy). These key insights can inform policy making.

To reduce microtargeting, a committee of the UK's House of Commons suggested "a minimum limit for the number of voters sent individual political messages [...] at a national level" (House of Commons, 2018, paragraph 142). Consequently, voters could not be targeted individually or in small groups. However, in games with microtargeting we show that even with our most limited message space of $\{-1, 1\}$ it is possible to manipulate an election. Customization of the message content at the individual level is not necessary as long as every voter can be attributed to a certain group with homogeneous characteristics and this group is not too small. Therefore, according to our results, it would be possible to both comply with the House of Commons' proposal and to effectively deceive voters and manipulate an election.

Arguably today's news platforms are best characterized by the *microtargeting game with voter unawareness*, which yields the worst welfare outcome of the four games analyzed.[23] Hence, a better intervention seems to promote policies reestablishing awareness about interest group's political positions, such as, compelling platforms to implement technologies by which users can easily recognize the identity of a specific message's original sender. It should help voters to assess the political position of a news item's originator, to update beliefs about the credibility of news received and, hence, to make voting decisions that are more aligned with their own interests.[24] This would particularly benefit moderate voters.

[23] Moreover, in the Appendix we derive an empirical hypothesis: if unawareness of voters about interest groups' political positions grows over time, e.g. through the proliferation of social media in political news reporting, we expect fewer moderate and more radical voters.

[24] For an example that moves in the direction of such a policy, see footnote 6.

A complementary policy proposal supported by our analysis of competing interest groups is to increase competition among senders of political news, for instance by increasing media plurality, supporting local news stations, and preventing large conglomerates from monopolizing news dissemination.

Our model relies on stylized assumptions. Confirmation bias is said to be a relevant phenomenon among voters (Plous, 1993), which might be included in our framework in future research. Similarly, deviations from the randomized draws and uniform distribution of messages, especially on the news platform's side (platform bias) and on the voters' side (endogenous news consumption) are promising extensions.[25] This requires further empirical work to assess how far these assumptions are acceptable simplifications of reality. Among the most pressing questions are: To which extent do voters fall prey to disinformation about politically relevant events depending on the information they have about interest group's political positions? To which extent can political interest groups actually make use of this weakness and manipulate election outcomes? To get clean results, such empirical testing could first be conducted in lab experiments. Then, to verify the external validity of lab results, they should be tested in the field. A lot of future work is waiting.

[25]Both characteristics were suggested by Piolatto and Schuett (2015).

# References


Allcott, Hunt and Matthew Gentzkow. 2017. "Social media and fake news in the 2016 election." *Journal of Economic Perspectives* 31 (2):211–36.

Alonso, Ricardo and Gerard Padró I Miquel. 2023. "Competitive capture of public opinion." NBER Working Paper 31414.

Austen-Smith, David and Jeffrey S Banks. 2002. "Costly signaling and cheap talk in models of political influence." *European Journal of Political Economy* 18 (2):263–280.

Baron, David P. 2006. "Persistent media bias." *Journal of Public Economics* 90 (1-2):1–36.

Besley, Timothy and Andrea Prat. 2006. "Handcuffs for the grabbing hand? Media capture and government accountability." *American Economic Review* 96 (3):720–736.

Binswanger, Johannes and Jens Prüfer. 2012. "Democracy, populism, and (un)bounded rationality." *European Journal of Political Economy* 28:358–372.

Chan, Jimmy and Wing Suen. 2009. "Media as watchdogs: The role of news media in electoral competition." *European Economic Review* 53 (7):799–814.

Downs, Anthony. 1957. "An economic theory of political action in a democracy." *Journal of Political Economy* 65 (2):135–150.

Faris, Robert, Hal Roberts, Bruce Etling, Nikki Bourassa, Ethan Zuckerman, and Yochai Benkler. 2017. "Partisanship, propaganda, and disinformation: Online media and the 2016 US presidential election." *Berkman Klein Center Research Publication* 6.

Farrell, Joseph and Robert Gibbons. 1989. "Cheap talk with two audiences." *American Economic Review* 79 (5):1214–1223.

Galeotti, Andrea, Christian Ghiglino, and Francesco Squintani. 2013. "Strategic information transmission networks." *Journal of Economic Theory* 148 (5):1751–1769.

Gentzkow, Matthew, Jesse M. Shapiro, and Daniel F. Stone. 2015. "Media bias in the marketplace: Theory." In *Handbook of Media Economics*, vol. 1B, edited by Simon Anderson, Joel Waldfogel, and David Stromberg. Elsevier, 623–645.

Germano, Fabrizio and Francesco Sobbrio. 2020. "Opinion dynamics via search engines (and other algorithmic gatekeepers)." *Journal of Public Economics* 187:104188.

Glaeser, Edward L., Giacomo A.M. Ponzetto, and Jesse M. Shapiro. 2005. "Strategic extremism: Why Republicans and Democrats divide on religious values." *The Quarterly Journal of Economics* 120 (4):1283–1330.

Goltsman, Maria and Gregory Pavlov. 2011. "How to talk to multiple audiences." *Games and Economic Behavior* 72 (1):100–122.

Grossman, Gene M and Elhanan Helpman. 2001. *Special Interest Politics*. MIT press.

Hersh, Eitan D. 2015. *Hacking the electorate: How campaigns perceive voters*. Cambridge University Press.

Hillygus, D. Sunshine and Todd G. Shields. 2014. *The persuadable voter*. Princeton University Press.

Hoffmann, Florian, Roman Inderst, and Marco Ottaviani. 2020. "Persuasion through selective disclosure: implications for marketing, campaigning, and privacy regulation." *Management Science* 66 (11):4958–4979.

House of Commons. 2018. "Disinformation and 'fake news': Interim Report." Digital, Culture, Media and Sport Committee. United Kingdom Parliament.

Janssen, Maarten C.W. and Mariya Teteryatnikova. 2017. "Mystifying but not misleading: when does political ambiguity not confuse voters?" *Public Choice* 172 (3-4):501–524.

Kalogeropoulos, Antonis and Nic Newman. 2017. "I Saw the News on Facebook." Reuters Institute for the Study of Journalism.

Kang, Hyunjin, Keunmin Bae, Shaoke Zhang, and S Shyam Sundar. 2011. "Source cues in online news: Is the proximate source more powerful than distal sources?" *Journalism & Mass Communication Quarterly* 88 (4):719–736.

Kartal, Melis and James Tremewan. 2018. "An offer you can refuse: the effect of transparency with endogenous conflict of interest." *Journal of Public Economics* 161:44–55.

Kavanagh, Jennifer, William Marcellino, Jonathan S. Blake, Shawn Smith, Steven Davenport, and Mahlet Gizaw. 2019. *News in a Digital Age: Comparing the Presentation of News Information over Time and Across Media Platforms*. RAND Corporation.

Kavanagh, Jennifer and Michael D Rich. 2018. *Truth decay: An initial exploration of the diminishing role of facts and analysis in American public life*. Rand Corporation.

Kosinski, Michal, David Stillwell, and Thore Graepel. 2013. "Private traits and attributes are predictable from digital records of human behavior." *Proceedings of the National Academy of Sciences* 110 (15):5802–5805.

Levy, Gilat and Ronny Razin. 2019. "Echo chambers and their effects on economic and political outcomes." *Annual Review of Economics* 11:303–328.

Li, Ming and Kristóf Madarász. 2008. "When mandatory disclosure hurts: Expert advice

and conflicting interests." *Journal of Economic Theory* 139 (1):47–74.

Liberini, Federica, Michela Redoano, Antonio Russo, Angel Cuevas, and Ruben Cuevas. 2020. "Politics in the Facebook Era. Evidence from the 2016 US Presidential Elections." Working paper.

Maréchal, Nathalie. 2017. "Networked authoritarianism and the geopolitics of information: Understanding Russian Internet policy." *Media and Communication* 5 (1):29–41.

Newman, Nic, Richard Fletcher, Anne Schulz, Sımge Andi, and Rasmus Kleis Nielsen. 2020. "Reuters Institute Digital News Report 2020." Reuters Institute for the Study of Journalism.

Petrova, Maria. 2012. "Mass media and special interest groups." *Journal of Economic Behavior & Organization* 84 (1):17–38.

Piolatto, Amedeo and Florian Schuett. 2015. "Media competition and electoral politics." *Journal of Public Economics* 130:80–93.

Plous, Scott. 1993. *The psychology of judgment and decision making.* Mcgraw-Hill Book Company.

Pons, Vincent and Clémence Tricaud. 2018. "Expressive voting and its cost: Evidence from runoffs with two or three candidates." *Econometrica* 86 (5):1621–1649.

Schipper, Burkhard C. and Hee Woo. 2019. "Political Awareness, Microtargeting of Voters, and Negative Electoral Campaigning." *Quarterly Journal of Political Science* 14 (1):41–88.

Schuessler, Alexander A. 2000. *A logic of expressive choice.* Princeton University Press.

Shapiro, Jesse M. 2016. "Special interests and the media: Theory and an application to climate change." *Journal of Public Economics* 144:91–108.

Sobbrio, Francesco. 2011. "Indirect lobbying and media bias." *Quarterly Journal of Political Science* 6:3–4.

———. 2014. "Citizen-editors' endogenous information acquisition and news accuracy." *Journal of Public Economics* 113:43–53.

Tucker, Joshua, Andrew Guess, Pablo Barberá, Cristian Vaccari, Alexandra Siegel, Sergey Sanovich, Denis Stukal, and Brendan Nyhan. 2018. "Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature." Hewlett Foundation.

Tyran, Jean-Robert and Alexander K. Wagner. 2019. "Experimental evidence on expressive voting." In *The Oxford Handbook of Public Choice*, vol. 2, edited by Roger D. Congleton, Bernard N. Grofman, and Stefan Voigt. Oxford University Press, 928–940.

Van der Meer, Tom W. G. 2017. "Political Trust and the 'Crisis of Democracy'." In *Oxford Research Encyclopedia of Politics*, edited by William R. Thompson. Oxford University Press.

Van Winden, Frans. 2004. "Interest group behavior and influence." In *The Encyclopedia of Public Choice*, edited by Charles Rowley and Friedrich Schneider. Springer Science & Business Media, 118–129.

Yu, Zhihao. 2005. "Environmental protection: A theory of direct and indirect competition for political influence." *The Review of Economic Studies* 72 (1):269–286.

Zhuravskaya, Ekaterina, Maria Petrova, and Ruben Enikolopov. 2020. "Political effects of the internet and social media." *Annual Review of Economics* 12:415–438.

# Appendix

## Formal Definition of Equilibrium & Technical Assumptions

**Definition A.1.** *A Perfect Bayesian Nash Equilibrium consists of a reporting strategy $m^*$ of an interest group, a voting strategy $a^*$ of a voter, and a belief $\mu^*$ satisfying:*

*(1) For all $m$, the voter chooses the best voting action $a^*$, which is defined as:*

$$a^*(y, m, c_y) = \begin{cases} L & \text{if } \mathbb{E}\left[U\left(a = L, y, \theta\right)|\mu^*\right] > \max\{\mathbb{E}\left[U\left(a = R, y, \theta\right)|\mu^*\right], c_y\} \\ R & \text{if } \mathbb{E}\left[U\left(a = R, y, \theta\right)|\mu^*\right] > \max\{\mathbb{E}\left[U\left(a = L, y, \theta\right)|\mu^*\right], c_y\} \\ 0 & \text{otherwise.} \end{cases} \tag{A.1}$$

*(2) Given (A.1), $m^*(z, \theta)$ is the news reporting strategy that maximizes the payoff of an interest group with ideology $z$.*

*(3) For all $m$, a voter updates her belief $\mu^*$ using Bayes' rule whenever possible.*

Because the games we analyze have multiple PBEs, we focus on the *Voter Welfare-Maximizing Perfect Bayesian Equilibrium (VWMPBE)*.[26] We will show that this coincides with the equilibrium in which most voters receive news that is informative to them.[27] News reporting is *informative* to a voter if there is a news item $m$ that is reported with positive probability on the equilibrium path such that $\mu(m) \neq p$, *i.e.* voters' beliefs about the state of the world are actually influenced by the message. Since we only allow for binary news reports, informativeness of news implies that $\mu(m = -1) \neq \mu(m = 1)$. For the sake of simplicity, we limit our attention to the VWMPBE in which $\mu^*(m = -1) \geq \mu^*(m = 1)$ for all voters and ignore essentially equivalent equilibria that are formed by permuting the messages $m = -1$ and $m = 1$. Hence, we focus on cases where receiving the message $m = -1$ makes more voters believe that the state of the world is actually $\theta = -1$ than if they receive the message $m = 1$ in equilibrium.

[26] This equilibrium also maximizes interest group payoffs. Since both voter welfare and interest group payoffs are maximized, we consider this the most 'sensible' equilibrium. Conventional equilibrium refinement criteria such as the Intuitive Criterion and the Divinity Criterion are not restrictive enough for our games. These criteria limit the number of equilibria by imposing some restrictions on the receivers' beliefs in the case that they receive out-of-equilibrium messages. In our games, we have a multitude of equilibria in which voters receive each message with a positive probability, such that there are no out-of-equilibrium messages.

[27] The Perfect Bayesian Equilibrium with the *lowest* voter welfare coincides with the *babbling* equilibrium, in which news is uninformative to each voter. As in all cheap talk games, all our four games have babbling equilibria. Voting actions and voter welfare are identical in the babbling equilibria of the four games.

We make two assumptions to eliminate multiple equilibria in which voting behavior is the same. First, if news are uninformative to a voter in equilibrium (*i.e.*, if $\mu^*(m=-1) = \mu^*(m=1) = p$) for *any* interest group ideology $z$, we assume that the interest group randomizes between $m=-1$ and $m=1$ with probability $p$. Second, if news are uninformative to a voter in equilibrium for ideology $z'$ but not for *all* $z \in [-h,h]$, we assume that the interest group with ideology $z'$ mimics the equilibrium reporting behavior of the ideologically closest interest group for which news reporting is informative.

**Technical Assumptions:** The following assumptions facilitate the equilibrium analysis.

A voter with belief $\mu'$ is *indifferent* between voting for parties $L$ and $R$ if she has ideological position

$$\hat{y}_{\{\mu'\}} = \frac{x_L + x_R}{2} + 2\mu' - 1, \tag{A.2}$$

which solves the equality $\mathbb{E}[U(a=L,y,\theta)\,|\mu'] = \mathbb{E}[U(a=R,y,\theta)\,|\mu']$.[28] It is *incentive compatible* for a voter $y$ with belief $\mu'$ to vote for party $L$ (party $R$) if she has a more left-wing (right-wing) ideology than the *indifferent voter*:

$$y < \hat{y}_{\{\mu'\}} \ (\mathbf{IC_L}), \tag{A.3}$$

and

$$y > \hat{y}_{\{\mu'\}} \ (\mathbf{IC_R}). \tag{A.4}$$

The ideological position of the indifferent voter with belief $\mu' \in [0,1]$, denoted by $\hat{y}_{\{\mu'\}}$, is only well-defined if $-b < \hat{y}_{\{\mu'\}} < b$. To ensure this, we make:

**Assumption A.1.** $-b+1 < x_L < x_R < b-1$.

This implies that each party is preferred by some voters in either state of the world.

A voter $y$ with belief $\mu'$ prefers voting for party $j$ over abstaining if her *participation constraint* is satisfied: That is, if

$$c_y < \mathbb{E}[U(a=j,y,\theta)\,|\mu'] \ (\mathbf{PC_j}). \tag{A.5}$$

We assume that the psychological gain from voting for a party with the voter's ideal policy position outweighs the mismatch cost, for all $y$, but is lower than the maximal voting cost, regardless of the voter's belief and the locations of party platforms:

**Assumption A.2.** $4tb^2 < g < \bar{c}$.

[28] Note that there can be multiple indifferent voters, as posterior beliefs might differ across voters.

Hence, ex ante each voter $y$ with belief $\mu \in [0,1]$ votes and abstains with positive probability. Since voting costs are independently drawn from a uniform distribution, Assumption A.2 allows to write the probability that a voter $y \in \left[-b, \hat{y}_{\{\mu'\}}\right)$ with belief $\mu'$ votes for party $L$ as $\mathbb{E}\left[U\left(a=L,y,\theta\right)|\mu'\right]/\bar{c}$. The probability that a voter $y \in \left(\hat{y}_{\{\mu'\}}, b\right]$ with belief $\mu'$ votes for party $R$ is given by $\mathbb{E}\left[U\left(a=L,y,\theta\right)|\mu'\right]/\bar{c}$. We denote the *turnout* for party $j$ by $\tau_j$ and define it as the expected share of all voters who vote for party $j$:

$$\tau_L = \int_0^1 \int_{-b}^{\hat{y}_{\{\mu'\}}} \mathbb{1}_{\{\mu_y=\mu'\}} \frac{\mathbb{E}\left[U\left(a=L,y,\theta\right)|\mu'\right]}{2b\bar{c}} dy d\mu', \tag{A.6}$$

$$\tau_R = \int_0^1 \int_{\hat{y}_{\{\mu'\}}}^{b} \mathbb{1}_{\{\mu_y=\mu'\}} \frac{\mathbb{E}\left[U\left(a=R,y,\theta\right)|\mu'\right]}{2b\bar{c}} dy d\mu', \tag{A.7}$$

where $\mu_y$ indicates the belief of voter $y$.[29] The probability that a voter $y$ turns out to vote for party $j$ is calculated for all possible voter beliefs, contributing only to total party turnout if the considered belief equals the voter's actual belief (*i.e.*, if $\mu_y = \mu'$). We can calculate party $L$'s *vote share* from the turnout for party $L$ and party $R$ as:

$$\nu_L = \frac{\tau_L}{\tau_L + \tau_R}. \tag{A.8}$$

It follows from (5) that the interest group is *indifferent* between a vote for party $L$ and $R$ at:

$$\hat{z}\left(\theta\right) = \frac{x_L + x_R}{2} - \theta, \tag{A.9}$$

which solves the equality $\left(x_L - x_n\right)^2 = \left(x_R - x_n\right)^2$. An interest group with ideological position $z$ favors party $L$ if $z \leq \hat{z}\left(\theta\right)$ and party $R$ if $z > \hat{z}\left(\theta\right)$.[30] Lastly, we assume that, ex ante, each party $L$ and $R$ has positive probability that the interest group favors it in both states:

**Assumption A.3.** $-h + 1 < x_L < x_R < h - 1$.

## Proof of Proposition 1

Before proving Proposition 1, we need to prove the following lemma.

**Lemma A.1.** *Party $L$'s vote share, $\nu_L$, increases in $\mu\left(\cdot\right)$ if all voters hold identical beliefs.*

*Proof of Lemma A.1.* Given that all voters hold belief $\mu'$, party $L$'s turnout is

$$\tau_L\left(a\right) = \int_{-b}^{\hat{y}_{\{\mu'\}}} \frac{\mathbb{E}\left[U\left(a=L,y,\theta\right)|\mu'\right]}{2b\bar{c}} dy. \tag{A.10}$$

[29] Janssen and Teteryatnikova (2017) use a similar formulation but only allow for identical voter beliefs.

[30] As a tie-breaker, we assume that the *indifferent interest group* acts as if it favors party $L$.

We use (A.2), integrate (A.10), and take its derivative with respect to $\mu$. This yields

$$\begin{aligned}\frac{d\tau_L}{d\mu} &= \frac{\mathbb{E}\left[U\left(a=L,\hat{y}_{\{\mu'\}},\theta\right)|\mu'\right]-t\left(x_L+b\right)^2+t\left(\hat{y}_{\{\mu'\}}-x_L\right)^2}{b\bar{c}} \\ &= \frac{g-t-2t\left(2\mu'-1\right)\left(\frac{x_L-x_R}{2}-2\mu'+1\right)-t\left(x_L+b\right)^2}{b\bar{c}},\end{aligned} \tag{A.11}$$

which is strictly positive if the following inequality holds:

$$g > t+2t\left(2\mu'-1\right)\left(\frac{x_L-x_R}{2}-2\mu'+1\right)+t\left(x_L+b\right)^2. \tag{A.12}$$

The right-hand side of (A.12) is maximized for

$$\mu' = \frac{4-x_R+x_L}{8}. \tag{A.13}$$

Plugging (A.13) into (A.12) gives

$$g > t+\frac{t\left(x_R-x_L\right)^2}{8}+t\left(x_L+b\right)^2, \tag{A.14}$$

which holds under Assumptions A.1 and A.2. Hence, $\tau_L\left(a\right)$ increases in $\mu\left(\cdot\right)$ for any $\mu'$.

It is analogous to show that $\tau_R\left(a\right)$ decreases in $\mu\left(\cdot\right)$ for any $\mu'$.

Since $d\nu_L\left(a\right)/d\tau_L\left(a\right) > 0$ and $d\nu_L\left(a\right)/d\tau_R\left(a\right) < 0$ (A.8), party $L$'s vote share $\nu_L\left(a\right)$ is increasing in $\mu\left(\cdot\right)$: $\frac{d\nu_L(a)}{d\mu} > 0$. □

*Proof of Proposition 1.* Each voter maximizes her expected utility, given her belief $\mu\left(m\right)$, by voting for party $j \in \{L,R\}$ if $IC_j$ ((A.3) and (A.4)) and $PC_j$ (A.5) hold, and by abstaining from voting otherwise (7). The interest group reports $m=-1$ if it favors party $L$ and $m=1$ if it favors party $R$ (see (A.9) and (8)). As depicted in Figure 1, it depends on the state of the world which party the interest group favors if $\hat{z}\left(\theta=1\right) < z \leq \hat{z}\left(\theta=-1\right)$. The group's favorite party is state-independent if $z \leq \hat{z}\left(\theta=1\right)$ or $z > \hat{z}\left(\theta=-1\right)$. Hence, following Bayes' rule, voters have the posterior beliefs specified in (9). Conditional on the interest group's ideology being $\hat{z}\left(\theta=1\right) < z \leq \hat{z}\left(\theta=-1\right)$, voters assign a higher probability to $\theta=-1$ if $m=-1$ than if $m=1$. As a result, the indifferent voter is located further to the right of the political spectrum for $m=-1$ than for $m=1$ (A.2) (see Figure 2). Lemma A.1 shows that $\nu_L\left(a\right)$ is increasing in $\hat{y}\left(\cdot\right)$, which implies that the interest group has no incentive to deviate from its reporting strategy if $\hat{z}\left(\theta=1\right) < z \leq \hat{z}\left(\theta=-1\right)$.

We can construct voters' out-of-equilibrium beliefs s.t. $\mu\left(m=-1, z \leq \hat{z}\left(\theta=1\right)\right) \leq p$ and $\mu\left(m=1, z > \hat{z}\left(\theta=-1\right)\right) \geq p$, which provides the interest group no incentive to deviate

for $z \leq \hat{z}(\theta = 1)$ and $z > \hat{z}(\theta = -1)$. Since $\mu^*(m = -1) > \mu^*(m = 1)$ for $\hat{z}(\theta = 1) < z \leq \hat{z}(\theta = -1)$, news reporting is informative in equilibrium. Thus, the game has an informative equilibrium in which the interest group follows the reporting strategy in (8) and voters choose their best voting actions and update their beliefs according to (7) and (9).

*Uniqueness:* Because we focus on equilibria in which $\mu^*(m = -1) \geq \mu^*(m = 1)$ for all voters, no other informative equilibrium exists. An interest group with ideology $\hat{z}(\theta = 1) < z \leq \hat{z}(\theta = -1)$ has an incentive to report $m = -1$ if $\theta = -1$ and $m = 1$ if $\theta = 1$ in any equilibrium in which $\mu(m = -1, \hat{z}(\theta = 1) < z \leq \hat{z}(\theta = -1)) > \mu(m = 1, \hat{z}(\theta = 1) < z \leq \hat{z}(\theta = -1))$. It then follows from Bayes' rule that beliefs are as specified in (9). An equilibrium with $\mu(m = -1, z) > \mu(m = 1, z)$ for $z \leq \hat{z}(\theta = 1)$ or $z > \hat{z}(\theta = -1)$ cannot exist, as it would induce the interest group to report the same news item in both states, such that voters' beliefs would be inconsistent with the interest group's reporting behavior.

*Voter-welfare maximizing* PBE: We need to show that voter welfare is higher in the informative equilibrium than in the babbling equilibrium (where $\mu^*(m = -1) = \mu^*(m = 1) = p$). Voter $y$'s expected payoffs in a babbling equilibrium are:

$$\mathbb{E}[U(a, y, \theta)|p] = \begin{cases} \frac{\mathbb{E}[U(a=L,y,\theta)|p]}{\bar{c}} \left( \mathbb{E}[U(a = L, y, \theta)|p] - \int_0^{\mathbb{E}[U(a=L,y,\theta)|p]} \frac{c_y}{\mathbb{E}[U(a=L,y,\theta)|p]} dc_y \right) \\ \text{if } y \leq y_{\{\mu=p\}} \\ \frac{\mathbb{E}[U(a=R,y,\theta)|p]}{\bar{c}} \left( \mathbb{E}[U(a = R, y, \theta)|p] - \int_0^{\mathbb{E}[U(a=R,y,\theta)|p]} \frac{c_y}{\mathbb{E}[U(a=R,y,\theta)|p]} dc_y \right) \\ \text{if } y > y_{\{\mu=p\}}, \end{cases} \tag{A.15}$$

where the first term (outside the brackets) is the probability that voter $y$ turns out to vote, the second term is her expected utility from voting for party $j \in \{L, R\}$, and the third term is her expected cost of voting, conditional on turning out to vote for party $j$.

Now, without loss of generality, consider the case that voter $y$ receives message $m = -1$. The gain in expected payoffs from moving from the babbling equilibrium to the informative equilibrium is as follows for a voter with ideology $y \leq \min\{x_L, \hat{y}_{\{p\}}\}$:

$$\begin{aligned} &\frac{\mathbb{E}[U(a = L, y, \theta)|p] - \mathbb{E}[U(a = L, y, \theta)|\mu(m = -1)]}{\bar{c}} \\ &\times \left( \int_{\mathbb{E}[U(a=L,y,\theta)|\mu(m=-1)]}^{\mathbb{E}[U(a=L,y,\theta)|p]} \frac{c_y}{\mathbb{E}[U(a = L, y, \theta)|p] - \mathbb{E}[U(a = L, y, \theta)|\mu(m = -1)]} dc_y \right. \\ &\left. - \mathbb{E}[U(a = L, y, \theta)|\mu(m = -1)] \right), \end{aligned} \tag{A.16}$$

which is strictly positive because $\mu(m = -1) > p$ and $U(a = L, y, \theta = -1) < U(a = L, y, \theta = 1)$

for $y \leq min\{x_L, \hat{y}_{\{p\}}\}$.

Correspondingly, the gain from news is as follows for a voter with ideology $x_L < y < \hat{y}_{\{p\}}$:

$$\frac{\mathbb{E}\left[U\left(a=L,y,\theta\right)|\mu\left(m=-1\right)\right]-\mathbb{E}\left[U\left(a=L,y,\theta\right)|p\right]}{\bar{c}}\times\left(\mathbb{E}\left[U\left(a=L,y,\theta\right)|\mu\left(m=-1\right)\right]\right.$$
$$\left.-\int_{\mathbb{E}[U(a=L,y)|p]}^{\mathbb{E}[U(a=L,y,\theta)|\mu(m=-1)]}\frac{c_y}{\mathbb{E}\left[U\left(a=L,y,\theta\right)|\mu\left(m=-1\right)\right]-\mathbb{E}\left[U\left(a=L,y\right)|p\right]}dc_y\right), \quad \text{(A.17)}$$

which is strictly positive because $\mu\left(m=-1\right)>p$ and $U\left(a=L,y,\theta=-1\right)>U\left(a=L,y,\theta=1\right)$ for $x_L < y < \hat{y}_{\{p\}}$.

The gain is as follows for a voter with ideology $\hat{y}_{\{p\}} < y \leq \hat{y}_{\{\mu_2\}}$:

$$\frac{\mathbb{E}\left[U\left(a=L,y,\theta\right)|\mu\left(m=-1\right)\right]}{\bar{c}}\left(\mathbb{E}\left[U\left(a=L,y,\theta\right)|\mu\left(m=-1\right)\right]\right.$$
$$\left.-\int_0^{\mathbb{E}[U(a=L,y,\theta)|\mu(m=-1)]}\frac{c_y}{\mathbb{E}\left[U\left(a=L,y,\theta\right)|\mu\left(m=-1\right)\right]}dc_y\right)$$
$$-\frac{\mathbb{E}\left[U\left(a=R,y,\theta\right)|p\right]}{\bar{c}}\left(\mathbb{E}\left[U\left(a=R,y,\theta\right)|\mu\left(m=-1\right)\right]-\right.$$
$$\left.\int_0^{\mathbb{E}[U(a=R,y,\theta)|p]}\frac{c_y}{\mathbb{E}\left[U\left(a=R,y,\theta\right)|p\right]}dc_y\right), \quad \text{(A.18)}$$

which is strictly positive because $\mu\left(m=-1\right)>p$ and $U\left(a=L,y,\theta=-1\right)>U\left(a=R,y,\theta=-1\right)$ for $\hat{y}_{\{p\}} < y \leq \hat{y}_{\{\mu_2\}}$.

The gain is as follows for a voter with ideology $\hat{y}_{\{\mu_2\}} < y < x_R$:

$$\frac{\mathbb{E}\left[U\left(a=R,y,\theta\right)|p\right]-\mathbb{E}\left[U\left(a=R,y,\theta\right)|\mu\left(m=-1\right)\right]}{\bar{c}}$$
$$\times\left(\int_{\mathbb{E}[U(a=R,y,\theta)|\mu(m=-1)]}^{\mathbb{E}[U(a=R,y,\theta)|p]}\frac{c_y}{\mathbb{E}\left[U\left(a=R,y,\theta\right)|p\right]-\mathbb{E}\left[U\left(a=R,y,\theta\right)|\mu\left(m=-1\right)\right]}dc_y\right.$$
$$\left.-\mathbb{E}\left[U\left(a=R,y,\theta\right)|\mu\left(m=-1\right)\right]\right), \quad \text{(A.19)}$$

which is strictly positive because $\mu\left(m=-1\right)>p$ and $U\left(a=R,y,\theta=-1\right)<U\left(a=R,y,\theta=1\right)$ for $\hat{y}_{\{\mu_2\}} < y < x_R$.

The gain is as follows for a voter with ideology $y > \max\{x_R, \hat{y}_{\{\mu_2\}}\}$:

$$\frac{\mathbb{E}\left[U\left(a=R,y,\theta\right)|\mu\left(m=-1\right)\right]-\mathbb{E}\left[U\left(a=R,y,\theta\right)|p\right]}{\bar{c}}\times\left(\mathbb{E}\left[U\left(a=R,y,\theta\right)|\mu\left(m=-1\right)\right]\right.$$
$$\left.-\int_{\mathbb{E}[U(a=R,y,\theta)|p]}^{\mathbb{E}[U(a=R,y,\theta)|\mu(m=-1)]}\frac{c_y}{\mathbb{E}\left[U\left(a=R,y,\theta\right)|\mu\left(m=-1\right)\right]-\mathbb{E}\left[U\left(a=R,y,\theta\right)|p\right]}dc_y\right), \tag{A.20}$$

which is strictly positive because $\mu\left(m=-1\right)>p$ and $U\left(a=R,y,\theta=-1\right)>U\left(a=R,y,\theta=1\right)$ for $y>\max\{x_R,\hat{y}_{\{\mu_2\}}\}$.

Hence, the informative equilibrium is the voter welfare-maximizing equilibrium. □

## Proof of Proposition 2

*Proof.* A voter's best voting action (10) and the interest group's reporting strategy (11) are the same as in the public game with voter awareness (see (7) and (8)). Under Assumption A.3, the interest group is more likely to favor party $L$ if $\theta=-1$ than if $\theta=1$, as is illustrated in Figure 1. For this reason, voters assign a higher probability to $\theta=-1$ if they receive news item $m=-1$ instead of $m=1$ (12). The ideological position of the indifferent voter lies further to the right of the spectrum if $m=-1$ than if $m=1$, which implies that party $L$ gets a larger vote share if $m=-1$ ((A.2) and Lemma A.1). Since the interest group only reports $m=-1$ if it favors party $L$ (11), no interest group type has an incentive to deviate. Thus, there is an equilibrium in which the interest group follows the strategy stated in (11) and the voters choose their voting action and update their beliefs according to (10) and (12).

The equilibrium is informative since $\mu*\left(m=-1\right)>p>\mu*\left(m=1\right)$ for each voter. Restricting ourselves to equilibria in which $\mu^*\left(m=-1\right)\geq\mu^*\left(m=1\right)$ for all voters, no other informative equilibrium exists for reasons analogous to the case of Proposition 1. This equilibrium is voter-welfare maximizing (see Proposition 1). □

## Proof of Proposition 3

*Proof.* As in the proof of Proposition 1, voters' best voting action (13) is the same as in the public games (see (7) and (10)). Reacting to that strategy, the interest group reports $m=-1$ if it favors party L and $m=1$ if it favors party R to voters with ideological position

$Y_1^U < y \leq Y_2^U$ (14),where

$$Y_1^U = min\{x_L, \hat{y}_{\{\mu_1\}}\}, \tag{A.21}$$
$$Y_2^U = max\{x_R, \hat{y}_{\{\mu_2\}}\}. \tag{A.22}$$

Assumption A.3 guarantees that the interest group is more likely to favor party $L$ if $\theta = -1$ than if $\theta = 1$, as depicted in Figure 1. Hence, voters with ideology $Y_1^U < y \leq Y_2^U$ assign a higher probability to $\theta = -1$ if $m_y = -1$ than if $m_y = 1$, according to Bayes' rule (15). The interest group engages in state-independent reporting to voters with ideology $y \leq Y_1^U$ or $y > Y_2^U$ (14), such that application of Bayes' rule induces these voters to ignore news (15). There is no incentive to deviate from its reporting behavior to voters with ideology $y \leq Y_1^U$ and $y > Y_2^U$ because both news items evoke the same beliefs and, hence, the same voting behavior among these voters. The interest group also cannot gain by changing its reporting behavior to the rest of the electorate. For voters with ideology $\hat{y}_{\{\mu_1\}} < y < \hat{y}_{\{\mu_2\}}$ it is, respectively, incentive compatible to vote for party $L$ if $m_y = -1$ and party $R$ if $m_y = 1$ (see (A.2),(A.3),(A.4), (13)), see Figure 3a. Thus, the interest group cannot increase its expected payoff by deviating from its reporting behavior for this group of voters. If $x_L < \hat{y}_{\{\mu_1\}}$, there are voters with ideology $x_L < y < \hat{y}_{\{\mu_1\}}$, for whom it is always incentive compatible to vote for party $L$ (see (A.2),(A.3),(15)), see Figure 3b. Their $PC_L$, however, is more likely to be satisfied for $m_y = -1$ than $m_y = 1$ ((2),(A.5),(15)). Since the interest group only reports $m_y = -1$ for $x_L < y < \hat{y}_{\{\mu_1\}}$ if party $L$ is its favorite party and $m_y = 1$ otherwise, the interest group cannot gain by changing its reporting behavior for these voters. In figure 3b, it is always incentive compatible for voters with $\hat{y}_{\{\mu_2\}} < y < x_R$ to vote for party $R$. For these, $PC_R$ is more likely to hold for $m_y = 1$ than $m_y = -1$ ((2),(A.5),(15)). Hence, the interest group can also not improve upon its reporting behavior to these voters ($m_y = -1$) if it favors party $L$ (and $m_y = 1$ if it favors party $R$ (14)). Therefore, the game has an informative equilibrium in which players follow the strategies described in (13) and (14) and voters update beliefs according to (15).

No more informative equilibrium exists. In any equilibrium with $\mu^*(m_y = -1) > \mu^*(m_y = 1)$ for $Y_1^U < y \leq Y_2^U$, the interest group maximizes its expected payoff by reporting as described in (14). It then follows from Bayes' rule that beliefs of voters with ideology $Y_1^U < y \leq Y_2^U$ are determined according to (15). News cannot be informative to voters with ideology $y \leq Y_1^U$ and $y > Y_2^U$. If $\mu(m_y = -1) > \mu(m_y = 1)$ for $y \leq Y_1^U$ or $y > Y_2^U$, the interest group would have an incentive to report $m_y = 1$ if $z \leq \hat{z}(\theta)$ and $m_y = -1$ if $z > \hat{z}(\theta)$. This implies that voters with ideology $y \leq Y_1^U$ and $y > Y_2^U$ would be *more* likely to receive $m_y = -1$ if $\theta = 1$ than if $\theta = -1$, which makes the reporting behavior

inconsistent with voters' beliefs.

Voter-welfare maximization: The expected payoff of voter $y$ in the babbling equilibrium (where $\mu^*(m=-1)=\mu^*(m=1)=p$) is given in (A.15). W.l.o.g., consider the case that voter $y$ receives message $m_y=-1$. Moving from the babbling equilibrium to the informative equilibrium, a voter's gain with ideology $Y_1^U < y \leq Y_2^U$ is given in equations (A.17)-(A.19) and is (weakly) positive. A voter with ideology $y \leq Y_1^U$ or $y > Y_2^U$ is exactly well-off with and without news because news is uninformative to her ($\mu(m_y=-1)=\mu(m_y=1)=p$ for $y \leq Y_1^U$ and $y > Y_2^U$). Since each voter $y$ is (weakly) better off, we conclude that the informative equilibrium we found is the voter welfare-maximizing equilibrium. □

## Proof of Proposition 4

*Proof.* The best voting action (see (16)) remains unchanged, as compared to the previously analyzed games (see (7), (10), and (13)). The interest group's reporting strategy (17) is the same as in the microtargeting game with voter unawareness (14). Given (17), it follows from Bayes' rule that voters' posterior beliefs are determined according to (18).

We need to show that nobody has an incentive to deviate from strategy profiles and beliefs as depicted in (16), (17), and (18). For voters with ideology $y \leq Y_1^A$ and $y > Y_2^A$, the interest group always induces the same posterior beliefs (18) regardless of the news it sends and, hence, has no incentive to deviate from its reporting behavior to these voters (17). In the proof of Proposition 3, we established that if voters with ideology $Y_1^A < y \leq Y_2^A$ assign a *higher* probability to $\theta=-1$, the interest group achieves a *higher* expected payoff for $z \leq \hat{z}(\theta)$ and a *lower* expected payoff for $z > \hat{z}(\theta)$. Thus, the interest group has no incentive to deviate from its reporting behavior if $\hat{z}(\theta=1) < z \leq \hat{z}(\theta=-1)$ ((17) and (18)). If we construct voters' out-of-equilibrium beliefs such that $\mu(m_y=-1, z \leq \hat{z}(\theta=1)) \leq p$ and $\mu(m_y=1, z > \hat{z}(\theta=-1)) \geq p$, the interest group also cannot gain from changing its equilibrium reporting behavior for $z \leq \hat{z}(\theta=1)$ and $z > \hat{z}(\theta=-1)$. Hence, there is an informative equilibrium with voters choosing voting actions as in (16) and updating their beliefs as in (18), and in which the interest group follows a strategy as in (17).

Analogous to the case of Proposition 3, no more informative equilibrium exists. As before, we consider the case that voter $y$ receives message $m_y=-1$, without loss of generality. The gain in expected payoffs for a voter with ideology $Y_1^A < y \leq Y_2^A$ from moving from the babbling equilibrium to the informative equilibrium is given in (A.17)-(A.19) and is (weakly) positive. A voter with ideology $y \leq Y_1^A$ or $y > Y_2^A$ is equally well-off with and without news because news is uninformative to her ($\mu(m_y=-1)=\mu(m_y=1)=p$ for $y \leq Y_1^A$ and $y > Y_2^A$). Since each voter $y$ has a (weakly) positive gain, the informative equilibrium we

found is the voter welfare-maximizing equilibrium. □

## Proof of Proposition 5

*Proof.* (1) Voters with ideology $Y_1^A < y \leq Y_2^A$ have exactly the same strictly positive payoff gains moving from the babbling equilibrium to the informative equilibrium in the public game with awareness and the microtargeting game with awareness ((A.15)-(A.20)). Voters with ideology $y \leq Y_1^A$ and $y > Y_2^A$, however, have strictly positive payoff gains from moving from the babbling equilibrium to the informative equilibrium in the public game with awareness but are equally well-off in the babbling equilibrium and the informative equilibrium in the microtargeting game with awareness ((A.15)-(A.20)).

Voters with $Y_1^U < y \leq Y_2^U$ have exactly the same strictly positive payoff gains moving from the babbling equilibrium to the informative equilibrium in the public game with unawareness and the microtargeting game with unawareness ( (A.15), (A.17)-(A.19)). Voters with $y \leq Y_1^U$ and $y > Y_2^U$, however, have strictly positive payoff gains if moving from the babbling equilibrium to the informative equilibrium in the public game with unawareness but are equally well-off in the babbling and the informative equilibria in the microtargeting game with unawareness ((A.15), (A.17)-(A.19)).

Thus, all voters yield weakly higher payoff and some voters have a strictly higher payoff in the public games than in the microtargeting games. Consequently, total welfare is strictly higher in the public games than in the microtargeting games (3).

(2) We first look at the gain in expected payoffs for voter $y$ from moving from the public game with voter unawareness to the public game with voter awareness. W.l.o.g., consider the case that voter $y$ receives message $m = -1$. If the interest group has $\hat{z}(\theta = 1) < z \leq \hat{z}(\theta = -1)$, the gain for voter $y$ is given in equations (A.16)-(A.20), with the exception that we need to substitute $p$ by $\mu^*(m = -1)$ (12) and $\mu(m = -1)$ by $\mu = 1$ (9). If an interest group has ideology $z \leq \hat{z}(\theta = 1)$ or $z > \hat{z}(\theta = -1)$, the gain for a voter with ideology $y \leq \min\{x_L, \hat{y}_{\{p\}}\}$, with $\mu(m = -1)$ given in (12), is:

$$\frac{\mathbb{E}\left[U\left(a = L, y, \theta\right)|p\right] - \mathbb{E}\left[U\left(a = L, y, \theta\right)|\mu\left(m = -1\right)\right]}{\bar{c}} \times \left( \mathbb{E}\left[U\left(a = L, y, \theta\right)|\mu\left(m = -1\right)\right] \right.$$
$$\left. - \int_{\mathbb{E}[U(a=L,y,\theta)|\mu(m=-1)]}^{\mathbb{E}[U(a=L,y,\theta)|p]} \frac{c_y}{\mathbb{E}\left[U\left(a = L, y, \theta\right)|p\right] - \mathbb{E}\left[U\left(a = L, y, \theta\right)|\mu\left(m = -1\right)\right]} dc_y \right), \quad \text{(A.23)}$$

which is strictly positive because $\mu(m = -1) > p$ and $U(a = L, y, \theta = -1) < U(a = L, y, \theta = 1)$ for $y \leq min\{x_L, \hat{y}_{\{p\}}\}$.

Analogous to this case, it is straightforward to show that a voter with ideology $y >$

$\min\{x_L, \hat{y}_{\{p\}}\}$ also gains from news.

Summarizing, the gain is strictly positive for each voter $y$ and any interest group ideology $z$. Consequently, total voter welfare is strictly higher in the public game with voter awareness than in the public game with voter unawareness (3).

Now we look at the expected payoff change for voter $y$ from *moving from the microtargeting game with voter unawareness* to the *microtargeting game with voter awareness*. For *stable moderate voters* (with ideology $Y_1^U < y \leq Y_2^U$), nothing changes as compared to the public games (see above). For *unstable moderate voters* (with ideology $Y_1^A < y \leq Y_1^U$ or $Y_2^U < y \leq Y_2^A$), the gain from awareness is given in equations (A.16) and (A.20), with $\mu(m=-1) = 1$ if $\hat{z}(\theta=1) < z \leq \hat{z}(\theta=-1)$ (18). If $z \leq \hat{z}(\theta=1)$ or $z > \hat{z}(\theta=-1)$, unstable moderate voters hold the same beliefs ((15) and (18)) and have, consequently, the same payoffs in both microtargeting games. Finally, *stable radical voters* (with ideology $y \leq Y_1^A$ or $y > Y_2^A$) hold the same beliefs for any $z$ ((15) and (18)) and, hence, get similar payoffs in both microtargeting games.

Thus, in microtargeting games the gain from awareness is (weakly) positive for all and strictly positive for some voters. Hence, total voter welfare is strictly higher in the microtargeting game with voter awareness than with voter unawareness. □

## Proof of Corollary 1

*Proof.* The welfare ranking follows directly from Proposition 5. □

## Proof of Proposition 6

We only consider cases where the electorate wants to elect party $L$ if $\theta = -1$ and party $R$ if $\theta = 1$. Formally, this restriction means that $\mathbb{E}[\tau_L|\mu=1] > \mathbb{E}[\tau_R|\mu=1]$ and $\mathbb{E}[\tau_L|\mu=0] < \mathbb{E}[\tau_R|\mu=0]$, where $\mathbb{E}[\tau_L|\mu=1]$ is the expected voter turnout for party $L$ given that voters are certain that $\theta = -1$, after having received messages. This restriction is met if the position of the indifferent voter lies on the right (left) side of the ideological spectrum (see (A.2), (A.6) (A.7)), which holds under the following assumption:

**Assumption A.4.** $-2 < x_L + x_R < 2$.

**Definition A.2.** *There is election flipping due to disinformation if one of the following two conditions holds:*

*(1)* $\mathbb{E}[\tau_L|p] > \mathbb{E}[\tau_R|p]$ *and* $\mathbb{E}[\tau_L|m^*(z,\theta), z] < \mathbb{E}[\tau_R|m^*(z,\theta), z]$ *if* $\theta = -1$.

*(2)* $\mathbb{E}[\tau_L|p] < \mathbb{E}[\tau_R|p]$ *and* $\mathbb{E}[\tau_L|m^*(z,\theta), z] > \mathbb{E}[\tau_R|m^*(z,\theta), z]$ *if* $\theta = 1$.

*Proof of Proposition 6.* Ex-ante probability of exposure to a malevolent interest group:

$$(1-p)\frac{x_L+x_R-2+2h}{4h}+p\frac{2h-x_L-x_R-2}{4h}. \tag{A.24}$$

This term reflects that only a *radical right* interest group (with ideology $z > \hat{z}(\theta=-1)$) would attempt to flip an election if $\theta=-1$, and only a *radical left* interest group (with ideology $z \leq \hat{z}(\theta=1)$) has an incentive to flip an election if $\theta=1$ ((A.9), Propositions 1-4).

(1) Public communication and voter unawareness: W.l.o.g. assume $\theta=-1$. Since $\tau_L-\tau_R$ increases in $\mu$ (Lemma A.1), Assumption A.4 implies that there is a belief $\hat{\mu}(g,t,x_L,x_R,b)$ for which turnout for party $L$ is equal to turnout for party $R$. Using (A.6) and (A.7) and plugging in $\hat{\mu}(g,t,x_L,x_R,b)$ for $\mu'$, we find that $\hat{\mu}(g,t,x_L,x_R,b)$ solves the following equation:

$$\begin{aligned} 2\frac{2}{3}\hat{\mu}^3(\cdot)-4\hat{\mu}^2(\cdot)+\left(\frac{g}{t}+1+b(x_R-x_L)-\frac{1}{2}\left(x_L^2+x_R^2\right)-b^2\right)\hat{\mu}(\cdot)= \\ \frac{g}{4t}(2-x_L-x_R)+(x_R-b-1)^3-(x_L+b-1)^3. \end{aligned} \tag{A.25}$$

Given that $\theta=-1$, an interest group would only have an incentive to manipulate an election by reporting $m=1$ if $z>\hat{z}(\theta=-1)$ ((5), (11)) and (A.9). An attempt to flip an election is successful if $\hat{\mu}(\cdot)<\mu(m=1)<p$. Using $\mu^*(m=1)$ (12), setting it equal to $\hat{\mu}(\cdot)$ and solving for $\hat{p}^{PU}$, we find that the interest group's election flipping potential is:

$$\hat{p}^{PU}=min\{\frac{(2(h+1)-x_L-x_R)\hat{\mu}(\cdot)}{2(h+2\hat{\mu}(\cdot)-1)-x_L-x_R},1\}, \tag{A.26}$$

which is strictly greater than $\hat{\mu}(\cdot)$ under Assumption A.3. Hence, there is a range of prior beliefs $p\in\left(\hat{\mu}(\cdot),\hat{p}^{PU}\right)$ for which the interest group can flip the election outcome with disinformation.

We now show that $\hat{p}^{MU}\geq\hat{p}^{PU}$. Suppose that $\hat{p}^{PU}<1$. Using Lemma A.1 and equations (5), (11), (A.6) and (A.7), we find the following inequality:

$$\frac{d\left(\mathbb{E}\left[\tau_R^{PU}|m^*(z,\theta),z>\hat{z}(\theta=-1)\right]-\mathbb{E}\left[\tau_L^{PU}|m^*(z,\theta),z>\hat{z}(\theta=-1)\right]\right)}{dp}<0. \tag{A.27}$$

If $\hat{p}^{PU}<1$, the following equality holds by definition:

$$\mathbb{E}\left[\tau_R^{PU}|m^*(z,\theta),z>\hat{z}(\theta=-1),p=\hat{p}^{PU}\right]=\mathbb{E}\left[\tau_L^{PU}|m^*(z,\theta),z>\hat{z}(\theta=-1),p=\hat{p}^{PU}\right]. \tag{A.28}$$

Using (12), (15), (A.6) and (A.7), we find that the following inequalities hold for any $p$:

$$\mathbb{E}\left[\tau_R^{MU}|m^*\left(z,\theta\right),z>\hat{z}\left(\theta=-1\right),p=\hat{p}^{PU}\right]>\mathbb{E}\left[\tau_R^{PU}|m^*\left(z,\theta\right),z>\hat{z}\left(\theta=-1\right),p=\hat{p}^{PU}\right], \tag{A.29}$$

$$\mathbb{E}\left[\tau_L^{MU}|m^*\left(z,\theta\right),z>\hat{z}\left(\theta=-1\right),p=\hat{p}^{PU}\right]<\mathbb{E}\left[\tau_L^{PU}|m^*\left(z,\theta\right),z>\hat{z}\left(\theta=-1\right),p=\hat{p}^{PU}\right]. \tag{A.30}$$

Hence, it follows from (A.29) and (A.30) that this inequality holds:

$$\mathbb{E}\left[\tau_R^{MU}|m^*\left(z,\theta\right),z>\hat{z}\left(\theta=-1\right),p=\hat{p}^{PU}\right]-\mathbb{E}\left[\tau_L^{MU}|m^*\left(z,\theta\right),z>\hat{z}\left(\theta=-1\right),p=\hat{p}^{PU}\right]>0, \tag{A.31}$$

which implies that $\hat{p}^{MU}>\hat{p}^{PU}$ if $\hat{p}^{PU}<1$.

If, instead, $\hat{p}^{PU}=1$, necessarily $\hat{p}^{PU}=\hat{p}^{MU}=1$, as $\hat{p}^{PU}$ and $\hat{p}^{MU}$ can only be $\leq 1$.

(2) Only an interest group with ideology $z>\hat{z}\left(\theta=-1\right)$ or $z\leq\hat{z}\left(\theta=1\right)$ might have an incentive to flip an election (see (5) and (A.9)). With voter awareness, voters hold on to their prior belief $p$ when they receive news from this interest group type (Propositions 1 and 4): following Definition 1, election flipping cannot occur. □

## Proof of Proposition 7

*Proof.* Consider first the VWMPBE of the public game with voter unawareness. Using (11) and (12) and applying Bayes' rule, voters' posterior belief is:

$$\mu^*\left(q,s\right)=\frac{\left(x_L+x_R+2h+2\right)^q\left(2h-2-x_L-x_R\right)^s p}{\left(x_L+x_R+2h+2\right)^q\left(2h-2-x_L-x_R\right)^s p+\left(x_L+x_R+2h-2\right)^q\left(2h+2-x_L-x_R\right)^s\left(1-p\right)} \tag{A.32}$$

Since (A.32) increases in $q$ (the number of messages $m=-1$) and decreases in $s$ (the number of messages $m=1$), an interest group has no incentive to deviate from its reporting behavior, regardless of the number of other interest groups. Using (A.1), a voter maximizes her expected utility by voting for party $j$ if $c_y<\mathbb{E}\left[U\left(a=j,y,\theta\right)|\mu^*\left(\cdot\right)\right]$. Hence, a voter *only* maximizes her expected utility if she has the posterior belief given in (A.32), which is the result of Bayesian updating and which takes *all* news items into account. It then follows that voter welfare increases with interest group competition (3).

The proof for the microtargeting game with voter unawareness is similar, with the exception that only moderate voters receive informative news and experience an increase in the expected payoff from voting due to interest group competition.

In the proofs of Proposition 1 and Proposition 4, we established that (moderate) voters only have a higher payoff in the informative equilibrium than in the babbling equilibrium if they face a moderate interest group. Using (A.9), we find that the expected number of moderate interest groups that a voter encounters is equal to $K/h$, which is increasing in $K$. Thus, all voters have a (weakly) higher payoff due to interest group competition in games with voter awareness. Hence, voter welfare increases with interest group competition (3). $\square$