

# Social Media and Democracy: Experimental Results\*

Freek van Gils<sup>†</sup>

Wieland Müller<sup>‡</sup>

Jens Prüfer<sup>§</sup>

## Abstract

Social media have become a main source of information for many voters. Political interest groups on social media platforms have the ability to (i) microtarget news based on individual-level voter data and (ii) obfuscate their identities, which can be exploited to spread disinformation. Two proposed interventions to prevent election manipulation by disinformation are a microtargeting ban and disclosure requirements. An empirical foundation for these interventions is missing. We experimentally study the effects of the implementation of a microtargeting ban and disclosure of interests in a social media environment on voting behavior. Our results show that mandatory disclosure of interests, in combination with or without a microtargeting ban, increases the efficiency of aggregate voter decision-making. However, only the combination of disclosure of interests and a microtargeting ban counteracts election manipulation. The implementation of a microtargeting ban without disclosure requirements has adverse effects.

**Keywords:** disinformation, social media, voting, microtargeting, disclosure of interests, laboratory experiments.

**JEL Codes:** C92, D72, D82, D83.

---

\*We are grateful to Tobias Klein, Boris van Leeuwen, Jan Potters, Jared Rubin, Martin Salm, Francesco Sobbrío, Sigrid Suetens, Mariya Teteryatnikova and seminar participants in Tilburg, Frankfurt, Passau, Lucerne, at the University of East Anglia, OFCOM, the TILEC Workshop on Economic Governance and Legitimacy, and the SIOE Conference (Toronto) for their comments. We thank Anna Osipenko for research assistance. Wieland Müller and Jens Prüfer gratefully acknowledge financial support from a research grant awarded by Data Science Center Tilburg.

<sup>†</sup>Netherlands Authority for Consumers and Markets (ACM), PO Box 16326, 2500 BH The Hague, the Netherlands; [freekvangils@outlook.com](mailto: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.

<sup>‡</sup>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](mailto:wieland.mueller@univie.ac.at).

<sup>§</sup>School of Economics and Centre for Competition Policy, University of East Anglia, and TILEC, Tilburg University; [j.prufer@uea.ac.uk](mailto:j.prufer@uea.ac.uk).

# 1 Introduction

In a democracy, political elections are the main institution legitimizing politicians to exercise power on behalf of the people. If politicians are perceived to rule legitimately, the government can save resources enforcing their policies by means of coercion because citizens accept those policies as rightful, even if they disagree with the specific contents, and protest less (Rubin, 2017, ch.2). Historically, empires tended to fall when their “subject people” did not support the rulers, anymore (Parsons, 2010).

Creating legitimacy for rulers by means of democratic elections requires that voters have sufficient and correct information about politically relevant events (*political news*). In the past few years, social media platforms have rapidly become dominant sources of political news. About a fifth of all U.S. adults and almost one half of those under thirty consume political news primarily through social media.<sup>1</sup> In most other countries, news consumption through social media is also substantial and rising.<sup>2</sup>

Using social media for the provision of political news leads to several fundamental problems, as we explain below. To improve our understanding of these problems and of the proposed policy interventions to tackle the problems, we study the effects of cheap-talk communication about payoff-relevant events from political interest groups to voters, both theoretically and experimentally. We do the latter within a (perfectly controlled) social media environment in the lab and contrast the results with alternative media environments that would arise if prominent policy proposals were introduced, as is already under way in the European Union: a ban on microtargeting techniques and mandatory disclosure requirements forcing social media platforms to make the original sender of political news more salient to

---

<sup>1</sup>See <https://www.journalism.org/2020/07/30/americans-who-mainly-get-their-news-on-social-media-are-less-engaged-less-knowledgeable/>.

<sup>2</sup>See [https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2022-06/Digital\\_News-Report\\_2022.pdf](https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2022-06/Digital_News-Report_2022.pdf). The European Parliament writes: “Political advertising is central to influencing how people vote, and may affect citizens’ perceptions of the legitimacy of their own political system, particularly when published in the run-up to elections. Rules governing political advertising are therefore key to guaranteeing citizens’ fundamental rights and the integrity of democratic processes.” ([https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/733592/EPRS\\_BRI\(2022\)733592\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/733592/EPRS_BRI(2022)733592_EN.pdf)).

voters.<sup>3</sup>

By using social media, voters generate lots of detailed information about their preferences and characteristics: so-called *user information* (Argenton and Prüfer, 2012). Platforms can infer a range of attributes from these data, most notably users' political views (Kosinski et al., 2013). A key characteristic of many news platforms, which is often at the heart of their business model, are microtargeted advertising services, which can be used by political interest groups to tailor news to preferences and characteristics of individual voters. *Microtargeting* allows political interest groups (or advertisers) to differentiate their news reports to influence voters' beliefs in their favor in each subgroup of the electorate.<sup>4</sup>

Next to microtargeting, a critical consequence of the use of algorithms by social media and other news platforms to draw messages from a large amount of political news providers is the *obfuscation of the original sender* of the news in the minds of voters: Kalogeropoulos and Newman (2017) find that 47% of digital news consumers recall the source of a news item that is accessed through social media. By contrast, the recall rate is 81% if the source is accessed directly. Kang et al. (2011) show that news portal users pay more attention to the identity of the portal than the original source when assessing the credibility of a news item.

Initially, scholars emphasized the pro-democratic role of social media. Social media have allowed new actors to reach voters with political information due to low barriers to entry, giving a voice to opponents of autocratic regimes (Zhuravskaya et al., 2020). In addition, social media have the potential to reach otherwise uninformed (and underrepresented) voter groups by means of news differentiation, i.e. microtargeting (Zuiderveen Borgesius et al., 2018). This optimistic view changed when actors with obscure motives exploited the low barriers to entry and microtargeting possibilities to spread disinformation among American

---

<sup>3</sup>See <https://www.europarl.europa.eu/news/en/press-room/20230123IPR68616/meps-toughen-rules-on-political-advertising>.

<sup>4</sup>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>).

voters, aiming to manipulate the outcome of the 2016 U.S. presidential election.<sup>5</sup> The low cost of creating automated social media accounts and the ability to post content using anonymous or impersonated accounts enable the manipulation of online content seen by users, which could lead to political persuasion (Zhuravskaya et al., 2020). The data that online platforms collect about their users could be (and have been) used to target specific groups to make such manipulations more effective.<sup>6</sup> Social media have also been blamed for creating an accommodating environment for antidemocratic forces (Tucker et al., 2017).<sup>7</sup>

Facing heavy public criticism, social media companies restricted political microtargeting and required senders of political messages to disclose their identity—initiatives that have been taken up in the political domain and have received strong public support as ways to ensure election integrity.<sup>8</sup> A solid empirical foundation for these interventions to avoid the manipulation of voter beliefs and election outcomes by disinformation is, however, missing.<sup>9</sup> It is hard to make valid causal claims about the impact of disinformation on voting behavior based on the (often highly aggregated) social media data available to researchers (Guess and Lyons, 2020), which complicates evaluations of the effectiveness of the implemented measures with observational data.<sup>10</sup>

To address these issues, we develop, theoretically analyze and experimentally test a series

---

<sup>5</sup>See for instance <https://www.washingtonpost.com/technology/2018/12/16/new-report-russian-disinformation-prepared-senate-shows-operations-scale-sweep/>.

<sup>6</sup>See, for instance, the New York Times story about Cambridge Analytica: <https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html>.

<sup>7</sup>In a meta study on digital media and political variables, covering 499 articles, Lorenz-Spreen et al. (2023) conclude: “declining political trust, advantages for populists, and growing polarization, are likely to be detrimental to democracy and were more pronounced in established democracies.”

<sup>8</sup>For instance, Google restricts targeting for election ads and Facebook requires that political ads are accompanied by a disclosure on who is paying for the advertising (<https://www.vox.com/recode/2020/9/29/21439824/online-digital-political-ads-facebook-google>). In a public consultation of the European Commission, 96% of the respondents expressed their support for disclosure rules and 83% of the respondents reported to be in favor of a restriction on political microtargeting ([https://ec.europa.eu/commission/presscorner/detail/en/IP\\_20\\_2250](https://ec.europa.eu/commission/presscorner/detail/en/IP_20_2250)).

<sup>9</sup>In their handbook on social media and democracy, Persily and Tucker (2020, p. 2) worry that “untested conventional wisdom based on folk theories of technology’s impact on democracy is leading to misguided reform proposals that may even worsen the problems they are attempting to solve.”

<sup>10</sup>“It remains the case that the employees of the platforms are the only ones who really know the scale of the problems widely attributed to them. Those of us on the outside must make do with the glimpses provided through publicly available data, which may or may not paint an accurate picture of what is actually going on.” Persily and Tucker (2020, p. xvii)

of games to study the effects of a ban on microtargeting and mandatory disclosure of interests within a social media environment on individual voting behavior and election outcomes. By controlling the factors influencing actual decision making of subjects in an experimental laboratory, we aim at a fundamental understanding of the forces and underlying mechanisms at play. This approach can both mimic the behavior of interest groups and voters, in a stylized and framing-free environment, and thereby “look” behind the curtain of proprietary data of social media firms, and also suggest causal relationships that are in line with the evidence, which are necessary to inform policy implications.

Our games comprise a politically motivated actor (“interest group”), who is informed about a binary state of the world and sends a cheap talk message concerning that state (the truth or disinformation) to uninformed voters, who subsequently make a voting decision.<sup>11</sup> The interest group and voters have one of two types, Majority (most common) or Minority (less common), which determines their payoffs in the different states of the world. In the first half of the experiment, the interest group and voters interact in a social media environment, which is characterized by two features: first, the interest group can report a message to the smaller Minority voter group, which may differ from the message directed to the larger Majority voter group. This element of the game mimics social media platforms’ microtargeted advertising services. Second, the type of the interest group is undisclosed (obfuscated) to voters. Voters only know that it is ex ante most probable that the interest group has type Majority. This game feature reflects that the sources of (political) information are much less salient and transparent on social media than in traditional media.

In the second half of the experiment, we then analyze the effects of two policy interventions—a ban on microtargeting and mandatory disclosure of the interest group’s type—and a combination of both. A *microtargeting ban* restricts the interest group to send the same (public) message to both voter groups, and voters know that everybody sees the same message.

---

<sup>11</sup>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 without lying/fake news.

Therefore, this intervention can also speak to the importance of an electorate’s shared beliefs about political events. The second intervention, *mandatory disclosure of interests*, reveals the interest group’s type to voters. As in the first half of the experiment, we let subjects interact repeatedly such that we can observe behavior that arises with experience. Our design allows to test our theoretical predictions about the effects of the interventions on voter decision-making after subjects have been locked into a social media environment, reflecting today’s informational status quo. We can then compare the outcomes in each of the three interventions with the outcomes of the benchmark treatment, the social media environment, in the second half of the experiment.

The experimental results show that mandatory disclosure of interests, with or without a microtargeting ban, increases the efficiency of aggregate voter decision-making. We find that voters are more likely to trust truthful messages and are less likely to give weight to false messages due to disclosure, which ultimately leads to more efficient voting actions. However, only the combination of disclosure of interests and a microtargeting ban counteracts election manipulation. A disclosure requirement on its own is insufficient to prevent interest groups from influencing the aggregate election outcome in their favor. A microtargeting ban even has adverse consequences if implemented in isolation because it reduces the power of an interest group to steer the election outcome on the basis of its (perfect) information about the state of the world. If the interest group is most likely to favor an efficient election outcome (as is mostly the case in our framework), this intervention is expected to reduce the efficiency of aggregate voter decision-making. Our experimental data confirm this prediction.

As robustness checks, we change the sequence of experimental games: in a new set of treatments, subjects start in an environment where microtargeting is impossible and where voters know the type of political interest group sending them a message. Only thereafter, in one treatment voters are exposed to microtargeting, in a second one the interest group type is obfuscated to voters, in a third treatment both changes occur simultaneously—and in a fourth treatment nothing changes. We find that the comparative statics confirm our main

results regarding the effects of microtargeting (or a ban thereof) and obfuscation of senders' interests (or disclosure thereof).

Finally, we let subjects start in an environment with only microtargeting or, respectively, with undisclosed interest group types, and then change the environments. We find that the results do not depend on whether we employ a between-subject or a within-subject design. We conclude that our main results regarding the implementation of a microtargeting ban and disclosure of interest reported above are highly robust.

Our experimental research approach has a number of advantages compared to research based on observational data, when studying voting behavior in different media environments. Firstly, a laboratory setting allows us to cleanly test our theoretical predictions under the exact conditions postulated in the assumptions underlying the model. Specifically, we control payoffs, communication modes, and the information sets of the interest group and voters. Secondly, we precisely observe the information voters are exposed to. Outside of the lab, it is impossible to observe all sources of information that voters use to make their voting decision. Thirdly, we can directly elicit subjects' beliefs, which allows us to study how communication affects voters' beliefs and how voting actions are driven by voter beliefs.

The paper is organized as follows. Section 2 reviews related literature. Section 3 describes the theoretical framework as well as the experimental design and procedures. Section 4 presents the experimental results. Section 5 discusses key design choices, while Section 6 concludes with policy implications. An (online) appendix contains a formal solution of our games and a comprehensive list of robustness checks, next to experimental instructions and background data.

## 2 Related literature

To the best of our knowledge, there is one other experimental paper on social media and voting.<sup>12</sup> [Pogorelskiy and Shum \(2019\)](#) report a laboratory experiment in which voters receive (biased) payoff-relevant news and decide whether or not to share it with other voters in their (partisan or complete) network before they cast a vote. The authors find that media bias always reduces the efficiency of collective decision-making. The ability to share news with *all* voters in the network (as opposed to solely voters belonging to the same partisan group) only enhances efficiency in the absence of media bias. The setting that [Pogorelskiy and Shum \(2019\)](#) consider is fundamentally different from ours. They consider voters’ strategic decision to share ‘hard’ (verifiable) news generated by a non-strategic media outlet in a network. In our games, the information provider is an interest group that can deliberately inform or misinform voters, who do not communicate with each other. Furthermore, in the experiment by [Pogorelskiy and Shum \(2019\)](#) voters know each others’ types and either share news with their entire network or not at all, whereas microtargeted news reporting and undisclosed ideological types are the key features of our experiment.

[Ziegler \(2023\)](#) does not focus on a social media environment explicitly but is related to our paper because he uses a lab experiment to test whether public or private messages are more persuasive for receivers and how this depends on the audience’s strategic environment. Specifically, he distinguishes between one environment, where a receiver’s incentive to choose an action increases in the number of other receivers choosing that action (“strategic complementarities”) and another one where the incentives decrease in the numbers of others choosing that action (“strategic substitutes”). Thereby, he implements a (mis)coordination problem among receivers. By contrast, in our games the payoffs of every receiver depend on the state of the world, the receiver’s voting action, and the receiver’s type. They are not interdependent. In our environment, as it turns out, receivers (especially Minority receivers)

---

<sup>12</sup>[Kartal and Tyran \(2022\)](#) experimentally study the effect of overconfidence and misinformation on information aggregation in elections but do not consider social media environments. [Sun et al. \(2021\)](#) analyze the effects of (traditional) media bias on voting behavior in a laboratory experiment.

use public communications technology to evaluate the truthfulness of a sender’s cheap talk message.

Our theoretical framework relates to a game of cheap talk communication with two groups of receivers by [Farrell and Gibbons \(1989\)](#). An informed sender reports one of two (public or private) messages about a binary state of the world to two uninformed receivers, who subsequently choose one of two actions. As is standard in cheap talk games, the payoff of the sender depends on the actions of both receivers and the true state, and the payoff of each receiver depends on the own action and the true state. Our framework differs in two ways from this cheap talk game. First, in all our games receivers can always guarantee themselves an intermediate payoff by choosing to abstain from voting, which yields the same payoff in both states of the world. Second, the type of the sender is undisclosed in two of our games.

Our study is also related to two cheap talk experiments in which the receiver is uncertain about the type of the sender. [Chung and Harbaugh \(2019\)](#) find mild evidence that a receiver benefits from disclosure of the sender’s bias. With disclosure, a receiver discounts the message of a biased sender. Without disclosure, a biased sender is more likely to lie and a receiver insufficiently discounts the message. These experimental findings are in contrast with the results of [Cain et al. \(2005\)](#), who find that receivers insufficiently discount messages of senders with a disclosed bias. Moreover, they find that a sender lies more in case of disclosure, which might be explained by strategic exaggeration or a feeling to be “morally licensed” to lie because of disclosure ([Cain et al., 2005](#)). Whereas [Chung and Harbaugh \(2019\)](#) and [Cain et al. \(2005\)](#) consider a set-up with one receiver, we study (public or microtargeted) communication to multiple receivers.

Lastly, our paper relates to an empirical literature on social media and voting behavior.<sup>13</sup> [Bond et al. \(2012\)](#) and [Jones et al. \(2017\)](#) show in a large-scale field experiment during the 2010 U.S. congressional elections that political mobilization messages on Facebook had a positive effect on voter turnout. [Liberini et al. \(2020\)](#) demonstrate that microtargeted polit-

---

<sup>13</sup>See [Zhuravskaya et al. \(2020\)](#) and [Persily and Tucker \(2020\)](#) for literature reviews about social media and political outcomes.

ical advertising on Facebook prior to the 2016 U.S. presidential elections persuaded moderate voters to vote for Trump, mobilized Republican supporters and demobilized Democratic supporters from turning out to vote. In a field experiment conducted in the weeks before the 2020 U.S. presidential elections, [Beknazar-Yuzbashev and Stalinski \(2022\)](#) find that political ads had an insignificant effect on average voter turnout but a substantial negative effect on turnout by Republicans. [Fujiwara et al. \(2021\)](#) and [Rotesi \(2019\)](#) exploit exogenous variation in regional Twitter adoption to study the effect of social media on voting behavior. [Fujiwara et al. \(2021\)](#) find that exposure to Twitter negatively affected the Republican vote share in the 2016 U.S. presidential elections, whereas the results of [Rotesi \(2019\)](#) suggest that Twitter exposure lowered the Democratic vote share in the 2012 U.S. presidential elections. [Allcott and Gentzkow \(2017\)](#) and [Guess et al. \(2020\)](#) look at the diffusion of misinformation on social media prior to the 2016 U.S. presidential elections and argue that it cannot have been decisive for the election outcome.

### 3 Theory, experimental design and procedures

#### 3.1 The four games

We consider a set-up with an electorate that consists of multiple voter groups with different sizes and interests.<sup>14</sup> Voters can vote for one of two parties, *party X* or *party Y*, or *abstain* from voting. Voting implies a payoff that depends on the state of the world, which is unobservable to voters. Prior to voting, the voters receive a message about the state from an

---

<sup>14</sup>[Van Gils et al. \(2020\)](#) discuss a related model, which mainly differs from this set-up in how voters and interest groups are modeled. Whereas [Van Gils et al. \(2020\)](#) consider a continuum of voter and interest group types which are located on a left-right ideological spectrum, we only allow for two interest group and voter types to keep the experimental games as simple as possible. In addition, in [Van Gils et al. \(2020\)](#) the ideal policy positions of the interest group and the voters depend on a binary state of the world which *objectively* favors either more left-wing or more right-wing policy. In our framework, the state of the world is interpreted *subjectively*: the interest group and voters always favor opposite actions in a given state. Despite these differences in the set-up of the models, the two frameworks generate the same main insights.

interest group that observes the true state.<sup>15</sup> The interest group shares the same interests as one of the voter groups. Its message is cheap talk (costless, non-binding and non-verifiable) and may be used to strategically inform or misinform voters. All these elements of the set-up are common knowledge.

For the sake of simplicity, our experimental games only contain two voter groups (*Majority* and *Minority*), one interest group (type *Majority* or *Minority*) and a binary state of the world (*One* or *Two*). The Majority voter group captures the ‘mainstream’ of the electorate and comprises two voters. The Minority group represents a smaller, ‘niche’ voter group and consists of one voter.<sup>16</sup> In line with the sizes of the respective voter groups, the interest group has type Majority with probability  $2/3$  and type Minority with probability  $1/3$ . The payoffs of the interest group and voters are displayed in Table 1a and Table 1b. As can be seen from Table 1a, the lowest and intermediate payoffs for the interest group are, respectively,  $1/3$  and  $2/3$  of the highest interest group payoff. For voters, the intermediate payoff is  $2/3$  of the highest voter payoff (Table 1b). The lowest payoff is slightly less than  $1/3$  of the highest payoff to induce voters who maximize their expected payoff to abstain in equilibria in which a message is uninformative (i.e., when a voter’s posterior belief equals the prior belief).<sup>17</sup>

We analyze environments that differ in terms of the interest group-voter communication mode—messages can be the same for all voters (**P**ublic) or tailored to voter types (**M**icrotargeted)—and the transparency regime, which concerns the type of the interest group that can be **D**isclosed or **U**ndisclosed to voters. Combining the two dimensions, we have four different games—PD, PU, MD and MU.

The timing of each game is as follows:

Stage 0: Nature determines the state (One or Two, both with probability  $1/2$ ) and the interest group type (Majority, with probability  $2/3$ , or Minority, with probability  $1/3$ ). The

---

<sup>15</sup>Interest groups usually have access to expert knowledge and resources to discover the true state. See Shapiro (2016) and Kartal and Tremewan (2018) for a more detailed justification of the assumption that the interest group is perfectly informed about the true state of the world.

<sup>16</sup>Voting game experiments with three voters are quite common. See Großer (2020) for an overview.

<sup>17</sup>If the lowest payoff would be equal to  $1/3$  of the highest payoff, every voting action would be possible in an equilibrium with uninformative messages.

Table 1: Payoffs

(a) Interest group payoffs (for each vote)				(b) Voter payoffs			
Type	Vote	State		Type	Vote	State	
		One	Two			One	Two
Majority	X	30	10	Majority	X	90	27
	Y	10	30		Y	27	90
	A	20	20		A	60	60
Minority	X	10	30	Minority	X	27	90
	Y	30	10		Y	90	27
	A	20	20		A	60	60

*Notes.* The payoff that an interest group receives from the interaction with a voter depends on its type (Majority or Minority), the voting action (party X (X), party Y (Y) or abstain (A)) and the state of the world (One or Two). The payoff that voters receive depends on their type (Majority or Minority), voting action (party X (X), party Y (Y) or abstain (A)) and the state of the world (One or Two).

interest group observes the state and its own type. Voters learn the interest group type in Disclosure games only.

Stage 1: The interest group selects one message (One or Two) for *all* voters in Public games or a separate message for *each* voter type in Microtargeting games.

Stage 2: Voters observe their own message and make a voting decision (party X, party Y or abstain (A)). All payoffs are realized as specified in Table 1a and Table 1b. The party that receives a majority of the votes wins the election.<sup>18</sup> There is no election winner if neither party obtains a majority of the votes.

The solution concept for our games is Perfect Bayesian Equilibrium (Fudenberg and Tirole, 1991). As is conventional in the cheap talk literature, we focus on the most informative equilibrium, in which the most information is transmitted from the interest group to voters.<sup>19</sup> Before we informally describe the equilibria of the games, we first introduce some

<sup>18</sup>Note that the payoffs of the interest group and voters only depend on *individual* voting actions and not on the *aggregate* election outcome, which is in line with the expressive voting theory (e.g. Schuessler, 2000). Political parties are not active players and do not receive any payoffs.

<sup>19</sup>All cheap talk games also have a completely uninformative ‘babbling’ equilibrium.

measures. The formal statement of the most informative equilibria of the four games and the corresponding proofs are presented in Appendix [A.1-A.5](#).

### 3.2 Voting efficiency and communication measures

Our primary interest is the *efficiency* of voter decision-making, both for individual voters and for the electorate as a whole. On the individual level, we consider voting behavior *efficient* (*inefficient*) if it yields the maximal (minimal) voter payoff. Abstention from voting (voting action A) is neither an efficient nor an inefficient vote. On the aggregate level, we consider an election outcome *efficient* (*inefficient*) if the party favored by a majority of the voters receives the majority (minority) of the votes. Specifically, an election outcome is efficient (inefficient) if party X wins (loses) in State One or if party Y wins (loses) in State Two. A *tie* (both parties receive the same number of votes) is neither efficient nor inefficient.

The efficiency of voting behavior is the result of the communication between the interest group and the voters. Our communication measure for the interest group is *truth*, which takes value one if a message is equal to the true state and value zero if the interest group reports the opposite of the true state. The communication measure for voters, *trust*, is equal to one if a voting action yields the maximal voter payoff under the assumption that the message received is truthful, and zero otherwise. This measure only captures whether a voter believes that the message from the interest group is correct if voters choose the own expected payoff-maximizing action (i.e., if voters do not have strong other-regarding preferences). Our data from an additional task of the experiment suggest that this implicit assumption is valid. Elicited voter beliefs are consistent with own payoff-maximizing actions in roughly 95% of the cases.

### 3.3 Equilibria

Let us first consider the equilibria of the Disclosure (D) games. As can be seen from Table [1a](#) and Table [1b](#), the incentives of a Majority interest group are aligned with the incentives

of Majority voters and misaligned with the incentives of the Minority voter. For a Minority interest group, the opposite holds true. For the **MD-game**, in which there is Microtargeted communication (M), these payoff structures imply in equilibrium that a message from a Majority (Minority) interest group is truthful and induces trust on the side of a Majority (Minority) voter. If the types of the interest group and voter do not match, the message is unrelated to the true state and the voting action is unrelated to the message (truth and trust with probability  $1/2$ , conditional on non-abstention).<sup>20</sup>

In the **PD-game**, the Disclosure game with Public communication (P), the (mis)alignment between the incentives of the interest group and the Majority voters determines whether there is truthful communication and trust in equilibrium because the Majority interest group is disciplined to be truthful to a Minority voter and a Minority interest group can no longer be consistently truthful to a Minority voter (truth with probability  $1/2$ ) due to the requirement that messages are the same for all voters.<sup>21</sup> Consequently, *all* voters trust a message from a Majority interest group and only trust a message from a Minority interest group with probability  $1/2$ , conditional on non-abstention.

Moving to the games without Disclosure (U), voters can no longer recognize the type of the interest group. In the equilibrium of the **MU-game**, Majority voters receive a truthful message from a Majority interest group type and a lie from a Minority interest group type. Majority voters trust their message, conditional on non-abstention, because the interest group is most likely to have the Majority type. Information transmission to Minority voters breaks down: messages are unrelated to the true state and voting actions are unrelated to the messages received (truth and trust with probability  $1/2$ , conditional on non-abstention).<sup>22</sup>

In the **PU-game**, a Majority interest group is disciplined to be truthful to all voters

---

<sup>20</sup>In the most informative equilibrium, an expected payoff-maximizing voter always abstains if a message is unrelated to the true state.

<sup>21</sup>Farrell and Gibbons (1989) call these cases, respectively, one-sided discipline and subversion.

<sup>22</sup>Suppose that a Minority voter trusts a message with a probability higher than  $1/2$ . A Majority interest group would best respond by reporting the opposite of the true state. Since an interest group has the Majority type with probability  $2/3$ , the Minority voter would be worse off than by ignoring the message, which cannot be the case in equilibrium. For an analogous reason, no equilibrium exists in which a Minority voter trusts a message with a probability lower than  $1/2$ .

and a Minority interest group lies to all voters. All voters trust their message, conditional on non-abstention, because they are most likely facing an interest group with the Majority type.

### 3.4 Predictions

We study whether the efficiency of voting actions in a social media environment with micro-targeted communication and undisclosed interest group types (MU-game) can be improved by implementing a microtargeting ban (PU-game), mandatory disclosure of interests (MD-game) or a combination of the two interventions (PD-game). Our predictions regarding the effects of these interventions follow directly from the most informative equilibria derived in the previous section.

**Microtargeting ban.** As the largest voter group, Majority voters are expected to always receive messages tailored to them, regardless of the communication technology in place. Consequently, a microtargeting ban should not affect the efficiency of their voting actions. Minority voters, on the other hand, are expected to benefit from a *discipline effect*: the ban forces an interest group to convey the same information to Minority voters as to Majority voters. Since it is more likely that an interest group has type Majority, this should increase Minority voters' payoffs from voting: there are two effects affecting the Minority voter, which work in opposite directions. On the one hand, a Minority interest group will now consistently tell a lie to the Minority voter, which is an unwanted by-product of the lie that the Minority interest group tells to the Majority voters. On the other hand, a Majority interest group will now tell the truth to the Minority voter, which is the unwanted by-product of the truth that the Majority interest group tells to the Majority voters. Since it is more likely that the interest group has type Majority, the message sent to a Minority voter is informative (in expectation) and should increase a Minority voter's payoff from voting.

Despite the fact that the microtargeting ban without disclosure of interests is expected to (weakly) improve efficiency of voters on the individual level, we predict that this intervention

has a negative effect on the aggregate level. Due to the discipline effect, the ban is expected to provide Minority voters with better information about the state of the world. Since Minority voters favor a party that is disliked by most voters, this makes it less likely that aggregate voter decision-making is efficient. Hypothesis 1 summarizes these predictions.

**Hypothesis 1** (Microtargeting ban).

- a. ***Voter payoffs.*** *A microtargeting ban increases the payoffs from voting for Minority voters but has no effect for Majority voters.*
- b. ***Election outcome.*** *A microtargeting ban decreases the efficiency of aggregate voter decision-making.*

**Mandatory disclosure of interests.** This intervention is expected to be beneficial for both voter groups because it prevents them from giving weight to messages sent by interest groups with misaligned incentives and avoids discounting of reliable messages. The aggregate effect is ambiguous. While disclosure is expected to stop the manipulation of Majority voters' beliefs by a Minority interest group, it is also expected to make information transmission to the Minority voter possible. Jointly, these two effects may enhance or decrease the efficiency of aggregate voter decision-making.

**Hypothesis 2** (Mandatory disclosure of interests).

- a. ***Voter payoffs.*** *Mandatory disclosure of interests increases the payoffs from voting for Majority voters and Minority voters.*
- b. ***Election outcome.*** *Mandatory disclosure of interests has an ambiguous effect on the efficiency of aggregate voter decision-making.*

**Microtargeting ban and mandatory disclosure of interests.** Due to mandatory disclosure of interests, Majority voters can accurately assess whether a message is trustworthy or not. The microtargeting ban has no added value for them. Minority voters benefit from

the combination of the two interventions: they receive more informative messages due to the microtargeting ban and are better able to recognize reliable messages due to disclosure of interests.

According to our theory, only a combination of the two interventions under study has an unambiguously positive expected effect on the efficiency of aggregate voter decision-making. A Minority interest group is unable to convey any *false* information to Majority voters due to mandatory disclosure of interests and any (consistently) *truthful* information to a Minority voter due to the microtargeting ban, which mitigates the risk of electing the on aggregate inefficient political party. In contrast, a Majority interest group is in equilibrium able to resolve all uncertainty about the state of the world and can thereby ensure an efficient election outcome.

**Hypothesis 3** (Microtargeting ban and mandatory disclosure of interests).

- a. **Voter payoffs.** Mandatory disclosure of interests in combination with a microtargeting ban increases the payoffs from voting for Majority voters and Minority voters.*
- b. **Election outcome.** Mandatory disclosure of interests in combination with a microtargeting ban increases the efficiency of aggregate voter decision-making.*

### 3.5 Experimental design

We implement the four games in a laboratory experiment using the parameter values specified in Section 3.1. Table 2 provides an overview of our treatments. Each session of each treatment contains twelve subjects and consists of two parts. Subjects first play forty rounds of one game (Part I) and then forty rounds of another game, as displayed in Table 2. The roles of all subjects are randomly determined for blocks of five consecutive rounds. After five rounds, new roles are assigned to all subjects that remain fixed for another block of five rounds. In every block of twenty rounds of a part of the experiment, subjects act five rounds in the role of a (Majority or Minority) interest group, ten rounds in the role of Majority

voter and five rounds as a Minority voter.<sup>23</sup> At the beginning of each round, subjects are randomly assigned to groups of four, containing one interest group, two Majority voters and one Minority voter. In every first and fifth round of a block of five rounds, we directly elicit voter beliefs about the true state of the world. Besides, in these rounds we also ask subjects in the role of interest group to predict the voting actions and subjects in the voter roles to predict the type of the interest group (if it was undisclosed). After Part II, subjects complete a lying aversion task, risk elicitation task and a survey.

To check robustness of our results, we study our research questions both from a within-subject and a between-subject perspective. The main goal is to study whether, starting from the social media environment, the implementation of (i) a microtargeting ban, (ii) mandatory disclosure of interests and (iii) a combination of both interventions improve the efficiency of voter decision-making. For this purpose, we use the first four treatments listed in Table 2 and report their results in Section 4.2. Subjects are first locked into the social media environment (MU), after which one of the three interventions (MD, PU or PD) is implemented or the status quo is maintained (MU).

Next, we study the effects when we start from a traditional media environment, where microtargeting is not possible and voters know the interest group’s objectives, and then move to a game with microtargeting or obfuscated objectives or both (treatments 5-8 in Table 2). Additionally, we use treatments 9-12 in Table 2 to check for experimental order effects because the game played in Part I could set focal points or expectations and thereby influence subjects’ behavior in Part II. To economize on space, we report on these results in Appendices A.6 to A.8. The main findings are highly robust: independent of whether we use a within-subject or a between-subject design, and independent of the specific game we let subjects start with, the effects of microtargeting (or a ban thereof) and of obfuscation (or disclosure) of interest group types always have the same effects as reported in Section 4.2.

---

<sup>23</sup>In the experiment, we use neutral wording to avoid framing. For instance, an interest group is called an A-player and a voter is called a B-player.

Table 2: Treatment overview

Treat- ment	Part I	Part II	Description	Number of sessions	Number of subjects
1	MU	MD	Mandatory disclosure	3	36
2	MU	PU	Microtargeting ban	3	36
3	MU	PD	Mandatory disclosure and microtargeting ban	3	36
4	MU	MU	Status quo microtargeting and undisclosed interests	3	36
5	PD	PU	Obfuscation of interests	3	36
6	PD	MD	Microtargeting implementation	3	36
7	PD	MU	Obfuscation of interests and microtargeting implementation	3	36
8	PD	PD	Status quo public communication and disclosed interests	3	36
9	MD	MU		3	36
10	MD	PD		3	36
11	PU	MU		3	36
12	PU	PD		3	36
			Total	36	432

### 3.6 Experimental procedures

The experiment was programmed in zTree ([Fischbacher, 2007](#)) and was conducted at the laboratory of the Vienna Center for Experimental Economics at the University of Vienna between October 2020 and July 2021. We organized 36 sessions in which 432 subjects participated. Subjects were recruited through the online recruitment system ORSEE ([Greiner, 2015](#)) and took part in only one session. Each session lasted about two and a half hours. Prior to the start of Part I of the experiment, subjects were informed that the experiment would consist of two parts, received written instructions for Part I and had to answer a series of tutorial questions correctly. After Part I, subjects received instructions for Part II and had to answer some additional tutorial questions correctly before Part II of the experiment started. At the end of the experiment, subjects were asked to complete a lying aversion task, a risk elicitation task and a survey. Detailed instructions for one of our treatments (MU-PU) can be found in [Appendix A.13](#). Subject earnings were 42.15 euros on average.

## 4 Experimental Analysis

### 4.1 Empirical strategy

Our aim is to analyze the effects of the three interventions (a microtargeting ban, mandatory disclosure of interests or a combination of the two measures) compared to the status quo (MU-game) *after* subjects have been locked into the social media environment. That is why we here only use data from the first four treatments listed in Table 2, excluding all treatments that do not start with an MU-game from our analysis. We restrict the analysis to the second half (rounds 21-40) of all games to purge the data of learning effects.<sup>24</sup> We estimate a random effects model for each of the measures explained in Subsection 3.2. This approach accounts for the dependency in our data which arises because we have repeated observations on each of 144 subjects.<sup>25</sup> The estimated models have the following form.<sup>26</sup>

$$y_{ith} = \alpha + \beta_1 PU + \beta_2 MD + \beta_3 PD + \beta_4 Part + u_i + \epsilon_{ith}, \quad (1)$$
$$i = 1, \dots, 144, \quad t = 21, \dots, 40, \quad h = 1, 2.$$

---

<sup>24</sup>Behavior in the first half is more volatile. Qualitatively, however, our core results are the same if we include all rounds.

<sup>25</sup>The subject-level random effect models do not (fully) account for potential session effects. We have, however, aimed to minimize session effects by the design of our experiment. To limit static session effects, we have randomized the order of our sessions and made sessions as homogeneous as possible. To minimize the risk of dynamic session effects, we have implemented random matching. Subjects were extensively informed about random matching and could only proceed to the experiment if they answered a test question about random matching correctly. Addressing potential session effects in the design instead of in the model estimation is advocated by Kim (2020), who studies experimental designs in which subjects make repeated decisions within the same session. As a further precaution to limit the risk of dynamic session effects distorting our results, we have restricted our analysis to the second block of twenty rounds, in which behavior is settled more than in the first twenty rounds. In the post-experimental questionnaire, we have asked respondents about their strategies in each of the experimental roles. An overwhelming majority of subjects describes a ‘fixed’ decision rule. There were only a few subjects who stated to follow a dynamic decision rule. Hence, we have no reason to believe that there were large dynamic session effects. This observation seems to be in line with experimental literature. In a study of static and dynamic session effects in laboratory experiments, Fréchette (2012) concludes that it is difficult to completely rule out dynamic session effects but that it is also hard to find many situations in which dynamic session effects seem to be enormous.

<sup>26</sup>For ease of interpretation, we present a linear probability model. We have also estimated subject-level random effects (ordered) probit models, which produce very similar estimates and gives rise to the same qualitative results.

In (1),  $y_{ith}$  is the outcome of one of our voting efficiency and communication measures in round  $t$  of part  $h$  for subject  $i$ .<sup>27</sup> The intercept  $\alpha$  is the aggregate outcome rate in the MU-game in Part II of the experiment. The regressors  $PU$ ,  $MD$  and  $PD$  represent the three interventions and are equal to one if an observation comes from the respective game and zero otherwise. The variable  $Part$  takes on value one if an observation comes from Part I of the experiment and zero otherwise. The subject-specific random effect  $u_i$  takes into account that subjects ( $i$ ) make repeated decisions in the periods ( $t$ ) of the parts ( $h$ ) of the experiment. Lastly, the error term is represented by  $\epsilon_{ith}$ .

## 4.2 Empirical results

**Microtargeting and undisclosed interests.** Tables 3 and 4 report the coefficient estimates of  $\alpha$ ,  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$  in (1) for all our outcome measures.

Table 3: Marginal effects of a microtargeting ban and disclosure of interests on truth-telling and trust

	Measure	Interaction		Status quo				Intervention				Combination					
		Interest group	Voters	Status quo		Microtargeting ban				Disclosure							
				Coeff.	St. dev.	Coeff.	St. dev.	Sig.	Pred.	Coeff.	St. dev.	Sig.	Pred.	Coeff.	St. dev.	Sig.	Pred.
1	Truth	Majority	Majority	0.88	0.05	-0.05	0.07		=	0.07	0.05		=	0.07	0.05		=
2			Minority	0.60	0.07	0.25	0.11	**	+	-0.06	0.12		=	0.34	0.09	***	+
3		Minority	Majority	0.42	0.10	-0.02	0.14		=	0.23	0.14		+	0.30	0.13	**	+
4			Minority	0.70	0.10	-0.30	0.14	**	-	0.23	0.11	**	+	0.02	0.13		=
5	Trust	Majority	Majority	0.74	0.05	-0.09	0.06		=	0.17	0.06	***	=	0.11	0.06	*	=
6			Minority	0.41	0.07	0.14	0.09		+	0.00	0.10		=	0.44	0.09	***	+
7		Minority	Majority	0.77	0.05	-0.17	0.08	**	=	-0.35	0.09	***	-	-0.45	0.08	***	-
8			Minority	0.38	0.08	0.25	0.11	**	+	0.43	0.12	***	+	0.16	0.12		=

*Notes.* This table reports the estimated outcome rates of the status quo as well as the marginal effects of the interventions microtargeting ban, disclosure of interests and a combination of these two interventions. The marginal effects are estimated using model 1 with robust standard errors, \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The symbols ‘+’, ‘-’, ‘=’, and ‘?’ in the column Pred. denote that the predicted marginal effect is, respectively, positive, negative, equal to zero, and ambiguous.

*How to read.* By means of illustration, in the first row it can be seen that the estimated probability of a Majority interest group being truthful to the Majority voters is 0.88, with a standard deviation of 0.05, in the status quo (with microtargeting and undisclosed interests). The estimated probability that a Majority interest group is truthful to Majority voters is 0.05 lower (i.e.,  $0.88 - 0.05 = 0.83$ ) in the treatment with the microtargeting ban. This difference is, however, not statistically significant at the 10%-level, which is in line with our theoretical prediction (=). Similarly, the estimated probabilities are 0.07 higher in the treatments with disclosure of interests and a combination of a microtargeting ban and disclosure of interests. These differences are also not statistically significant at the 10%-level.

In the benchmark, MU-game, as expected, we find higher truthfulness and trust if sender and receiver have the same type: a Majority interest group is more truthful to a Majority

<sup>27</sup>We consider truth and the efficiency of election outcomes (efficient, inefficient, tie) for subjects in the role of interest group and trust and the efficiency of vote choices (efficient, inefficient, abstention) for subjects in the role of voter. In addition, we also look at voter payoffs.

Table 4: Marginal effects of a microtargeting ban and disclosure of interests on voting behavior

Voter	Interest group	Vote choice	Status quo		Microtargeting ban				Intervention				Combination				
			Coeff.	St. dev.	Coeff.	St. dev.	Sig.	Pred.	Coeff.	St. dev.	Sig.	Pred.	Coeff.	St. dev.	Sig.	Pred.	
1	Majority	Majority	Efficient	0.64	0.05	-0.08	0.07		=	0.24	0.07	***	+	0.17	0.06	***	+
2			Inefficient	0.14	0.02	0.00	0.03		=	-0.08	0.03	***	-	-0.10	0.03	***	-
3			Abstention	0.24	0.05	0.05	0.07		=	-0.17	0.06	***	-	-0.10	0.06	*	-
4	Minority	Minority	Efficient	0.32	0.06	0.00	0.08		=	0.00	0.08		=	-0.03	0.08		=
5			Inefficient	0.43	0.06	0.00	0.08		=	-0.22	0.08	***	-	-0.17	0.08	**	-
6			Abstention	0.20	0.05	0.07	0.07		=	0.28	0.09	***	+	0.27	0.08	***	+
7	Minority	Majority	Efficient	0.33	0.06	0.16	0.09	*	+	-0.07	0.08		=	0.51	0.07	***	+
8			Inefficient	0.32	0.04	-0.13	0.06	**	-	-0.07	0.07		=	-0.28	0.05	***	-
9			Abstention	0.33	0.05	0.02	0.06		-	0.17	0.08	**	=	-0.21	0.08	***	-
10	Minority	Minority	Efficient	0.37	0.09	-0.05	0.11		=	0.41	0.14	***	+	0.01	0.12		=
11			Inefficient	0.16	0.06	0.23	0.10	**	+	-0.07	0.07		=	0.12	0.08		=
12			Abstention	0.49	0.07	-0.19	0.10	*	-	-0.34	0.11	***	-	-0.13	0.10		=

*Notes.* This table reports the estimated outcome rates of the status quo as well as the marginal effects of the interventions microtargeting ban, disclosure of interests and a combination of these two interventions. The marginal effects are estimated using model 1 with robust standard errors, \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The symbols '+', '-', '=', and '?' in the column Pred. denote that the predicted marginal effect is, respectively, positive, negative, equal to zero, and ambiguous.

voter than to a Minority voter (estimated coefficients of 0.88 and 0.60, respectively; rows 1 and 2 of Table 3).<sup>28</sup> For a Minority interest group, the reverse holds true (rows 3 and 4 of Table 3). As predicted by the most informative equilibrium, trust is lower for Minority voters than for Majority voters (rows 5 to 8 in Table 3).

In line with our theoretical predictions, a Majority voter is *more* likely to choose an efficient voting action, *less* likely to make an inefficient vote choice and *less* likely to abstain than a Minority voter if the interest group has type Majority (see rows 1 to 3 and 7 to 9 of Table 4). Consequently, a Majority voter obtains higher payoffs from voting than a Minority voter (rows 1 and 4 of Table 5). The outcomes with a Minority interest group are analogous to this case (see rows 4 to 6 and 10 to 12 of Table 4 and rows 2 and 5 of Table 5). Aggregate voter behavior is also as expected: election outcomes are less efficient with a Minority interest group than with a Majority interest group (Table 6).

**Microtargeting ban.** Due to the discipline effect of the public communication technology, a Majority interest group becomes *more* truthful and a Minority interest group *less* truthful to Minority voters, who display a higher level of trust in the message that they

<sup>28</sup>In this section, we solely analyze behavior and outcomes across games and do not provide statistical tests for our within-game comparative static predictions. An analysis of the four individual games can be found in Appendix A.8.

Table 5: Marginal effects of a microtargeting ban and disclosure of interests on voter payoffs

Voter	Interest group	Status quo		Microtargeting ban				Intervention				Combination				
		Coeff.	St. dev.	Coeff.	St. dev.	Sig.	Pred.	Coeff.	St. dev.	Sig.	Pred.	Coeff.	St. dev.	Sig.	Pred.	
1	Majority	Majority	74.63	1.87	-2.47	2.58		=	10.08	2.64	***	+	8.48	2.36	***	+
2		Minority	55.35	3.34	-0.78	4.31		=	7.19	4.14	*	+	4.97	4.02		+
3		All	67.68	1.59	-1.24	2.28		=	9.58	2.37	***	+	8.20	2.02	***	+
4	Minority	Majority	59.10	2.55	9.43	3.97	**	+	0.65	3.74		=	24.76	3.01	***	+
5		Minority	65.57	3.93	-9.58	5.71	*	-	14.37	5.50	***	+	-2.73	5.37		=
6		All	61.52	2.34	2.82	3.26		+	5.10	3.39		+	15.66	3.01	***	+

*Notes.* This table reports the estimated outcome rates of the status quo as well as the marginal effects of the interventions microtargeting ban, disclosure of interests and a combination of these two interventions on voter payoffs. The marginal effects are estimated using model 1 with robust standard errors, \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The symbols ‘+’, ‘-’, ‘=’, and ‘?’ in the column Pred. denote that the predicted marginal effect is, respectively, positive, negative, equal to zero, and ambiguous.

Table 6: Marginal effects of a microtargeting ban and disclosure of interests on election outcomes

Interest group	Election outcome	Status quo		Microtargeting ban				Intervention				Combination				
		Coeff.	St. dev.	Coeff.	St. dev.	Sig.	Pred.	Coeff.	St. dev.	Sig.	Pred.	Coeff.	St. dev.	Sig.	Pred.	
1	Majority	Efficient	0.72	0.05	-0.20	0.07	***	-	0.16	0.07	**	+	0.08	0.07		+
2		Inefficient	0.17	0.05	0.02	0.06		=	-0.09	0.06		-	-0.11	0.05	**	-
3		Tie	0.13	0.04	0.15	0.05	***	+	-0.08	0.04	*	-	0.03	0.05		-
4	Minority	Efficient	0.32	0.10	0.05	0.12		=	-0.08	0.14		=	-0.10	0.11		=
5		Inefficient	0.53	0.08	-0.07	0.11		-	-0.03	0.12		?	-0.23	0.11	**	-
6		Tie	0.15	0.04	0.02	0.05		+	0.12	0.08		?	0.33	0.06	***	+

*Notes.* This table reports the estimated outcome rates of the status quo as well as the marginal effects of the interventions microtargeting ban, disclosure of interests and a combination of these two interventions. The marginal effects are estimated using model 1 with robust standard errors, \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The symbols ‘+’, ‘-’, ‘=’, and ‘?’ in the column Pred. denote that the predicted marginal effect is, respectively, positive, negative, equal to zero, and ambiguous.

receive. All these effects but the higher trust of Minority voters in Majority interest groups are statistically significant (rows 2, 4, 6 and 8 of Table 3).

As a result, the voting behavior of Minority voters becomes more efficient and less inefficient if the interest group has type Majority and more inefficient otherwise (rows 7 to 12 of Table 4). In contrast to our prediction, the first effect is not sizeable enough to outweigh the latter effect. On average, the microtargeting ban neither has a significant impact on efficient and inefficient voting actions nor on abstention.<sup>29</sup> As a result, a Minority voter’s average payoffs are not significantly affected by the ban (row 6 of Table 5).

In line with our theory, we do not observe any significant differences in the truthfulness of messages to Majority voters (row 1 and 3 of Table 3). At odds with this reporting behavior, Majority voters display a lower level of trust after the introduction of the microtargeting

<sup>29</sup>The estimated coefficients (with standard deviations in parentheses) are, respectively 0.08 (0.07), -0.01 (0.05) and -0.04 (0.06) for efficient votes, inefficient votes and abstention.

ban (rows 5 and 7 of Table 3). This drop in trust, however, does not lead to a significant change in the efficiency of voting behavior of a Majority voters: the microtargeting ban has no significant effect on vote efficiency, vote inefficiency and abstention (rows 1 to 3 of Table 3). Similarly, a Majority voter’s payoffs do not significantly change due to the ban (rows 1 to 3 of Table 5).

In line with our prediction (Hypothesis 1b), the implementation of the microtargeting ban without mandatory disclosure of interests decreases the efficiency of aggregate voter decision-making. The ban takes away an instrument of the interest group to influence the election outcome in its favor. This results in less efficient election outcomes (Table 6).<sup>30</sup>

Our main findings are summarized in the following result.

**Result 1.**

- a. **Voter payoffs.** *A microtargeting ban has no statistically significant effect on the payoffs from voting for Majority and Minority voters.*
- b. **Election outcome.** *A microtargeting ban decreases the efficiency of aggregate voter decision-making significantly.*

**Mandatory disclosure of interests.** As expected, mandatory disclosure of interests only affects the truthfulness of messages sent by a Minority interest group (Table 3). Although our theory predicts a positive effect for messages to both Majority and Minority voters, only messages directed at the latter group become significantly more truthful. All our predictions regarding trust are borne out by the data (Table 3): a Majority voter trusts messages of a Majority (Minority) interest group more (less) than before the intervention. A Minority voter has more trust in messages from a Minority interest group than before.

The changes in interest group and voter behavior lead to improvements in voter decision-making for all voters (Table 4 and rows 1 to 3 and 5 of Table 5). On aggregate, we find that mandatory disclosure of interests leads to a significant increase in efficient election outcomes

---

<sup>30</sup>Note that we only observe a significant effect for election outcomes with a Majority interest group.

(Table 6). This improvement is driven by outcomes in elections with a Majority interest group (rows 1 to 3 of Table 6).

Summarizing our analysis, we state the following result.

**Result 2.**

- a. **Voting behavior.** Mandatory disclosure of interests significantly increases the payoffs from voting for Majority voters and Minority voters.*
- b. **Election outcome.** Mandatory disclosure of interests significantly increases the efficiency of aggregate voter decision-making.*

**Microtargeting ban and mandatory disclosure of interests.** Our predictions about the effects of this intervention on interest group behavior are all borne out by the data. The disciplining effect of the public communication technology causes the Majority interest group to be more truthful to a Minority voter (row 2 of Table 3) but has no effect on its reporting behavior to Majority voters (row 1). Due to mandatory disclosure of interests, both voter groups are able to recognize the Majority interest group. As a result, we see that trust increases for all voters (rows 5 and 6).

According to our theory, a Minority interest group can no longer deceive Majority voters by consistently reporting false messages due to disclosure of interests. Its equilibrium reporting strategy does not change with respect to a Minority voter. Both of these predictions are confirmed (rows 3 and 4 of Table 3). As expected, the intervention reduces trust of a Majority voter and has effect on trust of a Minority voter regarding messages from a Minority interest group (rows 7 and 8).

The changes in interest group and voter behavior due to the intervention unambiguously improve the efficiency of voter decision-making for (i) Majority voters, (ii) Minority voters and (iii) the electorate as a whole. On the individual level, voters choose more efficient and less inefficient voting actions (Table 4), which gives rise to higher payoffs for both voter groups (Table 5). On the aggregate level, we do not observe a significant rise in efficient

election outcomes (rows 1 and 4 of Table 6). However, the probability of having an inefficient election outcome decreases significantly (rows 2 and 5 of Table 6).

**Result 3.**

- a. **Voter payoffs.** *Mandatory disclosure of interests in combination with a microtargeting ban significantly increases the payoffs from voting for Majority voters and Minority voters.*
- b. **Election outcome.** *Mandatory disclosure of interests in combination with a microtargeting ban significantly increases the efficiency of aggregate voter decision-making.*

## 5 Discussion

**Competition among senders:** In our games and the experiment, there is only one political interest group that sends exactly one message to voters before the receivers take a voting action. This is a reduced-form setup that mimics the algorithmic random selection of one message out of a large set of messages provided by many heterogeneous interest groups, potentially customized to each voter’s characteristics. In practice, voters get many messages, from many senders, not only one.

We do not have evidence to speak to this case. However, the theoretical model suggests that the effects of competition among senders are different for each game.<sup>31</sup> In the *MD-game*, we found that interest groups will always send a truthful message to voters of the same type—and that voters will trust this message due to the disclosed identity of the interest group. Consequently, all uncertainty about the state of the world is resolved as soon as a voter encounters a message from an interest group with a matching type (regardless of how many messages are sent). Related, in the *PD-game* all uncertainty about the state of the world is resolved as soon as a voter encounters a message from an interest group with type Majority (regardless of how many messages are sent).

---

<sup>31</sup>Van Gils et al. (2020) study the effects of competing messages on voters’ beliefs and voting actions in a more general model.

Decision-making is harder for voters in games without disclosed interest group types. In the *MU-game*, for Majority voters, all messages affect beliefs but at a decreasing rate. There is convergence to the truth if the number of messages increases. For Minority voters, messages remain completely uninformative even if they receive many: they cannot trust any microtargeted message in a social media environment. Finally, in the *PU-game*, all messages affect the beliefs about the state of the world of all voters—but at a decreasing rate. There is convergence to the truth if the number of messages increases. Hence, in environments without disclosed sender types, competition among interest groups has a positive effect on voters’ information even if interest groups have divergent objectives and send different messages.

**Voter group size:** We studied cases where one voter group of the electorate (Majority) is significantly larger than another group (Minority). In reality, however, many elections are very tight.<sup>32</sup> To which extent do our results generalize to environments where both groups are roughly equal in size?

Here, too, the theoretical results depend on the game. In the *MD-game*, neither the size of the interest group nor the size of the voter group matters for the results. It only matters whether the voter type matches the interest group type. In the *PD-game*, the size of the interest group does not matter. The size of the voter group matters because a message will be tailored to the largest voter group. There are multiple equilibria if both voter groups have exactly the same size. However, the results presented above are qualitatively robust as long as one voter group is slightly larger than the other one (like 50,1% vs. 49,9%).

In both the *PU-game* and the *MU-game*, the sizes of the voter groups only matter if there is complete symmetry (as for the *PD-game*). The sizes of the interest groups determine the informativeness of the messages sent. If there is complete symmetry, messages are completely uninformative. In the *PU-game*, communication also breaks down if the interests of the largest interest group are aligned with the interests of the smallest voter group. If

---

<sup>32</sup>For instance, the general elections in Brazil in October 2022 ended with 50.9% of the votes for the challenging Presidential candidate and 49.1% for the incumbent ([https://en.wikipedia.org/wiki/2022\\_Brazilian\\_general\\_election](https://en.wikipedia.org/wiki/2022_Brazilian_general_election)).

this happened in the MU-environment, our results would flip: messages sent to the largest voter group would be uninformative and messages sent to the smallest voter group would be informative. However, as long as all voters groups are roughly presented equally among political interest groups, this combination is highly unlikely.

## 6 Conclusion

We have theoretically and experimentally studied the effects of two recently proposed policy measures aiming to prevent the manipulation of democratic elections in times of social media: a ban on microtargeted political messages that depend on the receiving voter’s characteristics or preferences, and mandatory disclosure of interests of the political interest groups sending messages to voters in a social media environment. Our results show that mandatory disclosure of interests, in combination with or without a microtargeting ban, increases the efficiency of aggregate voter decision-making. However, only the combination of disclosure of interests and a microtargeting ban counteracts election manipulation. The implementation of a microtargeting ban without mandatory disclosure of interests has an adverse effect: in our lab experiment, the frequency of election outcomes won by a party that is supported by a majority of the electorate is reduced significantly and substantially.

On the level of individual voter-decision making, mandatory disclosure of interests is found to be beneficial in combination with and without a microtargeting ban. Both Majority voters and Minority voters are better off due to these interventions. The implementation of a microtargeting ban without disclosure requirements, however, does not significantly improve individual voting actions. Our model predicts this surprising result: it shows that microtargeting hurts voters with non-mainstream preferences (here: Minority voters): they do not receive credible messages from any interest group type. Majority voters, by contrast, can count on the truthfulness of received messages, especially if they can identify the sender’s type and the sender has aligned political preferences. Consequently, if microtargeting is

banned, the intervention benefits Minority voters because they benefit from a discipline effect: they know that messages they receive are the same as Majority voters' and, hence, are trustworthy if originated from a Majority interest group (which occurs more often than not, even if the sender's type is undisclosed). This allows them to avoid inefficient voting decisions. However, as long as non-mainstream voters have opposed political objectives to majority voters, this hurts aggregate election efficiency.

Various thorough robustness checks confirm our main results (see Appendices A.6 to A.8).

These theoretical and experimental results have immediate implications for important ongoing legislative initiatives. The European Union, in particular, is active and has included various rules regarding the transparency of news senders and the use of microtargeting technologies in recent legislation. The EU's Digital Services Act (DSA) mandates *Very Large Online Platforms* to fight disinformation: they have to deliver (public) annual risk-assessment reports and risk-mitigation reports, including annual audits by independent parties. They are also subject to transparency requirements towards final consumers (DSA Art.24). Complementing the DSA, the Digital Markets Act (DMA) restricts the use of microtargeted advertising, especially before elections (Art 6(aa)) and requires more transparency towards advertisers/intermediaries (Art 6(g)).

On 24 January 2023, a Committee of the European Parliament adopted new rules on political advertising: "The changes made to the Commission's proposal require that only personal data explicitly provided for online political advertising can be used by advert providers. This creates a de facto ban on micro-targeting." Moreover, an online repository containing all online political advertisements and related data has to be established, which will make it much easier to obtain information on who is financing an advert, on how much was paid for it, and from where the money originated.<sup>33</sup>

In the US, the recently proposed Platform Accountability and Transparency Act would

---

<sup>33</sup>All details from <https://www.europarl.europa.eu/news/en/press-room/20230123IPR68616/mep-s-toughen-rules-on-political-advertising>.

require large social media firms to share data with qualified researchers, to make public content spread by large user accounts, and to inform the public about the platform’s content moderation policies and decisions.<sup>34</sup>

As all of these legislative initiatives are new or even ongoing, the next big challenges for policy makers—and for academic researchers—lie in the realms of implementation and enforcement. The fact, that large social media platforms have billions of users, makes every change of rules, and its monitoring and enforcement, a Herculean task. Bearing this in mind, it is somewhat ironic that a microtargeting ban, which we found to be only beneficial in combination with disclosure rules about interest-group identity, would be relatively easy to enforce: especially contracts between social media platforms and large advertisers could be monitored, and regulators could randomly check the number and types of messages sent to certain voter groups, in particular vulnerable ones. Governing such rules by the use of technology—in the form of apps, machine-learning algorithms, and tracking devices—may even partly automatize enforcement.

By contrast, implementing disclosure rules in the political and technological practice is significantly harder. In our experiment, subjects in the role of voters faced only two types of political interest groups and they clearly understood the objectives of both: either goals between the interest group and themselves were perfectly aligned or perfectly opposed. Reality is much more nuanced and multidimensional, such that even perfect disclosure might require stochastic strategies of voters making voting decisions under some remaining uncertainty. Moreover, biases abound (e.g., voters’ myopia, forgetfulness, or cognitive overload). Therefore, any disclosure rule would have to make the distance between a voter’s own objectives and the political interest group’s objectives clear with very simple tools. This should allow the voter to assess the credibility of this one message in a snapshot while keeping the result in mind until election day.

---

<sup>34</sup><https://www.brennancenter.org/our-work/research-reports/law-requiring-social-media-transparency-would-break-new-ground>

## References

- Allcott, H. and M. Gentzkow (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives* 31(2), 211–36.
- Argenton, C. and J. Prüfer (2012). Search engine competition with network externalities. *Journal of Competition Law and Economics* 8(1), 73–105.
- Battaglini, M. and U. Makarov (2014). Cheap talk with multiple audiences: An experimental analysis. *Games and Economic Behavior* 83, 147–164.
- Beknazar-Yuzbashev, G. and M. Stalinski (2022). Do social media ads matter for political behavior? a field experiment. *Journal of Public Economics* 214, 104735.
- Bond, R. M., C. J. Fariss, J. J. Jones, A. D. Kramer, C. Marlow, J. E. Settle, and J. H. Fowler (2012). A 61-million-person experiment in social influence and political mobilization. *Nature* 489(7415), 295–298.
- Cai, H. and J. T.-Y. Wang (2006). Overcommunication in strategic information transmission games. *Games and Economic Behavior* 56(1), 7–36.
- Cain, D. M., G. Loewenstein, and D. A. Moore (2005). The dirt on coming clean: Perverse effects of disclosing conflicts of interest. *The Journal of Legal Studies* 34(1), 1–25.
- Chung, W. and R. Harbaugh (2019). Biased recommendations from biased and unbiased experts. *Journal of Economics & Management Strategy* 28(3), 520–540.
- Drugov, M., R. Hernan Gonzalez, P. Kujal, and M. Troya-Martinez (2017). Cheap talk with two audiences: An experiment. Working paper, Available at <http://dx.doi.org/10.2139/ssrn.3086405>.
- Farrell, J. and R. Gibbons (1989). Cheap talk with two audiences. *American Economic Review* 79(5), 1214–1223.

- Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10(2), 171–178.
- Fréchette, G. (2012). Session-effects in the laboratory. *Experimental Economics* 15(3), 485–498.
- Fudenberg, D. and J. Tirole (1991). Perfect bayesian equilibrium and sequential equilibrium. *Journal of Economic Theory* 53(2), 236–260.
- Fujiwara, T., K. Müller, and C. Schwarz (2021). The effect of social media on elections: Evidence from the united states. Working paper, Available at <http://www.nber.org/papers/w28849>.
- Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association* 1(1), 114–125.
- Großer, J. (2020). Voting game experiments with incomplete information: a survey. In C. M. Capra, R. T. Croson, M. L. Rigdon, and T. S. Rosenblat (Eds.), *Handbook of Experimental Game Theory*, pp. 376–398. Edward Elgar Publishing.
- Guess, A. M. and B. A. Lyons (2020). Misinformation, disinformation, and online propaganda. In N. Persily and J. A. Tucker (Eds.), *Social Media and Democracy: The State of the Field, Prospects for Reform*, pp. 10–33. Cambridge University Press.
- Guess, A. M., B. Nyhan, and J. Reifler (2020). Exposure to untrustworthy websites in the 2016 US election. *Nature human behaviour* 4(5), 472–480.
- Holt, C. A. and S. K. Laury (2002). Risk aversion and incentive effects. *American Economic Review* 92(5), 1644–1655.
- Jones, J. J., R. M. Bond, E. Bakshy, D. Eckles, and J. H. Fowler (2017). Social influence and political mobilization: Further evidence from a randomized experiment in the 2012 US presidential election. *PloS one* 12(4), 1–9.

- Kalogeropoulos, A. and N. Newman (2017). I saw the news on Facebook. Reuters Institute for the Study of Journalism.
- Kang, H., K. Bae, S. Zhang, and S. S. 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, M. and J. Tremewan (2018). An offer you can refuse: the effect of transparency with endogenous conflict of interest. *Journal of Public Economics* 161, 44–55.
- Kartal, M. and J.-R. Tyran (2022). Fake news, voter overconfidence, and the quality of democratic choice. *American Economic Review* 112(10), 3367–97.
- Kim, D. (2020). Clustering standard errors at the “session” level. Working Paper, Available at <http://dx.doi.org/10.2139/ssrn.3635181>.
- Kosinski, M., D. Stillwell, and T. 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.
- Liberini, F., A. Russo, Á. Cuevas, and R. Cuevas (2020). Politics in the Facebook era. Evidence from the 2016 US presidential elections. Working Paper, Available at <https://ssrn.com/abstract=3584086>.
- Lorenz-Spreen, P., L. Oswald, S. Lewandowsky, and R. Hertwig (2023). A systematic review of worldwide causal and correlational evidence on digital media and democracy. *Nature Human Behaviour* 7, 74–101.
- Nyarko, Y. and A. Schotter (2002). An experimental study of belief learning using elicited beliefs. *Econometrica* 70(3), 971–1005.
- Parsons, T. (2010). *The rule of empires: those who built them, those who endured them, and why they always fall*. Oxford University Press.

- Persily, N. and J. A. Tucker (2020). *Social Media and Democracy: The State of the Field, Prospects for Reform*. Cambridge University Press.
- Pogorelskiy, K. and M. Shum (2019). News we like to share: How news sharing on social networks influences voting outcomes. Working paper, Available at <http://dx.doi.org/10.2139/ssrn.2972231>.
- Rotesi, T. (2019). Do social media matter? The impact of Twitter on political participation. Working paper.
- Rubin, J. (2017). *Rulers, Religion, and Riches: Why the West got rich and the Middle East did not*. Cambridge University Press.
- Schuessler, A. A. (2000). *A logic of expressive choice*. Princeton University Press.
- Shapiro, J. M. (2016). Special interests and the media: Theory and an application to climate change. *Journal of Public Economics* 144, 91–108.
- Sun, J., A. Schram, and R. Sloof (2021). Elections under biased candidate endorsements — an experimental study. *Games and Economic Behavior* 125, 141–158.
- Tucker, J., A. Guess, P. Barberá, C. Vaccari, A. Siegel, S. Sanovich, D. Stukal, and B. Nyhan (2018). Social media, political polarization, and political disinformation: A review of the scientific literature. Hewlett Foundation.
- Tucker, J. A., Y. Theocharis, M. E. Roberts, and P. Barberá (2017). From liberation to turmoil: Social media and democracy. *Journal of Democracy* 28(4), 46–59.
- Van Gils, F., W. Müller, and J. Prüfer (2020). Big data and democracy. TILEC Discussion Paper No. DP 2020-003, Available at <http://dx.doi.org/10.2139/ssrn.3556512>.
- Zhuravskaya, E., M. Petrova, and R. Enikolopov (2020). Political effects of the internet and social media. *Annual Review of Economics* 12, 415–438.

Ziegler, A. (2023). Persuading an audience: Testing information design in the laboratory. Working Paper, Available at [https://www.creedexperiment.nl/creed/pdffiles/ziegler\\_jump.pdf](https://www.creedexperiment.nl/creed/pdffiles/ziegler_jump.pdf).

Zuiderveen Borgesius, F., J. Möller, S. Kruikemeier, R. Ó Fathaigh, K. Irion, T. Dobber, B. Bodo, and C. H. de Vreese (2018). Online political microtargeting: promises and threats for democracy. *Utrecht Law Review* 14(1), 82–96.

## A Online Appendix

This appendix is organized as follows. Section A.1 describes our four games in formal, game-theoretic terms. Sections A.2 to A.5 characterize the most informative perfect Bayesian equilibrium, including its derivation, for each game.

As of Section A.6, we turn from theory to experimental results. To start thoroughly testing for the robustness of our results, we reverse the main text’s sequence of games in Section A.6. Here, subjects start in the traditional media environment (PD-game), where, for instance, every voter reads the same newspaper, watches the same TV-channel, or listens to the same radio station—and knows that everybody else gets the same, potentially slanted, information. Then we study how the introduction of microtargeting, the obfuscation of a message sender’s identity and objectives, and both changes simultaneously, affect individual and aggregate voting actions and outcomes. Where the main part of the paper relies on data from Treatments 1 to 4, this Section is based on data from Treatments 5 to 8 (cf. Table 2).

Section A.6, just as in the paper’s main text, follows a *within-subject* approach: every subject in the lab is exposed first to one game and then to another one, which allows to analyze changes in the behavior and beliefs of the same person. By contrast, Section A.7 applies a *between-subject* approach, which makes use of *all* Part-I games, excludes data from Part-II games, and therefore focuses on differences of different groups of subjects across the four games.

Section A.8 completely drops all order-related information. There we take data from Part I of all games (Treatments 1-12) and study behavior and outcomes in the *individual* games. Qualitatively, it turns out, the results reported in the main part of the paper are highly robust and not subject to order effects. For completeness, Section A.9 reports summary statistics of Part II of the experiment.

Section A.10 reports the distributions of elicited lying-task profiles and Section A.11 the distribution of risk-elicitation task profiles. As the main part of the paper only reports estimated marginal effects and omits the corresponding random effects panel regressions, to

save space, we present several selected regression results in Section [A.12](#). Finally, for exemplification, Section [A.13](#) contains the complete set of instructions for subjects participating in one treatment (Treatment 3 in Table [2](#)).

## A.1 A formal description of the four games

Before we derive the most informative equilibria of the four games (MD, MU, PD and PU), we first provide a formal description of the game ingredients.

There are two Majority voters and one Minority voter who make a voting decision  $d \in \{X, Y, \emptyset\}$ , where  $X$  ( $Y$ ) refers to voting for party  $X$  ( $Y$ ) and  $\emptyset$  to abstaining from voting. Voter payoffs, which are described for each voter type  $r \in \{Majority, Minority\}$  in Table [1b](#), depend on the state of the world  $\theta \in \{One, Two\}$ , with  $Pr(\theta = One) = Pr(\theta = Two) = 1/2$ . The true state is unobservable to voters. Prior to voting, the voters receive a costless, non-binding and non-verifiable (cheap talk) message  $m \in \{One, Two\}$  concerning the state. We denote by  $\mu(m)$  a voter's posterior belief that the realized state is *One*, conditional on the received message  $m$ . The message is sent by an interest group  $s \in \{Majority, Minority\}$ , with  $Pr(s = Majority) = 2/3$ , that perfectly observes the true state. The interest group's payoffs are displayed in Table [1a](#) in the main text.

The timing of our four games is as follows. First, nature determines  $\theta$  and  $s$ . The interest group observes  $\theta$  and  $s$ . Voters do not observe  $\theta$  but also learn  $s$  if there is disclosure. Then, the interest group selects  $m \in \{One, Two\}$  in public games or  $m_{Maj} \in \{One, Two\}$  for Majority voters and  $m_{Min} \in \{One, Two\}$  for Minority voters in microtargeting games. Lastly, each voter observes  $m$  (or  $m_{Maj}$  or  $m_{Min}$ ), updates belief  $\mu(\cdot)$  and makes voting decision  $d \in \{X, Y, \emptyset\}$ . All payoffs are realized and the party that receives a majority of the votes wins the election.

We assume that all players maximize their expected payoffs. The solution concept for our games is the Perfect Bayesian Equilibrium ([Fudenberg and Tirole, 1991](#)). We limit ourselves to the most informative equilibrium, in which the most information is transmitted from the

interest group to voters. A message is informative if  $\mu(m) \neq p$  (i.e., the posterior belief is different from the prior belief). Due to the binary nature of the messages sent by the interest group, informativeness implies that  $\mu(m = One) \neq \mu(m = Two)$ . To simplify our analysis, we restrict ourselves to equilibria in which  $\mu^*(m = One) \geq \mu^*(m = Two)$ .<sup>35</sup> For some of our games, a multitude of equilibria exist in which the reporting behavior to one or both voter groups is uninformative about the true state of the world. For the sake of simplicity, we capture all uninformative reporting strategies by  $\tilde{m}$ , which represents that the message sent is independent of the true state.

## A.2 Equilibrium MU-game

**Proposition 1.** *The following strategy profiles and beliefs constitute the most informative perfect Bayesian equilibrium of the MU-game:*

$$m^*(s, \theta) = \begin{cases} m_{Maj} = One \text{ and } m_{Min} = \tilde{m} & \text{if } \theta = One \\ m_{Maj} = Two \text{ and } m_{Min} = \tilde{m} & \text{if } \theta = Two \end{cases} \quad \text{for } s = Majority \quad (2)$$

$$m^*(s, \theta) = \begin{cases} m_{Maj} = Two \text{ and } m_{Min} = \tilde{m} & \text{if } \theta = One \\ m_{Maj} = One \text{ and } m_{Min} = \tilde{m} & \text{if } \theta = Two \end{cases} \quad \text{for } s = Minority$$

$$d^*(r, m_r) = \begin{cases} X & \text{if } m_{Maj} = One \\ Y & \text{if } m_{Maj} = Two \end{cases} \quad \text{for } r = Majority \quad (3)$$

$$d^*(r, m_r) = \emptyset \quad \text{for } r = Minority$$

---

<sup>35</sup>Equilibria formed by a permutation of the messages  $m = One$  and  $m = Two$  also exist.

$$\mu^*(r, m_r) = \begin{cases} \frac{2}{3} & \text{if } m_{Maj} = One \\ \frac{1}{3} & \text{if } m_{Maj} = Two \\ \frac{1}{2} & \end{cases} \begin{cases} \text{for } r = Majority \\ \\ \text{for } r = Minority \end{cases} \quad (4)$$

*Proof.* Given (2), it follows from Bayes' rule that voters' posterior beliefs are as specified in (4). Voters maximize their expected payoff, given the voter belief  $\mu^*(r, m_r)$ , by voting for party  $X$  ( $Y$ ) if (i) the expected payoff from voting for party  $X$  ( $Y$ ) outweighs the expected payoff from voting for party  $Y$  ( $X$ ) and (ii) the expected payoff from voting for party  $X$  ( $Y$ ) outweighs the payoff of abstention. Using (4) and Table 1b, we find that conditions (i) and (ii) are fulfilled if a voter follows the voting rule stated in (3). Using (3) and Table 1a, we find that an interest group cannot increase its payoff by deviating from the strategy stated in (2). Thus, the game has an equilibrium in which the interest group follows the strategy in (2) and voters vote according to (3) and update their beliefs according to (4). Since  $\mu^*(r = Majority, m_r = One) > \mu^*(r = Majority, m_r = Two)$ , the equilibrium is informative. In any equilibrium in which  $\mu^*(r = Majority, m_r = One) > \mu^*(r = Majority, m_r = Two)$ , the interest group has an incentive to follow the reporting strategy specified in (2) (see (3) and Table 1a). It then follows from Bayes' rule that the beliefs of a Majority voter must be as stated in (4). Hence, there does not exist an equilibrium in which more information is transmitted to Majority voters. In any equilibrium, messages must be uninformative to Minority voters (4). Suppose this is not the case (i.e.,  $\mu^*(r = Minority, m_r = One) > \mu^*(r = Minority, m_r = Two)$ ). A Majority interest group would then have an incentive to report the opposite of the true state to the Minority voter (see Tables 1a and 1b). Since the interest group has type Majority with probability  $2/3$ , this would be inconsistent with the equilibrium voter beliefs. Thus, the informative equilibrium described in (2)-(4) is also the most informative equilibrium.  $\square$

*Explanatory note.* The messaging strategy in (2) states that an interest group with type Majority reports to Majority voters that the state of the world is equal to One and that it sends an uninformative message (i.e., unrelated to the true state of the world) to the Minority voter if the true state of the world is One. If the true state of the world is Two, the Majority interest group reports to Majority voters that the state is Two and sends an uninformative message to the Minority voter. It is also stated in (2) that an interest group with type Minority reports to Majority voters that the state of the world is Two (One) if the true state is One (Two), and that the Minority interest group sends an uninformative message to the Minority voter in both states of the world. According to the voting rule in (3), Majority voters cast the vote X (Y) if they receive the message that the state of the world is One (Two). The Minority voter always abstains from voting, regardless of the message received. In (4), it can be seen that Majority voters believe that the true state of the world is One with probability  $2/3$  (and, hence, that the true state is Two with probability  $1/3$ ) if they receive a message stating that the state of the world is One. Upon receiving the message that the true state is Two, Majority voters believe that the state of the world is One with probability  $1/3$  and Two with probability  $2/3$ . A Minority voter believes that both states of the world are equally likely (the probability of the true state being One is  $1/2$ ), regardless of the message received.

### A.3 Equilibrium MD-game

**Proposition 2.** *The following strategy profiles and beliefs constitute the most informative perfect Bayesian equilibrium of the MD-game:*

$$m^*(s, \theta) = \begin{cases} \begin{cases} m_{Maj} = One \text{ and } m_{Min} = \tilde{m} & \text{if } \theta = One \\ m_{Maj} = Two \text{ and } m_{Min} = \tilde{m} & \text{if } \theta = Two \end{cases} & \text{for } s = Majority \\ \\ \begin{cases} m_{Maj} = \tilde{m} \text{ and } m_{Min} = One & \text{if } \theta = One \\ m_{Maj} = \tilde{m} \text{ and } m_{Min} = Two & \text{if } \theta = Two \end{cases} & \text{for } s = Minority \end{cases} \quad (5)$$

$$d^*(r, s, m_r) = \begin{cases} \begin{cases} X & \text{if } m_{Maj} = One \text{ and } s = Majority \\ Y & \text{if } m_{Maj} = Two \text{ and } s = Majority \\ \emptyset & \text{if } s = Minority \end{cases} & \text{for } r = Majority \\ \\ \begin{cases} Y & \text{if } m_{Min} = One \text{ and } s = Minority \\ X & \text{if } m_{Min} = Two \text{ and } s = Minority \\ \emptyset & \text{if } s = Majority \end{cases} & \text{for } r = Minority \end{cases} \quad (6)$$

$$\mu^*(r, s, m_r) = \begin{cases} \begin{cases} 1 & \text{if } m_{Maj} = One \text{ and } s = Majority \\ 0 & \text{if } m_{Maj} = Two \text{ and } s = Majority \\ \frac{1}{2} & \text{if } s = Minority \end{cases} & \text{for } r = Majority \\ \\ \begin{cases} 1 & \text{if } m_{Min} = One \text{ and } s = Minority \\ 0 & \text{if } m_{Min} = Two \text{ and } s = Minority \\ \frac{1}{2} & \text{if } s = Majority \end{cases} & \text{for } r = Minority \end{cases} \quad (7)$$

*Proof.* Given (5), it follows from Bayes' rule that voter beliefs are as specified in (7). Given (7) and using Table 1b, the maximization of the expected voter payoff implies that voters follow the voting rule in (6). Given (6) and using Table 1a, we find that no interest group can profitably deviate by changing the reporting strategy stated in (5). Thus, the game has an equilibrium in which the interest group acts according to (5) and the voters follows (6) and (7). This equilibrium is informative because  $\mu^*(r, m_r = One) > \mu^*(r, m_r = Two)$  if  $s = r$ . All uncertainty about the state of the world is resolved if  $s = r$ , which means that no further information transmission is possible. If  $s \neq r$ , it is not possible that any information is transmitted in equilibrium. If this would not be the case ( $\mu^*(r, s \neq r, m_r = One) > \mu^*(r, s \neq r, m_r = Two)$ ), the interest group (with type  $s \neq r$ ) has an incentive to report the opposite of the true state (see Table 1a, (6) and (5)), which would be inconsistent with equilibrium voter beliefs. Thus, the equilibrium described in (5)-(7) is the most informative equilibrium of the MD-game.  $\square$

## A.4 Equilibrium PU-game

**Proposition 3.** *The following strategy profiles and beliefs constitute the most informative perfect Bayesian equilibrium of the PU-game:*

$$m^*(s, \theta) = \begin{cases} \begin{array}{ll} One & \text{if } \theta = One \\ Two & \text{if } \theta = Two \end{array} & \text{for } s = Majority \\ \begin{array}{ll} Two & \text{if } \theta = One \\ One & \text{if } \theta = Two \end{array} & \text{for } s = Minority \end{cases} \quad (8)$$

$$d^*(r, m) = \begin{cases} X & \text{if } m = \text{One} \\ Y & \text{if } m = \text{Two} \end{cases} \quad \text{for } r = \text{Majority} \quad (9)$$

$$d^*(r, m) = \begin{cases} Y & \text{if } m = \text{One} \\ X & \text{if } m = \text{Two} \end{cases} \quad \text{for } r = \text{Minority}$$

$$\mu^*(m) = \begin{cases} \frac{2}{3} & \text{if } m = \text{One} \\ \frac{1}{3} & \text{if } m = \text{Two} \end{cases} \quad (10)$$

*Proof.* Given (8), it follows from Bayes' rule that voter beliefs are as specified in (10). Given (8) and using Table 1b, the maximization of the expected payoffs implies that voters follow the voting rule given in (9). Given (9) and using Table 1a, we find that no interest group can profitably deviate by changing the reporting strategy in (8). Thus, there is an equilibrium described by (8)-(10). Since  $\mu^*(m = \text{One}) > \mu^*(m = \text{Two})$ , the equilibrium is informative. The interest group has an incentive to report as stated in (8) in any equilibrium in which  $\mu^*(m = \text{One}) > \mu^*(m = \text{Two})$ . It then follows from Bayes' rule that the beliefs must be as given in (10). Thus, there does not exist a more informative equilibrium than the equilibrium stated here.  $\square$

## A.5 Equilibrium PD-game

**Proposition 4.** *The following strategy profiles and beliefs constitute the most informative perfect Bayesian equilibrium of the PD-game:*

$$m^*(s, \theta) = \begin{cases} m = One & \text{if } \theta = One \\ m = Two & \text{if } \theta = Two \end{cases} \quad \text{for } s = Majority \quad (11)$$

$$\tilde{m} \quad \text{for } s = Minority$$

$$d^*(r, s, m) = \begin{cases} X & \text{if } m = One \text{ and } s = Majority \\ Y & \text{if } m = Two \text{ and } s = Majority \quad \text{for } r = Majority \\ \emptyset & \text{if } s = Minority \end{cases} \quad (12)$$

$$\begin{cases} Y & \text{if } m = One \text{ and } s = Majority \\ X & \text{if } m = Two \text{ and } s = Majority \quad \text{for } r = Minority \\ \emptyset & \text{if } s = Minority \end{cases}$$

$$\mu^*(s, m) = \begin{cases} 1 & \text{if } m = One \text{ and } s = Majority \\ 0 & \text{if } m = Two \text{ and } s = Majority \\ \frac{1}{2} & \text{if } s = Minority \end{cases} \quad (13)$$

*Proof.* Given (11), it follows from Bayes' rule that voter beliefs are as specified in (13). Given (11) and using Table 1b, we find that voters follow the voting rule in (12) to maximize their expected payoff. Given (12) and using Table 1a, we find that no interest group can profitably deviate by changing the reporting strategy stated in (11). Thus, the strategy profiles and

beliefs described in (11)-(13) are an equilibrium. If  $s = Majority$ , all uncertainty about the state of the world is resolved, meaning that further information transmission is not possible. If  $s = Minority$ , it is not possible that any information is transmitted to a voter. If this would not be the case ( $\mu^*(r, s = Minority, m = One) > \mu^*(r, s = Minority, m = Two)$ ), the Minority interest group has an incentive to report the opposite of the true state (see Table 1a, (11) and (12)), which would be inconsistent with equilibrium voter beliefs. Thus, the equilibrium described in (11)-(13) is the most informative equilibrium of the PD-game.  $\square$

## A.6 Analysis across games: interventions in the traditional media environment

In addition to our main analysis, we study the effects of changes to the traditional media environment, characterized by public communication and disclosed interest group types, on the efficiency of voter decision-making. For this purpose, we analyze data from Treatments 5 to 8 (cf. Table 2). Specifically, we analyze the impact of (i) a microtargeting implementation (a shift from the PD-game to the MD-game), (ii) obfuscation of interests (a shift from the PD-game to the PU-game) and a combination of both changes (a shift from the PD-game to the MU-game) within our framework and test the following hypothesis.<sup>36</sup>

**Hypothesis 4** (Microtargeting implementation, obfuscation of interests and efficiency).

- a. **Voter payoffs.** *The payoffs from voting decrease for Minority voters and remain unchanged for Majority voters due to microtargeting implementation. Obfuscation of interests, in combination with or without microtargeting implementation, decrease voter payoffs for both voter groups.*
- b. **Election outcome.** *Microtargeting implementation, obfuscation of interests and a combination of the two interventions decrease the average efficiency of aggregate voter decision-making.*

---

<sup>36</sup>The predictions in Hypothesis 4 are implied by the most informative equilibria of the individual games, which are described in Section 3.3.

On the individual level, a microtargeting implementation is expected to decrease the efficiency of voter decision-making for Minority voters and obfuscation of interests is predicted to negatively affect both voter groups (Hypothesis 4a). This prediction can be seen as the mirror-image of Hypothesis 1a. On the aggregate level, any change to the traditional media environment is predicted to unavoidably lead to a drop in efficiency (Hypothesis 4b). This prediction is in line with Hypothesis 4b, which states that only the joint implementation of a microtargeting ban and mandatory disclosure of interests leads to an unambiguous improvement on the aggregate level.

We estimate a random effects model that is represented by the following equation:<sup>37</sup>

$$y_{ith} = \alpha + \beta_1 MD + \beta_2 PU + \beta_3 MU + \beta_4 Part + u_i + \epsilon_{ith}, \quad (14)$$

$$i = 1, \dots, 144, \quad t = 21, \dots, 40, \quad h = 1, 2.$$

As in equation (1),  $y_{ith}$  in (14) represents the outcome of one of our measures in round  $t$  of part  $h$  for subject  $i$ .

The intercept  $\alpha$  is the aggregate outcome rate in the PD-game in Part II of the experiment (i.e., the status quo). The variables  $MD$ ,  $PU$  and  $MU$  are treatment dummies which take on value one if the respective intervention is implemented. The estimates of the treatment effects ( $\beta_1$ ,  $\beta_2$  and  $\beta_3$ ) compared to the status quo ( $\alpha$ ) are reported in Table 7.<sup>38</sup> The binary variable  $Part$  is equal to one if an observation comes from Part I of the experiment and zero otherwise. Lastly, the model includes a subject-specific random effect  $u_i$  and error term  $\epsilon_{ith}$ .

**Microtargeting implementation.** We first consider the effects of a shift from the PD-game to the MD-game (see Tables 7-9). As expected (Hypothesis 4a), the acquired capacity of the interest group to send separate messages to the two voter groups only affects the payoffs of Minority voters (Table 8). The lower payoffs are caused by the changed

---

<sup>37</sup>We have also estimated subject-level random effects (ordered) probit models that generate similar estimates of the treatment effects but are harder to interpret than the (estimated) linear probability model.

<sup>38</sup>We only report marginal effects and leave out the regressions. The regression output for one of our measures (trust) can be found in Table 26 in the Appendix.

reporting behavior of the Majority interest group: being no longer restricted to report the same message as to Majority voters, the interest group significantly reduces the truthfulness of its messages directed to Minority voters (row 2 of Table 7). The increased ability of interest groups to influence election outcomes in their favor does, on average, not lead to a significant change in efficient and inefficient election outcomes: the rise in efficient election outcomes with a Majority interest group (estimate of 0.18) is undone by a comparable drop (estimate of  $-0.22$ ) in elections with a Minority interest group (rows 1 and 4 in Table 9).<sup>39</sup> Similarly, there is a slight but non-significant drop in election inefficiency if the interest group has type Majority and there is a sizeable increase in the case of a Minority interest group, which does not translate into a significant average effect (rows 2 and 5 of Table 9).<sup>40</sup>

**Obfuscation of interests.** In theory, a shift from the PD-game to the PU-game harms both Majority and Minority voters because they are no longer able to distinguish trustworthy from unreliable interest group types. Remarkably, we do not observe this effect in the experimental data. On average, there are no statistically significant effects on the payoffs from voting for both voter groups (rows 3 and 6 of Table 8). The absence of detrimental effects of obfuscation of interests is the result of excessive truth-telling by the Minority interest group in the PU-game (the actual average truth rate is 0.57, whereas the equilibrium prediction is 0; see Proposition 3. This excessive truthfulness is commonly observed in cheap talk experiments (e.g., Cai and Wang, 2006) and has been attributed to lying aversion. Our additional lying aversion task suggests, conservatively, that approximately 30% of our subjects have some aversion to lying (Table 22).<sup>41</sup> In line with our prediction, election efficiency decreases and election inefficiency increases if the interest group has type Majority (Hypothesis 4b and rows 1 and 2 of Table 9). None of the estimated coefficients for the Minority interest group (rows 4 and 5 of Table 9) are statistically significant.

---

<sup>39</sup>The estimated average effect on election efficiency is non-significant: the coefficient estimate is equal to 0.04 and the standard deviation equals 0.09.

<sup>40</sup>The estimated average effect is 0.07, with a standard deviation of 0.06.

<sup>41</sup>We believe that real interest groups are less likely to have an aversion to lying than the subjects in our experiment. Consequently, we expect that our estimated effects underestimate the true effects of obfuscation of interest on the efficiency of voting behavior.

**Microtargeting implementation and obfuscation of interests.** Our third and final intervention is a shift from the PD-game to the MU-game, which is expected to decrease the efficiency of voter decision-making (Hypothesis 4). On the individual level, the experimental data are in line with Hypothesis 4a: average voter payoffs decrease for both voter groups (rows 3 and 6 of Table 8). On the aggregate level, the support for our theoretical prediction (Hypothesis 4b) is milder. We find that election efficiency decreases and election inefficiency increases if the interest group has type Majority (rows 1 and 2 of Table 9) but none of the estimated coefficients for the Minority interest group are statistically significant (rows 4 and 5 of Table 9).

Based on the experimental findings reported in this section, we state the following key result.

**Result 4.**

- a. **Voter payoffs.** *Microtargeting implementation decreases the payoffs from voting for Minority voters but has no effect for Majority voters. Obfuscation of interests has no statistically significant effect on voter payoffs. Microtargeting implementation in combination with obfuscation of interests decreases the payoffs of both voter groups.*
- b. **Election outcome.** *On average, microtargeting implementation has no statistically significant effect on the efficiency of election outcomes. Obfuscation of interests, with or without microtargeting implementation, reduces the efficiency of election outcomes.*

Most importantly, Result 4 shows that the transition from a traditional media environment to a social media environment makes all voters worse off. In our experiment, both Majority and Minority voters failed to vote for their favorite party and cast votes for the opposing party more frequently due to the joint implementation of microtargeting technology and obfuscation of interests (Table 7), which confirms our theoretical prediction (Hypothesis 4a).

As expected (Hypothesis 4a), we find that the implementation of microtargeting technology only (negatively) affects Minority voters. However, we do not find that obfuscation of interests (on its own) decreases the efficiency of voter decision-making on the individual level, which is in contrast to our prediction (Hypothesis 4a and Result 4a). On the aggregate level, we find that obfuscation of interests, with or without microtargeting implementation, leads to less efficient and more inefficient election outcomes. The experimental data do not offer strong support for our prediction that any departure from the traditional media environment reduces the efficiency of aggregate voter decision-making (Hypothesis 4b): the average effects of microtargeting implementation are statistically insignificant.

Table 7: Marginal effects of microtargeting implementation and obfuscation of interests on voting behavior

Voter	Interest group	Vote choice	Status quo		Microtargeting implementation				Intervention				Combination				
			Coeff.	St. dev.	Coeff.	St. dev.	Sig.	Pred.	Obfuscation of interests		Obfuscation of interests		Coeff.	St. dev.	Sig.	Pred.	
1	Majority	Majority	Efficient	0.79	0.04	0.08	0.05	*	=	-0.13	0.07	*	-	-0.25	0.06	***	-
2		Inefficient	0.06	0.02	0.00	0.02	=	=	0.04	0.03	=	=	0.09	0.03	***	=	
3		Abstention	0.13	0.03	-0.07	0.04	*	=	0.10	0.06	*	+	0.19	0.06	***	+	
4	Minority	Majority	Efficient	0.25	0.05	-0.05	0.07	-	-	0.19	0.07	***	-	0.16	0.07	**	-
5		Inefficient	0.23	0.04	-0.05	0.06	+	+	0.14	0.07	**	+	0.12	0.07	*	+	
6		Abstention	0.50	0.07	0.13	0.09	+	+	-0.30	0.09	***	+	-0.25	0.10	**	+	
7	Minority	Majority	Efficient	0.75	0.05	-0.49	0.07	***	=	-0.08	0.09	-	-	-0.50	0.07	***	-
8		Inefficient	0.06	0.02	0.16	0.05	***	=	0.04	0.04	+	+	0.27	0.05	***	+	
9		Abstention	0.19	0.04	0.33	0.08	***	=	0.04	0.08	+	+	0.24	0.07	***	+	
10	Minority	Majority	Efficient	0.30	0.08	0.43	0.12	***	+	0.06	0.11	-	-	0.02	0.12		=
11		Inefficient	0.25	0.07	-0.11	0.10	-	-	0.00	0.10	+	+	-0.05	0.09		=	
12		Abstention	0.44	0.06	-0.32	0.09	***	-	-0.03	0.10	-	-	0.04	0.11		=	

*Notes.* This table reports the estimated outcome rates of the status quo as well as the marginal effects of the interventions microtargeting implementation, obfuscation of interests and a combination of these two interventions. The marginal effects are estimated using model 14 with robust standard errors, \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The symbols '+', '-', '=', and '?' in the column Pred. denote that the predicted marginal effect is, respectively, positive, negative, equal to zero, and ambiguous.

Table 8: Marginal effects of a microtargeting implementation and obfuscation of interests on voter payoffs

	Voter	Interest group	Status quo		Microtargeting implementation				Intervention				Combination			
			Coeff.	St. dev.	Coeff.	St. dev.	Sig.	Pred.	Coeff.	St. dev.	Sig.	Pred.	Coeff.	St. dev.	Sig.	Pred.
1	Majority	Majority	81.19	1.51	2.73	1.97		=	-5.00	2.71	*	-	-9.86	2.41	***	-
2		Minority	60.21	2.31	0.15	2.90		=	1.16	3.71		-	-0.34	3.43		-
3		All	73.95	1.45	2.13	1.78		=	-3.05	2.66		-	-6.20	2.41	***	-
4	Minority	Majority	80.20	2.06	-19.92	2.90	***	=	-3.93	3.33		=	-23.18	3.03	***	=
5		Minority	60.68	4.21	16.08	6.02	***	=	2.37	5.76		=	2.76	5.48		=
6		All	73.65	2.16	-7.61	2.96	**	=	-1.78	3.11		=	-13.47	2.86	***	=

*Notes.* This table reports the estimated outcome rates of the status quo as well as the marginal effects of the interventions microtargeting implementation, obfuscation of interests and a combination of these two interventions on voter payoffs. The marginal effects are estimated using model 14 with robust standard errors, \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The symbols '+', '-', '=', and '?' in the column Pred. denote that the predicted marginal effect is, respectively, positive, negative, equal to zero, and ambiguous.

Table 9: Marginal effects of microtargeting implementation and obfuscation of interests on election outcomes

	Interest group	Election outcome	Status quo		Microtargeting implementation				Intervention				Combination			
			Coeff.	St. dev.	Coeff.	St. dev.	Sig.	Pred.	Coeff.	St. dev.	Sig.	Pred.	Coeff.	St. dev.	Sig.	Pred.
1	Majority	Efficient	0.74	0.04	0.18	0.05	***	=	-0.14	0.07	**	-	-0.14	0.06	**	-
2		Inefficient	0.08	0.03	-0.02	0.04		=	0.10	0.04	**	+	0.12	0.05	***	+
3		Tie	0.18	0.04	-0.16	0.04	***	=	0.03	0.06		+	0.02	0.05		+
4	Minority	Efficient	0.35	0.08	-0.22	0.09	**	-	0.05	0.12		-	0.02	0.11		-
5		Inefficient	0.27	0.05	0.28	0.09	***	+	0.15	0.10		+	0.12	0.08		+
6		Tie	0.38	0.05	-0.07	0.07		-	-0.20	0.08	**	-	-0.13	0.09		-

*Notes.* This table reports the estimated outcome rates of the status quo as well as the marginal effects of the interventions microtargeting implementation, obfuscation of interests and a combination of these two interventions. The marginal effects are estimated using model 14 with robust standard errors, \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The symbols '+', '-', '=', and '?' in the column Pred. denote that the predicted marginal effect is, respectively, positive, negative, equal to zero, and ambiguous.

## A.7 Analysis of the social media environment: a between-subject approach

In this Section, we present the results of our main analysis with a *between-subject* instead of a *within-subject* approach.<sup>42</sup> This robustness check confirms our main finding: mandatory disclosure of interests, with or without a microtargeting ban, gives rise to more efficient aggregate voter decision-making. Only the combination of disclosure and a microtargeting ban effectively prevents election manipulation. Our findings on the individual level hold with one exception: in contrast to our prediction, mandatory disclosure of interests does *not* lead to a statistically significant improvement of a Minority voter's payoffs from voting.

<sup>42</sup>For the between-subject analysis, we make use of all Part-I games and exclude data from Part II because it may have been influenced by behavior in Part I of the experiment.

This deviation is caused by the Minority voter’s behavior if the interest group has type Majority. While our theory predicts that there is no difference in trust in the MU-game and in the MD-game, the Minority voter displays a significantly higher level of trust in the former than in the latter game (0.49 vs. 0.34; row 6 of Table 10).<sup>43</sup> The lower level of trust with disclosure results in lower average payoffs from voting, which are not undone by the improved information transmission between a Minority interest group and a Minority voter (rows 4 to 6 of Table 12).

The results of the between-subject analysis are presented in Tables 10-13.

Table 10: Marginal effects of a microtargeting ban and disclosure of interests on truth-telling and trust

	Measure	Interaction		Status quo		Microtargeting ban				Intervention				Combination			
		Interest group	Voters	Coeff.	St. dev.	Coeff.	St. dev.	Sig.	Pred.	Coeff.	St. dev.	Sig.	Pred.	Coeff.	St. dev.	Sig.	Pred.
1	Truth	Majority	Majority	0.89	0.02	-0.07	0.04	*	=	0.06	0.03	**	=	0.03	0.03		=
2			Minority	0.62	0.04	0.21	0.05	***	+	-0.09	0.06		=	0.31	0.04	***	+
3		Minority	Majority	0.48	0.06	0.03	0.10		=	0.12	0.08		+	0.16	0.07	**	+
4			Minority	0.83	0.04	-0.31	0.09	***	-	-0.07	0.07		+	-0.18	0.06	***	=
5	Trust	Majority	Majority	0.66	0.03	-0.04	0.05		=	0.22	0.04	***	=	0.18	0.03	***	=
6			Minority	0.49	0.03	0.05	0.06		+	-0.15	0.06	***	=	0.25	0.04	***	+
7		Minority	Majority	0.69	0.03	-0.08	0.05		=	-0.38	0.05	***	-	-0.28	0.04	***	-
8			Minority	0.51	0.04	0.11	0.06	*	+	0.27	0.06	***	+	0.01	0.06		=

*Notes.* This table reports the estimated outcome rates of the status quo as well as the marginal effects of the interventions microtargeting ban, disclosure of interests and a combination of these two interventions. The marginal effects are estimated using model 1, with the omission of the variable Part, with robust standard errors, \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The symbols ‘+’, ‘-’, ‘=’, and ‘?’ in the column Pred. denote that the predicted marginal effect is, respectively, positive, negative, equal to zero, and ambiguous.

<sup>43</sup>Our results of the main analysis did not show a drop in trust due to disclosure (0.41 vs. 0.41; row 6 of Table 3

Table 11: Marginal effects of a microtargeting ban and disclosure of interests on voting behavior

Voter	Interest group	Vote choice	Status quo		Microtargeting ban				Intervention				Combination				
			Coeff.	St. dev.	Coeff.	St. dev.	Sig.	Pred.	Coeff.	St. dev.	Sig.	Pred.	Coeff.	St. dev.	Sig.	Pred.	
1	Majority	Majority	Efficient	0.60	0.03	-0.04	0.05	=	0.24	0.04	***	+	0.18	0.04	***	+	
2			Inefficient	0.13	0.01	0.04	0.03	=	-0.05	0.02	**	-	-0.03	0.02	***	-	
3			Abstention	0.27	0.02	0.00	0.04	=	-0.18	0.04	***	-	-0.15	0.03	***	-	
4	Minority	Minority	Efficient	0.36	0.03	0.08	0.05	=	-0.06	0.05		=	-0.01	0.04		=	
5			Inefficient	0.40	0.03	-0.10	0.05	**	=	-0.16	0.04	***	-	-0.14	0.03	***	-
6			Abstention	0.25	0.02	0.03	0.04	=	0.22	0.05	***	+	0.15	0.04	***	+	
7	Minority	Majority	Efficient	0.39	0.03	0.11	0.05	**	+	-0.13	0.04	***	=	0.30	0.04	***	+
8			Inefficient	0.25	0.02	-0.09	0.04	**	-	0.02	0.04		=	-0.10	0.03	***	-
9			Abstention	0.36	0.03	-0.02	0.05		-	0.11	0.06	**	=	-0.20	0.04	***	-
10	Minority	Minority	Efficient	0.46	0.04	-0.03	0.07	=	0.18	0.07	***	+	-0.01	0.05		=	
11			Inefficient	0.19	0.03	0.12	0.06	**	+	0.04	0.05		=	0.07	0.04	*	=
12			Abstention	0.35	0.04	-0.09	0.06		-	-0.22	0.05	***	-	-0.06	0.05		=

*Notes.* This table reports the estimated outcome rates of the status quo as well as the marginal effects of the interventions microtargeting ban, disclosure of interests and a combination of these two interventions. The marginal effects are estimated using model 1, with the omission of the variable Part, with robust standard errors, \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The symbols '+', '-', '=', and '?' in the column Pred. denote that the predicted marginal effect is, respectively, positive, negative, equal to zero, and ambiguous.

Table 12: Marginal effects of a microtargeting ban and disclosure of interests on voter payoffs

Voter	Interest group	Status quo		Microtargeting ban				Intervention				Combination					
		Coeff.	St. dev.	Coeff.	St. dev.	Sig.	Pred.	Coeff.	St. dev.	Sig.	Pred.	Coeff.	St. dev.	Sig.	Pred.		
1	Majority	Majority	74.00	0.71	-3.01	1.29	**	=	8.78	1.08	***	=	5.98	0.97	***	=	
2			Minority	58.02	1.25	5.46	2.13	**	=	3.04	1.92		=	4.49	1.67	***	=
3			All	68.67	0.66	-0.19	1.14		=	6.86	1.06	***	=	5.48	0.90	***	=
4	Minority	Majority	63.44	1.13	5.95	1.87	***	=	-4.11	1.87	**	=	12.37	1.55	***	=	
5			Minority	67.43	1.54	-5.10	2.88	*	=	3.38	2.86		=	-2.85	2.28		=
6			All	64.77	0.91	2.27	1.58		=	-1.62	1.60		=	7.30	1.30	***	=

*Notes.* This table reports the estimated outcome rates of the status quo as well as the marginal effects of the interventions microtargeting ban, disclosure of interests and a combination of these two interventions on voter payoffs. The marginal effects are estimated using model 1, with the omission of the variable Part, with robust standard errors, \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The symbols '+', '-', '=', and '?' in the column Pred. denote that the predicted marginal effect is, respectively, positive, negative, equal to zero, and ambiguous.

Table 13: Marginal effects of a microtargeting ban and disclosure of interests on election outcomes

Interest group	Election outcome	Intervention														
		Status quo		Microtargeting ban				Disclosure				Combination				
		Coeff.	St. dev.	Coeff.	St. dev.	Sig.	Pred.	Coeff.	St. dev.	Sig.	Pred.	Coeff.	St. dev.	Sig.	Pred.	
1	Majority	Efficient	0.65	0.03	-0.11	0.05	**	-	0.22	0.04	***	+	0.04	0.04		+
2		Inefficient	0.17	0.02	0.04	0.04		=	-0.10	0.03	***	-	-0.04	0.02	*	-
3		Tie	0.19	0.02	0.07	0.04	**	+	-0.12	0.02	***	-	0.00	0.03		-
4	Minority	Efficient	0.35	0.04	0.10	0.07		=	-0.08	0.07		=	0.01	0.06		=
5		Inefficient	0.48	0.05	-0.17	0.08	**	-	-0.01	0.07		?	-0.17	0.06	***	-
6		Tie	0.17	0.03	0.07	0.05		+	0.10	0.05	*	?	0.16	0.04	***	+

*Notes.* This table reports the estimated outcome rates of the status quo as well as the marginal effects of the interventions microtargeting ban, disclosure of interests and a combination of these two interventions. The marginal effects are estimated using model 1, with the omission of the variable Part, with robust standard errors, \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The symbols '+', '-', '=', and '?' in the column Pred. denote that the predicted marginal effect is, respectively, positive, negative, equal to zero, and ambiguous.

## A.8 Analysis individual games

Based on the most informative equilibria of the four games, we derive Hypothesis 5 regarding the efficiency of voting behavior within the individual games

**Hypothesis 5** (Efficiency within games).

- a. **Voter payoffs.** *In Public games, there is no difference in the payoffs from voting for Minority and Majority voters. In Microtargeting games, Minority voters have lower payoffs than Majority voters.*
- b. **Election outcome.** *Aggregate voter decision-making is on average less efficient in games with a Minority interest group than in games with a Majority interest group.*

Our predictions regarding the efficiency of voter decision-making follow directly from the predicted actions of the interest group (truth) and voters (trust), which jointly determine how well-informed (predicted) voting actions are. In Public games, a Minority voter receives the same information and is expected to display the same level of trust as Majority voters.

Hence, a public communication technology is expected to give rise to the same payoffs for Minority and Majority voters, regardless of the disclosure regime in place (Hypothesis 5a). This equality is expected to disappear if the interest group gets access to a microtargeting technology (Hypothesis 5a). In the equilibrium of the MD-game, truth and trust occur with a higher probability if the interest group type and voter type match. Consequently, we expect that voter decision-making yields a lower (higher) payoff for a Minority voter than a Majority voter if the interest group has type Majority (Minority). Also in the MU-game, Minority voters are expected to be worse (better) off with a Majority (Minority) interest group compared to Majority voters. On the one hand, Minority voters do not receive (consistently) truthful information from a Majority interest group and therefore have a lower efficiency of voter decision-making. On the other hand, Minority voters are not led astray by a Minority interest group, which enables them to choose more-efficient voting actions than Majority voters. Since it is most likely that the interest group has the Majority type, the average voter payoffs are expected to be lower for a Minority voter than a Majority voter in both Microtargeting games.

Our final prediction about efficiency in individual games concerns the aggregate outcomes. Whereas a Majority interest group obtains a higher payoff from votes for the efficient election winner, a Minority interest group has an incentive to persuade voters to vote for the party that is inefficient on aggregate. Hence, we expect that an election with a Minority interest group is less likely to have the efficient outcome than an election with a Majority interest group (Hypothesis 5b).

For our study of the individual games, we estimate a linear random effects model for each of our measures.<sup>44</sup> All our models have the following form.

---

<sup>44</sup>For ease of interpretation of the coefficients, we present a linear probability model. Our estimations of subject-level random effects (ordered) probit model and a three-level mixed effects probit model produce qualitatively similar results.

$$y_{it} = \alpha + \beta Type_{it} + u_i + \epsilon_{it}, \tag{15}$$

$$i = 1, \dots, 432, \quad t = 21, \dots, 40.$$

In this model, an observation for one of our outcome variables is given by  $y_{it}$ , with subject indicator  $i$  and round indicator  $t$ . The intercept  $\alpha$  represents the aggregate outcome rate for a Majority (interest group or voter) type. The dummy variable  $Type_{it}$  is equal to one if subject  $i$  has type Minority in round  $t$  and zero otherwise. Furthermore, (15) includes a subject-specific random effect  $u_i$  and error term  $\epsilon_{it}$ .

Our coefficient of interest is  $\beta$ , which quantifies the difference in the outcome rate between a Minority and a Majority (interest group or voter) type. As in Section 4, we restrict our analysis to the second half (rounds 21-40) of the Part-I games. Table 14 displays our results regarding interest group behavior (truth) in each of the four games.<sup>45</sup> The table contains summary statistics (the mean of the session averages and the standard deviation of the mean) by interest group type, as well as the observed and predicted marginal effect of an interest group type change (from Majority to Minority) on the probability of truth being equal to one.<sup>46</sup> This marginal effect corresponds to the coefficient  $\beta$  in Equation 15. We present our results about individual voter behavior (Tables 15 and 16) and aggregate election outcomes (Table 18) in the same way.<sup>47</sup>

---

<sup>45</sup>Due to space constraints, we only present the estimated marginal effects and omit the random effects panel regressions. By means of illustration, we present the regression output for one of our measures (trust) in Table 24 in the Appendix.

<sup>46</sup>In addition to this table, we present the observed interest group strategy profiles in Table ??.

<sup>47</sup>The distribution of individual voting behavior profiles is presented in Table ??.

Table 14: Summary statistics and marginal effects: truth-telling by interest groups - Part I

1			Majority interest group		Minority interest group		Marginal effect			
2	Game	Measure	Mean	St. Dev.	Mean	St. Dev.	Coeff.	S. E.	Sig.	Pred.
3	MD	Truth to Majority voter	0.95	0.03	0.60	0.11	-0.35	0.06	***	-
4		Truth to Minority voter	0.52	0.11	0.76	0.19	0.23	0.08	***	+
5	MU	Truth to Majority voter	0.89	0.10	0.48	0.19	-0.41	0.06	***	-
6		Truth to Minority voter	0.62	0.14	0.82	0.10	0.21	0.05	***	=
7	PD	Truth	0.93	0.06	0.65	0.16	-0.28	0.05	***	-
8	PU	Truth	0.83	0.12	0.52	0.32	-0.31	0.09	***	-

*Notes.* This table displays summary statistics and marginal effects of truth-telling by interest groups for each of the four games. We report the mean of the session averages and the standard deviation of the mean for each interest group type. The marginal effects (differences between the minority interest group type and the majority interest group type) are estimated using model 15 with cluster-robust standard errors, \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The symbols ‘+’, ‘-’, and ‘=’ in the column Pred. denote that the predicted marginal effect is, respectively, positive, negative, and equal to zero.

Table 15: Summary statistics and marginal effects: trust of voters - Part I

			Majority voter		Minority voter		Marginal effect				
	Game	Measure	Interest group	Mean	St. Dev.	Mean	St. Dev.	Coeff.	S. E.	Sig.	Pred.
1	MD	Trust	Majority	0.88	0.05	0.35	0.15	-0.53	0.06	***	-
2			Minority	0.30	0.12	0.77	0.11	0.49	0.07	***	+
3	MU	Trust	Majority	0.67	0.09	0.49	0.10	-0.17	0.03	***	-
4			Minority	0.69	0.11	0.51	0.15	-0.18	0.04	***	-
5	PD	Trust	Majority	0.84	0.08	0.75	0.09	-0.09	0.02	***	=
6			Minority	0.41	0.12	0.52	0.17	0.08	0.04	*	=
7	PU	Trust	Majority	0.61	0.18	0.54	0.24	-0.07	0.04	*	=
8			Minority	0.62	0.14	0.63	0.20	0.01	0.06		=

*Notes.* This table displays summary statistics and marginal effects of trust by voters for each of the four games. We report the mean of the session averages and the standard deviation of the mean for each voter type. The marginal effects (differences between the minority voter type and the majority voter type) are estimated using model 15 with cluster-robust standard errors, \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The symbols ‘+’, ‘-’, and ‘=’ in the column Pred. denote that the predicted marginal effect is, respectively, positive, negative, and equal to zero.

Table 16: Summary statistics and marginal effects: efficiency of individual voter decision-making - Part I

	Game	Vote choice	Interest group	Majority voter		Minority voter		Marginal effect			
				Mean	St. Dev.	Mean	St. Dev.	Coeff.	S. E.	Sig.	Pred.
1	MD	Efficient	Majority	0.84	0.06	0.27	0.09	-0.57	0.04	***	-
2			Minority	0.29	0.10	0.62	0.21	0.34	0.07	***	+
3			All	0.66	0.07	0.39	0.12	-0.27	0.03	***	-
4		Inefficient	Majority	0.08	0.03	0.27	0.07	0.19	0.03	***	+
5			Minority	0.23	0.07	0.23	0.16	0.00	0.05		-
6			All	0.13	0.04	0.26	0.08	0.13	0.02	***	+
7		Abstention	Majority	0.08	0.05	0.46	0.11	0.37	0.05	***	+
8			Minority	0.48	0.07	0.15	0.10	-0.33	0.06	***	-
9			All	0.22	0.05	0.36	0.10	0.14	0.03	***	+
10	MU	Efficient	Majority	0.60	0.13	0.39	0.07	-0.21	0.04	***	-
11			Minority	0.36	0.15	0.46	0.14	0.10	0.04	***	+
12			All	0.52	0.09	0.41	0.07	-0.11	0.03	***	-
13		Inefficient	Majority	0.13	0.07	0.25	0.06	0.12	0.03	***	+
14			Minority	0.39	0.16	0.19	0.07	-0.21	0.03	***	-
15			All	0.21	0.05	0.23	0.05	0.02	0.02		+
16		Abstention	Majority	0.27	0.09	0.36	0.10	0.09	0.03	***	+
17			Minority	0.25	0.10	0.35	0.11	0.11	0.04	***	+
18			All	0.26	0.07	0.36	0.09	0.09	0.02	***	+
19	PD	Efficient	Majority	0.78	0.10	0.69	0.09	-0.09	0.03	***	=
20			Minority	0.36	0.12	0.45	0.16	0.08	0.05	*	=
21			All	0.64	0.09	0.61	0.09	-0.03	0.03		=
22		Inefficient	Majority	0.10	0.05	0.15	0.05	0.05	0.02	**	=
23			Minority	0.25	0.11	0.27	0.09	0.01	0.04		=
24			All	0.15	0.06	0.19	0.04	0.04	0.02	**	=
25		Abstention	Majority	0.12	0.07	0.16	0.08	0.04	0.02	**	=
26			Minority	0.40	0.08	0.29	0.14	-0.07	0.03	**	=
27			All	0.21	0.07	0.20	0.09	-0.01	0.02		=
28	PU	Efficient	Majority	0.55	0.20	0.49	0.25	-0.06	0.04		=
29			Minority	0.45	0.19	0.42	0.26	-0.03	0.07		=
30			All	0.52	0.19	0.47	0.24	-0.05	0.04		=
31		Inefficient	Majority	0.17	0.08	0.16	0.08	-0.01	0.04		=
32			Minority	0.30	0.17	0.31	0.18	0.01	0.06		=
33			All	0.21	0.10	0.21	0.10	0.00	0.03		=
34		Abstention	Majority	0.28	0.13	0.35	0.17	0.07	0.04	*	=
35			Minority	0.25	0.07	0.28	0.17	0.01	0.05		=
36			All	0.27	0.11	0.32	0.16	0.05	0.03		=

*Notes.* This table displays summary statistics and marginal effects of the efficiency of individual voter decision-making for each of the four games. We report the mean of the session averages and the standard deviation of the mean for each voter type. The marginal effects (differences between the minority voter type and the majority voter type) are estimated using model 15 with cluster-robust standard errors, \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The symbols '+', '-', and '=' in the column Pred. denote that the predicted marginal effect is, respectively, positive, negative, and equal to zero.

Table 17: Summary statistics and marginal effects: voter payoffs - Part I

Game	Interest group	Majority voter		Minority voter		Marginal effect				
		Mean	St. Dev.	Mean	St. Dev.	Coeff.	S. E.	Sig.	Pred.	
1	MD	Majority	82.78	2.68	59.33	3.75	-23.39	1.73	***	—
2		Minority	61.06	5.09	70.80	11.05	10.19	3.22	***	+
3		All	75.54	3.00	63.15	5.46	-12.39	1.55	***	—
4	MU	Majority	74.00	6.05	63.44	2.65	-10.23	1.76	***	—
5		Minority	58.02	9.54	67.43	5.97	9.74	1.93	***	+
6		All	68.67	3.69	64.77	2.40	-3.91	1.31	***	—
7	PD	Majority	79.98	4.26	75.81	3.59	-4.21	1.32	***	=
8		Minority	62.51	6.84	64.58	6.91	2.10	2.49		=
9		All	74.15	4.27	72.06	3.08	-2.09	1.27		=
10	PU	Majority	70.99	8.50	69.39	9.95	-1.62	2.19		=
11		Minority	63.47	11.11	62.33	12.80	-1.21	3.76		=
12		All	68.49	8.87	67.03	10.50	-1.45	1.93		=

*Notes.* This table displays summary statistics and marginal effects of voter payoffs for each of the four games. We report the mean of the session averages and the standard deviation of the mean for each voter type. The marginal effects (differences between the minority voter type and the majority voter type) are estimated using model 15 with cluster-robust standard errors, \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The symbols ‘+’, ‘-’, and ‘=’ in the column Pred. denote that the predicted marginal effect is, respectively, positive, negative, and equal to zero.

Table 18: Summary statistics and marginal effects: efficiency of aggregate voter decision-making - Part I

Game	Election outcome	Majority interest group		Minority interest group		Marginal effect				
		Mean	St. Dev.	Mean	St. Dev.	Coeff.	S. E.	Sig.	Pred.	
1	MD	Efficient	0.86	0.06	0.27	0.10	-0.60	0.05	***	—
2		Inefficient	0.07	0.03	0.47	0.10	0.40	0.06	***	+
3		Tie	0.07	0.03	0.27	0.14	0.20	0.05	***	+
4	MU	Efficient	0.65	0.10	0.35	0.18	-0.30	0.05	***	—
5		Inefficient	0.17	0.09	0.48	0.17	0.31	0.05	***	+
6		Tie	0.19	0.06	0.17	0.12	-0.01	0.03		=
7	PD	Efficient	0.69	0.12	0.36	0.15	-0.33	0.04	***	—
8		Inefficient	0.13	0.05	0.31	0.12	0.19	0.04	***	+
9		Tie	0.19	0.10	0.33	0.11	0.14	0.04	***	+
10	PU	Efficient	0.54	0.18	0.45	0.13	-0.09	0.07		—
11		Inefficient	0.20	0.10	0.31	0.16	0.10	0.07		+
12		Tie	0.26	0.11	0.24	0.07	-0.02	0.06		=

*Notes.* This table displays summary statistics and marginal effects of the efficiency of aggregate voter decision-making for each of the four games. We report the mean of the session averages and the standard deviation of the mean for each interest group type. The marginal effects (differences between the minority interest group type and the majority interest group type) are estimated using model 15 with cluster-robust standard errors, \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The symbols ‘+’, ‘-’, and ‘=’ in the column Pred. denote that the predicted marginal effect is, respectively, positive, negative, and equal to zero.

**MD-game.** In line with our theoretical prediction, we find that a Minority interest group is, in comparison to a Majority interest group, significantly *less* truthful to a Majority voter (estimated marginal effect of  $-0.35$ ; row 1 of Table 14) and significantly *more* truthful to a Minority voter (estimated coefficient of  $0.23$ ; row 2 of Table 14). Considering voters' actions (Table 15), we find that voters best respond to this reporting behavior: a Minority voter is significantly *less* likely to trust a Majority interest group (estimated coefficient of  $-0.53$ ; row 1) and significantly *more* likely to trust a Minority interest group (estimated marginal effect of  $0.49$ ; row 2) than a Majority voter. On average, the differences in the efficiency of voting actions of Majority and Minority voters confirm our theoretical predictions: compared to a Majority voter, a Minority voter's average vote efficiency is significantly *lower* (row 3 of Table 16) and average vote inefficiency is significantly *higher* (row 6), which leads to lower average payoffs from voting for the Minority voter (row 3 of Table 17). As can be seen from Table 18, our findings on aggregate voter decision-making are consistent with our theory: election efficiency is significantly lower (row 1) and election inefficiency is significantly higher (row 2) if the interest group has type Minority.

**MU-game.** We expect that a Majority interest group reports truthfully and a Minority interest group lies to Majority voters. Consistent with this prediction, we see a sizeable and significant negative marginal effect for truth to Majority voters (row 3 of Table 14). The comparative static prediction about communication to the Minority voter is refuted. Our theory implies that no information transmission can take place between an interest group and a Minority voter if the type of the interest group is undisclosed and messages are micro-targeted.<sup>48</sup> Hence, the predicted mean of truth to a Minority voter is  $0.5$  for both a Minority and a Majority interest group. We observe, however, that both interest group types are more truthful than predicted and that a Minority interest group is significantly more truthful to

---

<sup>48</sup>See footnote 22 for an explanation.

a Minority voter than a Majority interest group (row 4 of Table 14).<sup>49</sup> Trust is significantly lower for Minority voters than for Majority voters (rows 3 and 4 of Table 15). This behavior is prescribed by the most informative equilibrium but is inconsistent with the observed high level of truthfulness of the messages. Since all voters receive informative messages (the probability of receiving a truthful message is higher than 0.5), both Majority and Minority voters best respond by trusting a message received, conditional on non-abstention.

Compared to Majority voters, a Minority voter has a significantly lower average vote efficiency (row 12 of Table 16), which is predicted by our theory. There is no statistically significant difference in average vote inefficiency between Majority and Minority voters (row 15 of Table 16). Considering vote inefficiency by interest group type, it becomes clear that a Minority voter has (as predicted) a significantly higher vote inefficiency with a Majority interest group type and a significantly lower vote inefficiency with the Minority interest group type than a Majority voter (rows 13 and 14 of Table 16). In line with our theory, a Minority voter obtains a significantly lower average voter payoff than a Majority voter (row 6 of Table 17). Aggregate effects are as expected: election efficiency (inefficiency) is significantly lower (higher) with a Minority interest group than with a Majority interest group (rows 4 and 5 of Table 18).

**PD-game.** According to our theory, the (mis)match of the interest group’s type and the Majority voters determines whether the true state of the world is reported in Public games. Indeed, the average truth is significantly lower for the Minority interest group than the Majority interest group in the PD-game (row 5 of Table 14). In contrast to our prediction, Minority voters have significantly lower trust than Majority voters if the interest group has

---

<sup>49</sup>The predicted equilibrium outcome of no information transmission to a Minority voter can be achieved if the interest group follows a mixing strategy in which the message sent is sometimes truthful and sometimes a lie but always independent of the true state. Our lying aversion task suggests that a sizeable share of our subjects is unwilling to tell a lie, which increases, *ceteris paribus*, the informativeness of a message received by a Minority voter. Then, it is a best response of a Minority voter to trust the message received, which implies that a (non-lying averse) subject could profitably deviate from the mixing strategy by reporting the true state in the role of Minority interest group and the ‘wrong’ state in the role of Majority interest group. The observed individual interest group strategy profiles, reported in Table ??, suggest that some subjects have followed this deviation from the equilibrium strategy, which could explain why we observe differences in reporting behavior between Majority and Minority interest groups.

type Majority and vice versa with a Minority interest group (rows 5 and 6 of Table 15). This asymmetry in trust for the two voter types is consistent with the findings of Battaglini and Makarov (2014), who established that “It appears that receivers pay more attention to their “direct relationship” with the sender and partially ignore the other receiver”.

It is important to note that the differences in voting behavior between Majority and Minority voters do not translate into large efficiency differences. For the average vote efficiency, the estimated marginal effect is negative but statistically insignificant (row 21 of Table 16). For the average vote inefficiency, the difference between Minority and Majority voters is statistically significant but small (estimated difference of 0.04; row 24 of Table 16). As predicted, there is no statistically significant difference in the average payoffs from voting for Minority and Majority voters (row 9 of Table 17). The data confirm our predictions regarding the efficiency of aggregate voter decision-making: election efficiency is significantly lower and election inefficiency is significantly higher with a Minority interest group than with a Majority interest group (rows 5 and 6 of Table 18).

**PU-game.** Our experimental findings confirm predicted interest group behavior: a Minority interest group is significantly less truthful than a Majority interest group (estimated marginal effect of  $-0.31$ ; row 6 of Table 14). For trust, we find a similar deviation from the equilibrium as in the PD-game: a Minority voter has lower average trust than a Majority voter (estimated difference of  $-0.05$ ;  $p < 0.10$ ).<sup>50</sup> Again, we do not find support for large efficiency differences between the Minority and Majority voters: none of the differences in vote efficiency and vote inefficiency of the voter types are significant (rows 30 and 33 of Table 16). Consequently, voter payoffs are not significantly different for Majority and Minority voters (rows 10 to 12 of Table 17). Surprisingly, we find that election efficiency and inefficiency are not significantly different for Minority and Majority interest groups (rows

---

<sup>50</sup>Considering trust by interest group type, we find that a Minority voter has lower trust than a Majority voter with a Majority interest group (estimated marginal effect of  $-0.08$ ) but that there are no statistically significant differences if the interest group has type Minority (rows 7 and 8 of Table 15. This difference in marginal effects depending on interest group types has arisen by chance (recall that voters do not observe the interest group type in the PU-game). In the PU-game in Part II of the experiment, Minority voters have lower trust than Majority voters, regardless of the interest group type.)

10 and 11 of Table 18). This finding does not, however, appear to be robust. Considering the PU-games in Part II of the experiment, we find negative marginal effects for election efficiency (estimate of  $-0.18$ ;  $p < 0.01$ ) and a positive marginal effect for election inefficiency (estimate of  $0.25$ ;  $p < 0.01$ ), which is in line with our predictions.

Overall, the observed efficiency of voter decision-making within the four games is reasonably close to our predictions stated in Hypothesis 5. We derive the following result based on the discussion of the individual games above.

**Result 5.**

- a. **Voter payoffs.** In Public games, voter payoffs are similar for Minority and Majority voters. In Microtargeting games, Minority voters have lower payoffs than Majority voters.*
  
- b. **Election efficiency.** The efficiency of aggregate voter decision-making is lower in games with a Minority interest group than in games with a Majority interest group, with the exception of the PU-game.*

In line with our theory, the efficiency of voter decision-making is similar for Minority and Majority voters in games with public communication (Hypothesis 5a and Result 4a). This equality disappears in communication environments with microtargeting technology: as predicted, Minority voters are on average less likely to choose efficient and more likely to choose inefficient voting actions than Majority voters (Hypothesis 5a and Result 4a). These findings are consistent with the experimental results on information transmission in cheap talk games with public and private communication and disclosed sender types by Battaglini and Makarov (2014) and Drugov et al. (2017). Our set-up has also allowed us to study games with *undisclosed* sender types. Qualitatively, our results regarding the efficiency of individual-level voter decision-making are the same in environments with and without disclosure of the sender's interests.

Our theory performs well on the aggregate level: election outcomes were less likely to be efficient and more likely to be inefficient with a Minority interest group than a Majority interest group in three of the four games (Hypothesis 5b and Result 4b).

## A.9 Summary statistics of Part II

In this section, we report the summary statistics of the Part-II games.

Table 19: Summary statistics: truth-telling by interest groups - Part II

Game	Measure	Majority interest group		Minority interest group	
		Mean	St. Dev.	Mean	St. Dev.
MD	Truth to Majority voter	0.94	0.05	0.59	0.19
	Truth to Minority voter	0.57	0.04	0.93	0.07
MU	Truth to Majority voter	0.89	0.07	0.43	0.14
	Truth to Minority voter	0.55	0.12	0.78	0.17
PD	Truth	0.95	0.06	0.67	0.11
PU	Truth	0.84	0.15	0.48	0.21

*Notes.* This table displays summary statistics of truth-telling by interest groups for each of the four games in Part II of the experiment. We report the mean of the session averages and the standard deviation of the mean for each interest group type.

Table 20: Summary statistics: efficiency of individual voter decision-making - Part II

	Game	Vote choice	Interest group	Majority voter		Minority voter	
				Mean	St. Dev.	Mean	St. Dev.
1	MD	Efficient	Majority	0.90	0.09	0.27	0.07
2			Minority	0.27	0.10	0.73	0.14
3			All	0.69	0.08	0.42	0.06
4		Inefficient	Majority	0.06	0.04	0.23	0.03
5			Minority	0.20	0.06	0.12	0.12
6			All	0.11	0.04	0.19	0.05
7		Abstention	Majority	0.04	0.05	0.50	0.10
8			Minority	0.53	0.12	0.16	0.09
9			All	0.21	0.06	0.39	0.07
10	MU	Efficient	Majority	0.57	0.12	0.31	0.09
11			Minority	0.34	0.09	0.40	0.15
12			All	0.49	0.10	0.34	0.09
13		Inefficient	Majority	0.15	0.08	0.27	0.09
14			Minority	0.38	0.09	0.15	0.05
15			All	0.22	0.07	0.23	0.07
16		Abstention	Majority	0.28	0.08	0.42	0.11
17			Minority	0.28	0.07	0.45	0.16
18			All	0.28	0.08	0.43	0.10
19	PD	Efficient	Majority	0.83	0.13	0.77	0.15
20			Minority	0.34	0.10	0.38	0.19
21			All	0.67	0.11	0.64	0.12
22		Inefficient	Majority	0.05	0.04	0.06	0.04
23			Minority	0.23	0.06	0.24	0.11
24			All	0.11	0.02	0.12	0.04
25		Abstention	Majority	0.11	0.10	0.16	0.12
26			Minority	0.44	0.14	0.37	0.17
27			All	0.22	0.10	0.23	0.12
28	PU	Efficient	Majority	0.60	0.21	0.56	0.22
29			Minority	0.37	0.17	0.33	0.19
30			All	0.52	0.18	0.49	0.19
31		Inefficient	Majority	0.12	0.06	0.15	0.08
32			Minority	0.41	0.13	0.32	0.13
33			All	0.22	0.06	0.21	0.07
34		Abstention	Majority	0.28	0.18	0.28	0.18
35			Minority	0.22	0.11	0.35	0.15
36			All	0.26	0.15	0.31	0.16

*Notes.* This table displays summary statistics of the efficiency of individual voter decision-making for each of the four games. We report the mean of the session averages and the standard deviation of the mean for each voter type.

Table 21: Summary statistics: efficiency of aggregate voter decision-making, Part II

Game	Election outcome	Majority interest group		Minority interest group	
		Mean	St. Dev.	Mean	St. Dev.
MD	Efficient	0.90	0.08	0.18	0.15
	Inefficient	0.07	0.04	0.53	0.11
	Tie	0.03	0.05	0.29	0.09
MU	Efficient	0.64	0.12	0.33	0.13
	Inefficient	0.19	0.11	0.47	0.15
	Tie	0.18	0.06	0.20	0.11
PD	Efficient	0.78	0.12	0.33	0.16
	Inefficient	0.08	0.05	0.28	0.10
	Tie	0.14	0.09	0.40	0.12
PU	Efficient	0.57	0.14	0.38	0.14
	Inefficient	0.19	0.11	0.44	0.10
	Tie	0.25	0.08	0.18	0.10

*Notes.* This table displays summary statistics of the efficiency of aggregate voter decision-making for each of the four games. We report the mean of the session averages and the standard deviation of the mean for each interest group type.

## A.10 Lying task profiles

Here we provide an overview of the lying task profile distributions for each of the four games in Part I (Table 22).

Table 22: Distribution of lying task profiles

Game	Always	Mix	Never
MD	0.72	0.14	0.14
MU	0.70	0.15	0.15
PD	0.71	0.13	0.16
PU	0.61	0.18	0.21

*Notes.* We classify a subject's actions in the lying task as "Always" if the subject lies twice, "Mix" if the subject lies one, and "Never" if the subject does not lie.

In all games of the experiment, we observe more truth-telling than predicted. This deviation may be caused by lying aversion. The results of our lying aversion task suggest that some subjects were, indeed, unwilling to tell a lie that would have increased their payoffs. In Table 22, we observe that roughly 30% of all subjects told the truth at least once (cf. the *Mix* and *Never* columns), while telling a lie twice resulted in the highest payoff.

## A.11 Risk elicitation task

In Section A.13.4, we present our risk elicitation task based on Holt and Laury (2002). Subjects make ten consecutive choices between Option A and Option B. Option A guarantees a safe payoff which is equal to the intermediate voter payoff in our games. Option B is a gamble, with possible payoffs equal to the highest and lowest payoffs in our games. The payoffs are the same for all choices but the probability of winning the high payoff if option B is chosen increases with each choice (see the list in Section A.13.4). Hence, the expected payoff of Option B increases. We classify subjects into three categories based on the number of times that they have chosen the ‘safe’ Option A.<sup>51</sup> The first category of risk-seekers choose option A 1-3 times. The second category of risk-neutral subjects choose option A 4-7 times. The third category consists of risk-averse subjects who choose option A 8-10 times. An overview of the risk profiles for each of the four games in Part I of the experiment is presented in Table 23.

Table 23: Distribution of risk profiles

Game	Risk-seeking	Risk-neutral	Risk-averse
MD	0.21	0.60	0.29
MU	0.18	0.58	0.35
PD	0.18	0.60	0.31
PU	0.15	0.58	0.35

*Notes.* Each entry displays the frequency of a risk profile (column) in a Part-I game (row).

Propositions 1 to 4 assume risk-neutral subjects. In contrast to risk-neutral subjects, risk-seeking subjects would not abstain but vote for either Party X or Party Y if their beliefs are equal to 0.5. Risk-averse subjects are different from risk-neutral subjects because they choose abstaining instead of voting if their beliefs are equal to 1/3 or 2/3. Hypotheses 1 to 3 rely upon the assumption that there are at least some voters that are not risk-averse. This assumption is valid because our risk elicitation task reported on in Table 23 shows that

<sup>51</sup>We only allow subjects to switch from Option A to Option B once.

roughly 60% of the subjects have a risk-neutral profile.

## A.12 Regressions

Due to space constraints, we only report estimated marginal effects and omit the corresponding random effects panel regressions in the main text. By means of illustration, we present the random effects panel regressions of trust in the MD-game, the social media environment, and the traditional media environment in this section. Regressions for our other variables of interest are available upon request.

Table 24: Random effects panel regressions of trust in the MD-game

	Majority interest group	Minority interest group
Type	-0.525*** (0.056)	0.485*** (0.067)
Intercept	0.875*** (0.030)	0.304*** (0.044)
Variance of the random effect	0.171	0.189
Residual variance	0.343	0.403
Intraclass correlation	0.200	0.177
Observations	720	360
Subjects	72	72

*Notes.* The dependent variable in both models is trust. Type is a binary variable that takes on value 1 if the voter has type Minority and 0 otherwise. The model allows for subject-specific random effects. Robust standard errors, which are clustered at the subject level, are displayed in parentheses, \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table 25: Random effects panel regressions of trust in the social media environment

	Majority interest group		Minority interest group	
	Majority voter	Minority voter	Majority voter	Minority voter
MD	0.172*** (0.064)	-0.002 (0.101)	-0.350*** (0.088)	0.432*** (0.123)
PD	0.105* (0.061)	0.445*** (0.088)	-0.449*** (0.076)	0.160 (0.116)
PU	-0.089 (0.063)	0.141 (0.091)	-0.174** (0.076)	0.252** (0.108)
Part	-0.077 (0.045)	0.076 (0.069)	-0.070 (0.049)	0.121 (0.080)
Intercept	0.738*** (0.047)	0.411*** (0.067)	0.774*** (0.051)	0.385*** (0.080)
Variance of the random effect	0.248	0.248	0.244	0.287
Residual variance	0.324	0.262	0.267	0.341
Intraclass correlation	0.358	0.415	0.404	0.399
Observations	1,920	960	960	480
Subjects	144	144	144	144

*Notes.* The dependent variable in both models is trust. The regressors MD, PD and PU are equal to one if an observation comes from the respective game and zero otherwise. Part is a binary variable that takes on value 1 if an observation comes from Part I of the experiment and zero otherwise. The model allows for subject-specific random effects. Robust standard errors, which are clustered at the subject level, are displayed in parentheses, \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table 26: Random effects panel regressions of trust in the traditional media environment

	Majority interest group		Minority interest group	
	Majority voter	Minority voter	Majority voter	Minority voter
MD	0.089** (0.039)	-0.509*** (0.084)	-0.186** (0.086)	0.361*** (0.108)
MU	-0.281*** (0.060)	-0.444*** (0.074)	0.260*** (0.097)	-0.105 (0.122)
PU	-0.149** (0.059)	-0.092 (0.080)	0.334*** (0.086)	0.092 (0.104)
Part I	-0.019 (0.023)	-0.055 (0.036)	-0.003 (0.065)	0.056 (0.071)
Intercept	0.858*** (0.031)	0.801*** (0.044)	0.415*** (0.064)	0.462*** (0.071)
Variance of the random effect	0.224	0.257	0.286	0.311
Residual variance	0.337	0.323	0.352	0.391
Intraclass correlation	0.314	0.372	0.389	0.388
Observations	1,920	960	960	480
Subjects	144	144	144	144

*Notes.* The dependent variable in both models is trust. The regressors MD, MU and PU are equal to one if an observation comes from the respective game and zero otherwise. Part is a binary variable that takes on value 1 if an observation comes from Part I of the experiment and zero otherwise. The model allows for subject-specific random effects. Robust standard errors, which are clustered at the subject level, are displayed in parentheses, \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

## A.13 Instructions

In this section, we provide the complete set of instructions for subjects participating in the MU-PD treatment, which consists of instructions for Part I (Section A.13.1), instructions for Part II (Section A.13.2), instructions for the lying aversion task (Section A.13.3), instructions for the risk elicitation task (Section A.13.4), tutorial questions for Part I (Section A.13.4), tutorial questions for Part II (Section A.13.4) and the questionnaire (Section A.13.4).<sup>52</sup>

The description of the belief elicitation task is borrowed from Nyarko and Schotter (2002). The description of the risk elicitation task is based on Holt and Laury (2002).

### A.13.1 Part I

#### INSTRUCTIONS

Welcome to this experiment! Please remain silent during the experiment and do not speak to other participants. If you have a question or need assistance of any kind, please raise your hand and a member of our staff will come to you.

In the experiment, you will be asked to make a series of decisions. If you follow the instructions carefully, you can earn a considerable amount of money, depending on your decisions and those made by other participants. All payoffs in the experiment are denominated in “Points.” At the end of the experiment, your earnings in Points will be converted into Euros at the exchange rate indicated below and will be paid to you in cash.

The experiment will consist of two parts, called Part I and Part II, and some additional tasks. We will first explain the rules of Part I. The rules of Part II will be explained after the end of Part I.

Part I will consist of 40 rounds. In each round there will be a Decision Task and in some rounds there will also be a Prediction Task. We will first explain the Decision Task.

---

<sup>52</sup>The set of instructions for the other treatments are available upon request.

## Decision Task

The events in each round are as follows:

1. At the beginning of each round, you will be randomly assigned to a group of four participants. In each group, one participant will act in role A and three participants will act in role B.
  - (a) The A-participant can be of two types: With a 2 out of 3 chance the type of the A-participant in your group will be “**Type 1**” and with a 1 out of 3 chance the type of the A-participant in your group will be “**Type 2**.” An A-participant with “**Type 1**” will be called A1 and an A-participant with “**Type 2**” will be called A2.
  - (b) Of the three B-participants, two participants will act in role B1 and one participant will act in role B2.
2. After groups were formed, a chance move determines the “state” for your group: With a 1 out of 2 chance the state will be “**Heads**” and with a 1 out of 2 chance the state will be “**Tails**.” This chance move is independent of the type of the A-participant in your group, meaning that there is no relation between the draw of the state and the type of the A-participant in your group.
3. The A-participant will then be informed about the state (either “**Heads**” or “**Tails**”). Then the A-participant will choose one message (either “**Heads**” or “**Tails**”) that is sent to the two B1-participants and one message (either “**Heads**” or “**Tails**”) that is sent to the B2-participant. The two messages can be the same or they can be different and may or may not be the same as the state.
4. The three B-participants will then be informed about their own message sent by the A-participant (that is, B1-participants will only learn the message sent to them and the B2-participant will only learn about the message sent to him/her). B-participants

will **not** be informed about the state and **not** about the type of the A-participant. Then all three B-participants simultaneously and independently will choose between action “**X**”, action “**Y**” or action “**Z**.”

5. The payoffs in each round are described in the tables on the next page.

### Payoffs

The A-participant will receive a payoff from the interaction with **each** of the three B-participants. That is, the payoff of the A-participant will be the sum of the payoffs he/she receives from the interaction with each of the two B1-participants and the B2-participant.

The payoff an A-participant receives from the interaction with a B-participant depends on the state, the type of the A-participant and the action chosen by a B-participant. The payoffs of the A-participant from the interaction with a B1- or B2-participant are shown below in the Tables labelled “The payoffs of an A1-participant” and “The payoffs of an A2-participant.”

The payoff a B-participant receives depends on the state and the own action by a B-participant. The payoffs of a B1-participant and a B2-participant from the interaction with the A-participant are shown below in the Tables labelled “The payoffs of a B1-participant” and “The payoffs of a B2-participant.” Note that the payoffs of any of the B-participants do not depend on the type of the A-participant or the message chosen by the A-participant.

---

<b>The payoffs of an A1-participant</b>				<b>The payoffs of an A2-participant</b>			
	Decision of a B-participant (B1 or B2)				Decision of a B-participant (B1 or B2)		
State	X	Y	Z	State	X	Y	Z
Heads	30	10	20	Heads	10	30	20
Tails	10	30	20	Tails	30	10	20

---

---

The payoffs of a B1-participant				The payoffs of a B2-participant			
Decision of a B1-participant				Decision of the B2-participant			
State	X	Y	Z	State	X	Y	Z
Heads	90	27	60	Heads	27	90	60
Tails	27	90	60	Tails	90	27	60

---

## **Number of rounds, role assignment and matching**

Number of rounds: Part I of the experiment consists of 40 rounds.

Role assignment: You will have to make decisions both as an A- and as a B-participant, alternating in the following way: The roles of all participants are randomly determined for blocks of five consecutive rounds. After five rounds, new roles are assigned to all participants that remain fixed for another block of five rounds. For example, a participant who had the role of an A-participant (either A1 or A2) for a block of the past five rounds, will have the role of a B-participant (either B1 or B2) for the next block of five rounds (if the experiment is not over before this). Since there are more B1-participants than either A- or B2-participants, it can happen that you will act in the role of a B1-participant in ten consecutive rounds. In any case, each participant will act ten rounds in role A (A1 or A2), 20 rounds in role B1 and ten rounds in role B2. Your computer screen shows you in every round which role you have in that round (A1, A2, B1 or B2).

Matching: All participants in the room are assigned to a set of twelve participants. This set composition will remain the same throughout the entire experiment. At the beginning of each round and within a set of twelve, the computer will randomly match participants into groups of four (one A-participant (A1 or A2), two B1-participants and one B2-participant). The matching is random, meaning that there is no relation between the participants you have been matched with last round (or any other previous round) and the participants whom you will be matched with in the current round.

### **Prediction Task**

In the first and the last round of a block of five rounds, you will not only make a decision as described above, but you will also be asked to predict the choices that will be made by the B-participants (if you act in role A) or the state and the type of the A-participant that were drawn at the beginning of the round (if you act in role B). For this prediction task, you can earn additional Points depending upon how good your prediction was. Since A- and

B-participants have to make predictions about different things, we will explain the prediction tasks for A- and B-participants separately.

### **Prediction Task for A-participants**

After an A-participant made a decision in the first and the last round of a block of five rounds about which message to send to the B-participants and before the B-participants make their choices, an A-Participant will be asked to indicate what actions he/she thinks the two B1-participants and the B2-participant will choose in the current round.

For this purpose, an A-participant will see two tables of the following form on the screen. A table on the top for the prediction of the actions that will be chosen by the two B1-participants and a table on the bottom for the prediction of the action that will be chosen by the B2-participant.

#### **The Tables for an A-participant's Prediction Task**

**Which actions do you think the two B1-participants will choose?**

- Both B1 will choose X
- Both B1 will choose Y
- Both B1 will choose Z
- One B1 will choose X, One B1 will choose Y
- One B1 will choose X, One B1 will choose Z
- One B1 will choose Y, One B1 will choose Z

**Which action do you think the B2-participant will choose?**

- B2 will choose X
- B2 will choose Y
- B2 will choose Z

Regarding the choices of the two B1-participants, there are six different possibilities that the two B1-participants may decide after receiving the message from the A participant. They can either both choose the same action (X, Y or Z) or they can both choose a different action (1X and 1Y, or 1X and 1Z, or 1Y and 1Z).

- An A-participant has to indicate which of the six cases will apply by clicking the button in the corresponding line in the table on top of this screen.
- If an A participant's prediction was correct, the A-participant will earn 30 Points for this prediction task. If the prediction was not correct, the A-participant will earn 0 Points from the prediction task.

Regarding the choice of the B2-participant, there are three different possibilities that the B2-participant may decide after receiving the message from the A participant (X, Y or Z).

- An A-participant has to indicate which of the three cases will apply by clicking the button in the corresponding line in the table on the bottom of this screen.
- If an A participant's prediction was correct, the A-participant will earn 15 Points for this prediction task. If the prediction was not correct, the A-participant will earn 0 Points from the prediction task.

### **Prediction Task for B-participants**

After a B-participant made a decision in the first and the last round of a block of five rounds about which action to choose after receiving a message from the A-participant, a B-participant will be asked to predict the state and the type of the A-participant.

More precisely, a B-participant will be asked to indicate

- what he/she thinks is the chance that the state is Heads and what is the chance that the state is Tails, and

- what he/she thinks is the chance that the A-participant is of Type 1 and what is the chance that the A-participant is of Type 2.

To make a prediction, a B-participant will see two tables of the following form on the screen. A table on the top for the prediction of the state and a table on the bottom for the prediction of the type of the A-participant.

### The Tables for a B-participant's Prediction Task

<p><b>Which state do you think was selected?</b></p> <p>(The two numbers need to sum up to 100.)</p> <table style="width: 100%; border: none;"> <tbody> <tr> <td style="width: 50%; text-align: center; padding: 5px;">           Chance (in %) that the state is <b>Heads</b> <input style="width: 50px; height: 15px;" type="text"/> </td> <td style="width: 50%; text-align: center; padding: 5px;">           Chance (in %) that the state is <b>Tails</b> <input style="width: 50px; height: 15px;" type="text"/> </td> </tr> </tbody> </table>		Chance (in %) that the state is <b>Heads</b> <input style="width: 50px; height: 15px;" type="text"/>	Chance (in %) that the state is <b>Tails</b> <input style="width: 50px; height: 15px;" type="text"/>
Chance (in %) that the state is <b>Heads</b> <input style="width: 50px; height: 15px;" type="text"/>	Chance (in %) that the state is <b>Tails</b> <input style="width: 50px; height: 15px;" type="text"/>		
<p><b>Which type do you think the A-participant has?</b></p> <p>(The two numbers need to sum up to 100.)</p> <table style="width: 100%; border: none;"> <tbody> <tr> <td style="width: 50%; text-align: center; padding: 5px;">           Chance (in %) that the A-participant is of <b>Type 1</b> <input style="width: 50px; height: 15px;" type="text"/> </td> <td style="width: 50%; text-align: center; padding: 5px;">           Chance (in %) that the A-participant is of <b>Type 2</b> <input style="width: 50px; height: 15px;" type="text"/> </td> </tr> </tbody> </table>		Chance (in %) that the A-participant is of <b>Type 1</b> <input style="width: 50px; height: 15px;" type="text"/>	Chance (in %) that the A-participant is of <b>Type 2</b> <input style="width: 50px; height: 15px;" type="text"/>
Chance (in %) that the A-participant is of <b>Type 1</b> <input style="width: 50px; height: 15px;" type="text"/>	Chance (in %) that the A-participant is of <b>Type 2</b> <input style="width: 50px; height: 15px;" type="text"/>		

For example, say you think there is an 80% chance that the state is Heads, and hence a 20% chance that the state is Tails. This indicates that you believe that Heads is more likely than Tails by a considerable margin. If this is your belief about the likely state, then write 80 into the box labelled “Chance (in %) that the state is Heads” and write 20 into the box

labelled “Chance (in %) that the state is Tails.” Note that the two numbers must add up to 100.

Similarly, say you think there is a 67% chance that the A-participant is of Type 1 and a 33% chance that the A-participant is of Type 2, then write 67 into the box “Chance (in %) that the A-participant is of Type 1” and 33 into the box “Chance (in %) that the A-participant is of Type 2.” Note that also these two numbers must add up to 100.

At the end of the round, the computer will compare your predictions regarding the state and the type of the A-participant with the actual state and the type of the A-participant. You can earn up to 20 Points for each of your predictions. While the precise explanation we provide below of how you will be paid for the two prediction tasks might look difficult, it can be summarized by two basic principles:

- **The better your prediction, that is, the more percentage points your prediction assigns to the actual state / the actual type of the A-participant and the fewer percentage points your prediction assigns to the state that was not selected / the type not selected for the A-participant, the higher your payoff from each of the two prediction tasks.**
- **Since your prediction is made before you learn what the actual state / the actual type of the A-participant is, the best thing you can do to maximize your expected payoffs from the predictions is to simply state your true beliefs about the chances that the state is either Heads or Tails / the type of the A-participant is either Type 1 or Type 2. Any other predictions will decrease the payoffs you can expect to earn from the predictions.**

Please see the last page of these instructions if you are interested in the precise explanation of how a B-participant will be paid for the Prediction Task.

**Feedback after each round**

At the end of each round, the feedback screen will show the following information about what happened in your own group during the round:

- The feedback screen of an A-participant will show the state, the own type, the message sent to the two B1-participants, the message sent to the B2-participant, the actions taken by all three B-participants and the own payoff.
- The feedback screen of a B-participant will show the state, the type of the A-participant, the own message sent by the A-participant, the own action and the own payoff.

### **Monetary Earnings**

Your payoff for Part I of the experiment is equal to the sum of your own payoffs in all rounds. For every 175 Points earned in the Decision and the Prediction Task you will be paid 1 EUR.

### **Summary**

To summarize, the events in each round are as follows:

1. The computer randomly matches participants into groups of four. Each group contains one A-participant. With a 2 out of 3 chance the type of the A-participant in your group is “Type 1” (A1) and with a 1 out of 3 chance the type of the A-participant in your group is “Type 2” (A2). Each group also contains three B-participants (two B1-participants and one B2-participant). All participants are informed about their own role (A1, A2, B1 or B2).
2. The computer randomly determines the state. With a 1 out of 2 chance the state in your group will be “Heads” and with a 1 out of 2 chance the state in your group will be “Tails.” The chance move that determines the state is independent of the A-participant’s type in your group.

3. The A-participant is informed about the state. Then the A-participant sends one message to the two B1-participants (either “Heads” or “Tails”), and one message to the B2-participant (either “Heads” or “Tails”). The two messages sent can be the same or they can be different.
4. Each B-participant is informed about his/her own message sent by the A-participant. B-participants are not informed about the state or the type of the A-participant. Then each B participant chooses between action “X”, action “Y” and action “Z”.
5. In some rounds you will be asked to either predict the choices that will be made by the B-participants (if you act in role A) or the drawn state and the A-participants’s type (if you act in role B).
6. Payoffs result as described above.

## Precise explanation of how the B-participants will be paid for the Prediction Task

Let us here give you the precise explanation of how you will be paid for the B-participant's Prediction Task. We do so with regard to the prediction of the state. An analogous explanation applies to the prediction of the A-participant's type.

Suppose you predict that the state is Heads with an 80% chance and Tails with a 20% chance. Two cases can apply now. Case 1 is that the actual state is Heads and Case 2 is that the actual state is Tails. For each of these two cases we will explain now how your payoff from the state Prediction Task will be computed.

- **Case 1:** Suppose that the actual state is Heads. In this case, you will get

$$\text{Payoff} = 20 - \frac{(1 - 0.80)^2}{0.1} - \frac{(0.20)^2}{0.1}.$$

This formula means that you will be paid a fixed amount of 20 Points from which we will subtract an amount which depends on how inaccurate your prediction was. To do this, we will take the number you assigned to Heads (in this case 80% on Heads), divide it by 100 (to get 0.80), subtract it from 1, square it and divide the result by 0.1. We will then take the number you assigned to the state that is not the actual state (in this case the 20% you assigned to Tails), divide it by 100 (to get 0.20), square it, and divide the result by 0.1. These two numbers will then be subtracted from the 20 Points we initially gave you to determine your prediction payoff.

- **Case 2:** Suppose that the actual state is Tails. In this case, you will get

$$\text{Payoff} = 20 - \frac{(1 - 0.20)^2}{0.1} - \frac{(0.80)^2}{0.1}.$$

This formula means that you will again be paid a fixed amount of 20 Points from which we will subtract an amount which depends on how inaccurate your prediction was. To

do this, we will take the number you assigned to Tails (in this case 20% on Tails), divide it by 100 (to get 0.20), subtract it from 1, square it, and divide the result by 0.1. We will then take the number you assigned to state that is not the actual state (in this case the 80% you assigned to Heads), divide it by 100 (to get 0.80), square it, and divide the result by 0.1. These two numbers will then be subtracted from the 20 Points we initially gave you to determine your prediction payoff.

Note that the worst you can do under this payoff scheme is to declare that you predict that there is a 100% chance that a certain state has been selected and assign 100% to that state when in fact the other state was selected. Here your payoff from prediction would be 0 Points. On the other hand, the best you can do is to predict correctly and assign 100% to the actual state that had been selected at the beginning of the round. Here your payoff will be 20 Points.

An analogous explanation applies to your payment for the prediction of the A-participant's type.

### **A.13.2 Part II**

#### **INSTRUCTIONS for Part II**

Part II of the experiment will also consist of 40 rounds. Again, in each round there will be a Decision Task and in some rounds there will also be a Prediction Task. The experimental situation in Part II is related to the one in Part I. However, there are some changes. The changes in comparison to Part I of the experiment are highlighted in bold font.

#### **Decision Task**

The events in each round are as follows:

1. At the beginning of each round, you will be randomly assigned to a group of four participants. In each group, one participant will act in role A and three participants will act in role B.

- (a) The A-participant can be of two types: With a 2 out of 3 chance the type of the A-participant in your group will be “**Type 1**” and with a 1 out of 3 chance the type of the A-participant in your group will be “**Type 2**.” An A-participant with “**Type 1**” will be called A1 and an A-participant with “**Type 2**” will be called A2.
- (b) Of the three B-participants, two participants will act in role B1 and one participant will act in role B2.
2. After groups were formed, a chance move determines the “state” for your group: With a 1 out of 2 chance the state will be “**Heads**” and with a 1 out of 2 chance the state will be “**Tails**.” This chance move is independent of the type of the A-participant in your group, meaning that there is no relation between the draw of the state and the type of the A-participant in your group.
  3. The A-participant will then be informed about the state (either “**Heads**” or “**Tails**”). **Then the A-participant will choose one message (either “Heads” or “Tails”) that is sent to all three B-participants.** The message may or may not be the same as the state.
  4. **The three B-participants will then be informed about the message sent by the A-participant and about the type of the A-participant.** B-participants will not be informed about the state. Then all three B-participants simultaneously and independently will choose between action “**X**”, action “**Y**” or action “**Z**.”
  5. The payoffs are exactly as in Part I of the experiment, and are described in the tables on the next page.

### **Payoffs**

The A-participant will receive a payoff from the interaction with **each** of the three B-participants. That is, the payoff of the A-participant will be the sum of the payoffs he/she

receives from the interaction with each of the two B1-participants and the B2-participant.

The payoff an A-participant receives from the interaction with a B-participant depends on the state, the type of the A-participant and the action chosen by a B-participant. The payoffs of the A-participant from the interaction with a B1- or B2-participant are shown below in the Tables labelled “The payoffs of an A1-participant” and “The payoffs of an A2-participant.”

The payoff a B-participant receives depends on the state and the own action by a B-participant. The payoffs of a B1-participant and a B2-participant from the interaction with the A-participant are shown below in the Tables labelled “The payoffs of a B1-participant” and “The payoffs of a B2-participant.” Note that the payoffs of any of the B-participants do not depend on the type of the A-participant or the message chosen by the A-participant.

---

<b>The payoffs of an A1-participant</b>				<b>The payoffs of an A2-participant</b>			
	Decision of a				Decision of a		
	B-participant (B1 or B2)				B-participant (B1 or B2)		
State	X	Y	Z	State	X	Y	Z
Heads	30	10	20	Heads	10	30	20
Tails	10	30	20	Tails	30	10	20

---

---

---

**The payoffs of a B1-participant**

	Decision of a <b>B1-participant</b>		
State	X	Y	Z
Heads	90	27	60
Tails	27	90	60

---

---

---

---

**The payoffs of a B2-participant**

	Decision of the <b>B2-participant</b>		
State	X	Y	Z
Heads	27	90	60
Tails	90	27	60

---

---

## **Number of rounds, role assignment and matching**

Number of rounds: Part II of the experiment consists of 40 rounds.

Role assignment: As in Part I, you will have to make decisions both as an A- and as a B-participant, alternating in blocks of five rounds as described in the instructions for Part I. Again, your computer screen shows you in every round which role you have in that round (A1, A2, B1 or B2).

Matching: Also as in Part I, **at the beginning of each round**, the computer will randomly match participants into groups of four (one A-participant (A1 or A2), two B1-participants and one B2-participant).

### **Prediction Task**

In the first and the last round of a block of five rounds, you will not only make a decision as described above, but you will also be asked to predict the choices that will be made by the B-participants (if you act in role A) or **only** the state that was drawn at the beginning of the round (if you act in role B). For this prediction task, you can earn additional Points depending upon how good your prediction was.

#### **Prediction Task for A-participants**

The instructions and payoff rules for the Prediction Task for A-participants are exactly the same as in Part I of the experiment.

#### **Prediction Task for B-participants**

The instructions and payoff rules for the Prediction Task for B-participants regarding the state are exactly the same as in Part I of the experiment.

### **Feedback after each round**

At the end of each round, the feedback screen will show the following information about what happened in your own group during the round:

- The feedback screen of an A-participant will show the state, the own type, the message sent to the B-participants, the actions taken by all three B-participants and the own payoff.
- The feedback screen of a B-participant will show the state, the type of the A-participant, the message sent by the A-participant, the own action and the own payoff.

### Monetary Earnings

Your payoff for Part II of the experiment is equal to the sum of your own payoffs in all rounds. For every 175 Points earned in the Decision and the Prediction Task you will be paid 1 EUR.

### Summary

To summarize, the events in each round of Part II are as follows:

1. The computer randomly matches participants into groups of four. Each group contains one A-participant. With a 2 out of 3 chance the type of the A-participant in your group is “Type 1” (A1) and with a 1 out of 3 chance the type of the A-participant in your group is “Type 2” (A2). Each group also contains three B-participants (two B1-participants and one B2-participant). All participants are informed about their own role (A1, A2, B1 or B2).
2. The computer randomly determines the state. With a 1 out of 2 chance the state in your group will be “Heads” and with a 1 out of 2 chance the state in your group will be “Tails.” The chance move that determines the state is independent of the A-participant’s type in your group.
3. The A-participant is informed about the state. **Then the A-participant sends one message to the three B-participants (either “Heads” or “Tails”).**

4. **Each B-participant (B1 or B2) is informed about the message chosen by the A-participant and about the type of the A-participant.** B-participants will not be informed about the state. Then each B-participant chooses between action “X”, action “Y” and action “Z”.
5. In some rounds you will be asked to either predict the choices made by the B-participants (if you act in role A) or the drawn state (if you act in role B).
6. Payoffs result as described above.

### A.13.3 Lying Aversion Task

We ask you to also make decisions in the following situation, which is related to the Decision Tasks in Part I and II of the experiment. However, there are various changes. For every 20 Points earned you will be paid 1 EUR.

The events in each of two rounds will be as follows:

1. You will act in the role of an A2-participant. In contrast to the situations in Part I and Part II of the experiment, you will only be matched with **one computerized B1-participant**.
2. A state will be selected. The state can be **Heads** or **Tails**.
3. You will then be informed about the state (either Heads or Tails) and you will then choose a message that is sent to the computerized B1-participant. The message can either be “**Heads**” or “**Tails**” and may or may not be the same as the state.
4. The computerized B1-participant will then be informed about your message and will choose between action “X” and action “Y”. (Action “Z” is not available to B1.) The

payoff table of the computerized B1-participant looks as follows.

		Decision of the computerized B1-participant	
		X	Y
Your	Heads	30	10
Message	Tails	10	30

The computerized B1 participant will assume that **the state is the same as the message you sent** and will automatically choose the action that maximizes its own payoff. That is:

- (a) If the message you sent is Heads, B1 will choose action X.
- (b) If the message you sent is Tails, B1 will choose action Y.

5. Your payoffs in each round are described in the following table:

		State	
		Heads	Tails
Your	Heads	10	30
Choice	Tails	30	10

That is:

- If the state is Heads and you choose Heads, you earn 10 Points.
- If the state is Heads and you choose Tails, you earn 30 Points.
- If the state is Tails and you choose Heads, you earn 30 Points.
- If the state is Tails and you choose Tails, you earn 10 Points.

You will make a decision as an A2-participant in two rounds, once for each of the two possible states (Heads or Tails). The computer will randomly determine the state for the first round and will select the other state in the second round.

If you have questions regarding these instructions, please raise your hand. Otherwise click the button below:

I have read and understood the instructions.

#### **A.13.4 Risk Elicitation Task**

We ask you to make some more decisions as follows.

The decision table on the right shows ten decisions tasks numbered from 1 to 10. Each decision task is a paired choice between “Option A” and “Option B.” You will have to make a choice between Option A and Option B in all ten decision tasks. However, only one of them will be used in the end to determine your earnings. For every 30 Points earned in this part of the experiment you will be paid 1 EUR. Before you start making your ten choices, let us explain how your choices will affect your earnings for this part of the experiment.

After you have made your ten decisions, the computer will first randomly draw a number between 1 and 10, where each of the ten numbers is equally likely to be drawn. The randomly drawn number determines which of your ten decisions will be relevant for payment.

Your payment will then be determined as follows. The computer will randomly draw a second number, this time between 1 to 100, where each of the 100 numbers is equally likely to be drawn.

Please look at decision 1 at the top of the table.

- Option A pays 60 Points with certainty.
- Option B pays 90 Points if the drawn number is 35 or lower, and it pays 27 Points if the drawn number is 36 or higher.

The other decisions are similar, except that as you move down the table the chances of getting 90 Points in Option B increases.

For each of the ten decisions, you will be asked to choose Option A or Option B by clicking on the appropriate box.

Are there any questions? If yes, please raise your hand. If not, please click the “I have read and understood the instructions” button.

The list:

	Option A	Option B
1	60 Points with certainty	90 Points with a chance of 35% 27 Points with a chance of 65%
2	60 Points with certainty	90 Points with a chance of 40% 27 Points with a chance of 60%
3	60 Points with certainty	90 Points with a chance of 45% 27 Points with a chance of 55%
4	60 Points with certainty	90 Points with a chance of 50% 27 Points with a chance of 50%
5	60 Points with certainty	90 Points with a chance of 55% 27 Points with a chance of 45%
6	60 Points with certainty	90 Points with a chance of 60% 27 Points with a chance of 40%
7	60 Points with certainty	90 Points with a chance of 65% 27 Points with a chance of 35%
8	60 Points with certainty	90 Points with a chance of 70% 27 Points with a chance of 30%
9	60 Points with certainty	90 Points with a chance of 75% 27 Points with a chance of 25%
10	60 Points with certainty	90 Points with a chance of 80% 27 Points with a chance of 20%

### A.13.5 Tutorial Questions Part I

Before we start with the experiment, we would like you to answer a few questions that are meant to review the rules of today's experiment. You will not be able to continue until you have answered all questions correctly, but may make as many attempts as you need. If you don't know the correct answer to a question, please first read the relevant part of the instructions again. If you then still have difficulties answering a question, please raise your hand.

1. Suppose you are an A1-participant and the state is "Tails". What is your payoff from the interaction with a B-participant who chooses Y?

2. Suppose you are an A2-participant and the state is "Heads". What is your total payoff if two B-participants choose Y and one B-participant chooses Z?

3. Suppose you are a B1-participant and you receive the message "Tails". Do you know with certainty what payoff you get if you choose action X?

- Yes  
 No

4. Suppose you are a B2-participant and the state is "Heads". What is your payoff if you choose action X?

5. Suppose you are a B-participant. Does your payoff depend on the actions of the other B-participants?

- Yes  
 No

6. How many A-participants will be in your group?

7. How many B1-participants will be in your group?

8. How many B2-participants will be in your group?

9. Of how many rounds will Part I of the experiment consist?

10. What is the chance at the beginning of a round that the state in your group will be "Heads"?

 out of 

OK

11. What is the chance at the beginning of a round that the A-participant in your group will have "Type 1"?

out of

12. Will the state that will be determined at the beginning of a round depend on the A-participant's type?

- Yes  
 No

13. Suppose you are assigned to be an A2-participant in the beginning of Part I of the experiment. How many consecutive rounds will you act in this role in Part I of the experiment?

14. For how many rounds will you act as an A-participant in Part I of the experiment?

15. For how many rounds will you act as a B1-participant in Part I of the experiment?

16. For how many rounds will you act as a B2-participant in Part I of the experiment?

17. Is the following statement true or false? At the beginning of each round, the computer will randomly match participants into groups of four (one A-participant (A1 or A2), two B1-participants and one B2-participant).

- True  
 False

18. Does a message sent by an A-participant need to be the same as the state?

- Yes  
 No

19. Will the A-participant be informed about the state before he/she sends a message?

- Yes  
 No

20. Will the B-participants be informed about the state before they choose an action?

- Yes  
 No

21. Will the B-participants be informed about the type of the A-participant before they choose an action?

- Yes  
 No

22. Suppose you are a B-participant and you receive a message (either "Heads" or "Tails"). Do you know with certainty now what message the other B-participants have received?

- Yes  
 No

OK

## A.13.6 Tutorial Questions Part II

Before we continue with the experiment, we would like you to answer a few questions that are meant to review the rules of today's experiment. You will not be able to continue until you have answered all questions correctly, but may make as many attempts as you need. If you don't know the correct answer to a question, please first read the relevant part of the instructions again. If you then still have difficulties answering a question, please raise your hand.

1. Will the B-participants be informed about the type of the A-participant before they choose an action?

- Yes  
 No

2. Suppose you are a B-participant and you receive a message (either "Heads" or "Tails"). Do you know with certainty now what message the other B-participants have received?

- Yes  
 No

OK

## A.13.7 Questionnaire

Questionnaire

**Background questions**

1. What is your age?

2. What is your gender?

male  
 female  
 other

3. What is your field of study?

4. Which year did you enroll in University for the first time?

5. What is your first (native) language?

Questionnaire

**Questions about Part I of the experiment**

If you have been an A1-participant during Part I of the experiment, please answer the following question.

1. As an A1-participant in Part I, was there a rule according to which you made your choices in the Decision task? If yes, could you please describe this rule?

If you have been an A2-participant during Part I of the experiment, please answer the following question.

2. As an A2-participant in Part I, was there a rule according to which you made your choices in the Decision task? If yes, could you please describe this rule?

3. As a B1-participant in Part I, was there a rule according to which you made your choices in the Decision task? If yes, could you please describe this rule?

4. As a B2-participant in Part I, was there a rule according to which you made your choices in the Decision task? If yes, could you please describe this rule?

Continue

Questionnaire

**Questions about Part II of the experiment**

If you have been an A1-participant during Part II of the experiment, please answer the following question.

1. As an A1-participant in Part II, was there a rule according to which you made your choices in the Decision task? If yes, could you please describe this rule?

If you have been an A2-participant during Part II of the experiment, please answer the following question.

2. As an A2-participant in Part II, was there a rule according to which you made your choices in the Decision task? If yes, could you please describe this rule?

3. As a B1-participant in Part II, was there a rule according to which you made your choices in the Decision task? If yes, could you please describe this rule?

4. As a B2-participant in Part II, was there a rule according to which you made your choices in the Decision task? If yes, could you please describe this rule?

Continue