Social Media and Democracy: Experimental Results*

Freek van Gils[†] Wieland Müller[‡] Jens Prüfer[†]

January 25, 2022

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. In a first set of treatments of our laboratory experiment, we study the effects of the implementation of a microtargeting ban and disclosure of interests in a social media environment on voting behavior. We find that disclosure of interests is essential to reduce the frequency of inefficient election outcomes. We also show that a microtargeting ban is redundant for 'mainstream' voters but is needed to improve voter decision-making of a smaller, 'niche' voter group. These findings are consistent with the result of our second set of treatments, in which we show that the transition from a traditional media environment to a social media environment makes all voters worse off. Lastly, we show that 'niche' voters are equally well off as 'mainstream' voters in games with public communication (one message for all voters) but stay behind in games with a microtargeted communication technology (one message for each voter group).

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

JEL Codes: C92, D72, D82, D83.

*We are grateful to Boris van Leeuwen, Jan Potters, Martin Salm, Sigrid Suetens and seminar participants in Tilburg for their comments. We thank Anna Osipenko for research assistance. Jens Prüfer gratefully acknowledges financial support from a research grant awarded by Data Science Center Tilburg.

[†]Department of Economics, CentER, TILEC, Tilburg University, PO Box 90153, 5000 LE Tilburg, the Netherlands; f.w.vangils@uvt.nl, j.prufer@uvt.nl.

[‡]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.

1 Introduction

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.[1] In most other countries, news consumption through social media is also substantial and on the rise.[2]

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 microtargeting, the use of highly granular data on voter preferences and characteristics to match relevant political messages to individual voters (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 voters, aiming to manipulate the outcome of the 2016 U.S. presidential election.[3] Increasingly, social media were blamed for creating an accommodating environment for antidemocratic forces (Tucker et al., 2017).

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.[4] A solid empirical foundation for these interventions to avoid the manipulation of voter beliefs and election outcomes by disinformation is, however, missing.[5] 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. To address this issue, we develop, theoretically analyze

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

[2] See https://www.digitalnewsreport.org/survey/2020/overview-key-findings-2020/.

[3] See for instance https://www.washingtonpost.com/technology/2018/12/16/new-report-russian-disinformation-prepared-senate-shows-operations-scale-sweep/.

[4] 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).

[5] 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."

and experimentally test a series of games to study the effects of the implementation of a microtargeting ban and mandatory disclosure of interests within a social media environment on individual voting behavior and election outcomes.

Our games contain a politically motivated actor ("interest group") which is informed about a binary state of the world and sends a cheap talk message concerning the state (the truth or disinformation) to uninformed voters who subsequently make a voting decision. 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 that may differ from the message directed to the larger Majority voter group. This element of the game mimics social media platforms' microtargeted advertising services, which allow interest groups (and advertisers) to differentiate their message on the basis of individual-level data about (political) preferences and characteristics.[6] Second, the type of the interest group is undisclosed 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.[7] After subjects have repeatedly interacted (and have become experienced) within the social media environment, we implement one of three interventions or maintain the same (social media) environment in the second half of the experiment. The first intervention, which resembles the political microtargeting ban, restricts the interest group to send the same (public) message to both voter groups. Our second intervention, which is akin to mandatory disclosure of interests, reveals the interest group type to voters. The third intervention combines the microtargeting ban and mandatory disclosure of interests. 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 us 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 by comparing the outcomes in each of the three interventions with the outcomes in the social media environment in the second half of the experiment.

We evaluate the effectiveness of the interventions within our framework for individual

[6]See for instance Facebook's Custom Audience (https://www.facebook.com/business/help/744354708981227).

[7]Kalogeropoulos and Newman (2017) find that 47% of digital news consumers recall the source of a news item that is accessed through social media. 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.

(Majority and Minority) voters as well as the electorate as a whole. For the individual level, we examine whether voters vote for the 'efficient' (own payoff-maximizing) party and avoid voting for the opposing (inefficient) party more frequently after the implementation of the interventions. For the aggregate level, we consider whether the interventions promote efficient election winners, which are favored by a majority of the electorate, and prevent inefficient election outcomes. Our experimental results show that mandatory disclosure of interests, with or without a microtargeting ban, significantly and sizeably reduces the rate of inefficient election outcomes. We find that voters are more likely to trust truthful messages and less likely to give weight to false messages due to disclosure, which ultimately leads to more efficient voting actions. Second, we find that the joint implementation of a microtargeting ban and mandatory disclosure of interests improves the efficiency of voter decision-making for both Majority and Minority voters, while disclosure of interests without a microtargeting ban is mostly beneficial to Majority voters. The rationale behind this finding is that a microtargeting ban 'disciplines' an interest group to reveal the same information to the smaller Minority voter group as to the larger Majority group, which is more likely to receive truthful information due to the higher prevalence of the Majority interest group type.[8] The third and final result of this set of treatments is that a microtargeting ban has adverse aggregate consequences if implemented in isolation. The ban 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 the case in our framework), this intervention is expected to reduce the efficiency of aggregate voter decision-making. Our experimental data confirm this prediction.

Our results suggest that a social media policy aimed to ensure that election outcomes reflect voters' true preferences should include mandatory disclosure of interests for the senders of political information. In addition, the implementation of disclosure rules without a microtargeting ban might have distributional consequences: smaller voter groups may not benefit as much from the intervention as the larger electoral mainstream.

For many people, social media have replaced traditional news outlets such as TV channels and newspapers as a main source of (political) news (Newman et al., 2021). Whereas traditional media report the same content to all their consumers and generally have recognizable identities, social media can differentiate its supply of information for individual users and may obfuscate the original source of the information provided (Kalogeropoulos and Newman, 2017). To analyze the effects of the transition from a traditional media environment to a social media environment on voter decision-making, we implement a second set of treatments in the laboratory. The set-up of this experiment is similar to our first set

[8]The discipline effect of public communication has first been suggested by Farrell and Gibbons (1989).

of treatments. An interest group and voters first (repeatedly) interact in a traditional media environment (with public communication and transparent interest group types), after which we maintain the same environment or introduce one of the following three interventions: (i) the implementation of a microtargeting technology, which enables the interest group to send a different message to the a Minority voter as to Majority voters, (ii) obfuscation of interests, which implies that the type of the interest group becomes undisclosed to voters and (iii) a combination of (i) and (ii), which represents the social media environment. The main finding of this second set of treatments is that all voters become worse off due to the transition from the traditional media environment to the social media environment: the efficiency of voter decision-making decreases for both the larger Majority voter group and the smaller Minority voter group.

As pointed out by Strömberg (2004), a media environment is not neutral and gives rise to "a specific distribution of informed and uninformed voters." The third and final goal of this paper is to study the efficiency of voter decision-making of Minority and Majority voters *within* each of the four media environments, which differ from each other in the communication mode—public (one message for all voters) or microtargeted (one message for each voter type)—and the transparency of the interest group type, which is disclosed or undisclosed to voters. We find that the efficiency of voter decision-making is similar for Majority and Minority voters in public media environments. In microtargeted media environments, Minority voters choose less-efficient voting actions than Majority voters. These findings suggest that microtargeting technology may especially have a (negative) impact on voting behavior of smaller, 'niche' voter groups. Our results are consistent with experimental findings by Battaglini and Makarov (2014) and Drugov et al. (2017), who studied information transmission in cheap talk games with public and private (microtargeted) communication and *transparent* sender types. We add to this literature by considering games with *undisclosed* sender (interest group) types and find similar results.

Our experimental study of voting behavior in different media environments has a number of advantages compared to research based on observational data. First, 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. Second, we precisely observe to what information voters are exposed. Outside of the lab, it is impossible to observe all sources of information that voters use to make their voting decision. Third, we can directly elicit beliefs, which allows us to study how communication affects voter beliefs and how voting actions are driven by voter beliefs.

Our 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 our analysis of the individual games. Sections 5 and 6 discuss our analysis across games. In Section 5, we study our interventions within the social media environment. Section 6 contains our analysis of the traditional media environment and the effects of departures from this environment. Section 7 concludes.

2 Related literature

To the best of our knowledge, there is one other experimental paper on social media and voting.[9] 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.

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.

Battaglini and Makarov (2014) implement the framework by Farrell and Gibbons (1989) in the laboratory. Their findings are consistent with the theoretical predictions about the pos-

[9] Kartal and Tyran (2020) 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.

sibility of information transmission in public and private (microtargeted) cheap talk games with different payoff structures. Drugov et al. (2017) confirm these results for experimental cheap talk games with five possible states, messages and actions. We consider both public and private games in which the type of the sender is undisclosed, which alters the optimal strategy of all players.

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.[10] 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 political 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. Fujiwara et al. (2020) and Rotesi (2019) exploit exogenous variation in regional Twitter adoption to study the effect of social media on voting behavior. Fujiwara et al. (2020) 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.

[10] See Zhuravskaya et al. (2020) and Persily and Tucker (2020) for literature reviews about social media and political outcomes.

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.[11] 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 interest group that observes the true state.[12] 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.[13] 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).[14]

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

[11] 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.

[12] 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.

[13] Voting game experiments with three voters are quite common. See Großer (2020) for an overview.

[14] 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)

		State	
Type	Vote	One	Two
Majority	X	30	10
	Y	10	30
	A	20	20
Minority	X	10	30
	Y	30	10
	A	20	20

(b) Voter payoffs

		State	
Type	Vote	One	Two
Majority	X	90	27
	Y	27	90
	A	60	60
Minority	X	27	90
	Y	90	27
	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).

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 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.[15]. 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

[15]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.

equilibrium, in which the most information is transmitted from the interest group to voters.[16] Before we informally describe the equilibria of the games, we first introduce some 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 make use of the binary variables *vote efficiency* and *vote inefficiency* to measure whether a voter chooses a voting action that yields the maximal (minimal) voter payoff or not. Abstention from voting (voting action A) is neither an efficient nor an inefficient vote. Aggregate efficiency is measured by the variables *election efficiency* and *election inefficiency*, which indicate whether the party favored by a majority of the voters receives the majority (minority) of the votes. Concretely, an election outcome is efficient (inefficient) if party X wins (loses) in State One or if party Y wins (loses) in State Two. An election outcome is neither efficient nor inefficient if both parties receive the same number of votes.

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 Predictions for individual games

In this section, we describe interest group behavior (truth) and voter behavior (trust) in the most informative equilibria of our four games. We then derive a hypothesis concerning the efficiency of voter decision-making in each game.

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

[16] All cheap talk games also have a completely uninformative 'babbling' equilibrium.

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).[17]

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.[18] 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).[19]

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

Based on these most informative equilibria of the four games, we derive Hypothesis 1 regarding the efficiency of voting behavior. To keep our hypotheses brief, we do not discuss vote efficiency and vote inefficiency separately. If we state a predicted increase in the efficiency of individual voter decision-making, we mean to convey that we expect an *increase*

[17] In the most informative equilibrium, an expected payoff-maximizing voter always abstains if a message is unrelated to the true state.

[18] Farrell and Gibbons (1989) call these cases, respectively, one-sided discipline and subversion.

[19] 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 $1/2$.

in vote efficiency and a *decrease* in vote inefficiency. Analogously, a predicted increase in the efficiency of aggregate voter decision-making implies that we expect both an *increase* in election efficiency and a *decrease* in election inefficiency.

Hypothesis 1 (Efficiency within games)**.**

a. ***Vote efficiency.*** *In Public games, there is no difference in the average efficiency of voter decision-making for Minority and Majority voters. In Microtargeting games, voter decision-making is on average less efficient for Minority voters than Majority voters.*

b. ***Election efficiency.*** *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 equality in the efficiency of voter decision-making of Minority and Majority voters, regardless of the disclosure regime in place (Hypothesis 1a). This equality is expected to disappear if the interest group gets access to a microtargeting technology (Hypothesis 1a). 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 is less (more) efficient 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 efficiency of voter decision-making is 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 1b).

3.4 Predictions across games: interventions social media environment

The transmission of political information from interest groups to voters on social media platforms is, arguably, characterized by the possibility to microtarget news and obfuscate interests. These key elements of the social media environment are captured in the MU-game. The main goal of our experiment is to study the effect of the implementation of (i) a microtargeting ban (a shift from the MU-game to the PU-game), (ii) mandatory disclosure of interests (a shift from the MU-game to the MD-game) and (iii) a combination of the two interventions (a shift from the MU-game to the PD-game) on the efficiency of voting actions. Hypothesis 2 summarizes our theoretical predictions regarding these interventions. These predictions are based on the equilibria of the four games, which are described in Section 3.3.

Hypothesis 2 (Microtargeting ban, mandatory disclosure of interests and efficiency)**.**

a. ***Vote efficiency.*** *A microtargeting ban increases the average efficiency of voter decision-making for Minority voters and has no effect for Majority voters. Mandatory disclosure of interests, in combination with or without a microtargeting ban, increases the average efficiency of voting for both voter groups.*

b. ***Election efficiency.*** *The average efficiency of aggregate voter decision-making decreases due to a microtargeting ban without mandatory disclosure of interests but increases as a result of a microtargeting ban in combination with disclosure. The effect of mandatory disclosure of interests without a microtargeting ban is ambiguous.*

Hypothesis 2a contains the predicted effects of the interventions on the efficiency of individual voters. 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, which enables them to choose, on average, better-informed voting actions. Mandatory disclosure of interests, in combination with or without a microtargeting ban, 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.

Only a combination of the two interventions under study has an unambiguously positive expected effect on the efficiency of aggregate voter decision-making (Hypothesis 2b). According to our theory, 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. Mandatory disclosure of interests without a microtargeting ban has an ambiguous expected effect on the aggregate level. Whereas 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. Although the microtargeting ban without disclosure of interests has a (weakly) positive expected effect for 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.

3.5 Predictions across games: interventions traditional media environment

In the final part of our paper, 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. 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 the two changes (a shift from the PD-game to the MU-game) within our framework and test the following hypothesis.[20]

Hypothesis 3 (Microtargeting implementation, obfuscation of interests and efficiency)**.**

a. ***Vote efficiency.*** *The average efficiency of voter decision-making decreases for Minority voters and remains unchanged for Majority voters due to microtargeting implementation. Obfuscation of interests, in combination with or without microtargeting implementation, decreases the average efficiency of voting for both voter groups.*

b. ***Election efficiency.*** *Microtargeting implementation, obfuscation of interests and a combination of the two interventions decrease the average efficiency of aggregate voter decision-making.*

[20]The predictions in Hypothesis 3 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 3a). This prediction can be seen as the mirror-image of Hypothesis 2a. On the aggregate level, any change to the traditional media environment is predicted to unavoidably lead to a drop in efficiency (Hypothesis 3b). This prediction is in line with Hypothesis 2b, 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 refer to the previous Section for an explanation of our predictions.

3.6 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.[21] 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.

Recall that the aim of our experiment is threefold. The first main goal is to study whether the implementation of (i) a microtargeting ban, (ii) mandatory disclosure of interests and (iii) a combination of the three interventions that alter the social media environment improve the efficiency of voter decision-making. For this purpose, we use the first four treatments listed in Table 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). Our second main goal is to analyze the effect of changes to the traditional media environment (PD) on the efficiency of voting behavior. In this part of our study, we

[21] 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.

make use of the second four treatments from Table 2: all subjects start in the traditional media environment (PD) and then either move to an alternative environment (PU, MD or MU) or stay in the same game (PD). Our third and final main goal is to test our theoretical predictions *within* each of the four games. We use all treatments in Table 2 but exclude data from Part-II games because this part of the experiment may have been influenced by Part-I behavior, which would distort our findings.

Table 2: Treatment overview

Part I	Part II	Description	Number of sessions	Number of subjects
MU	MD	Mandatory disclosure	3	36
MU	PU	Microtargeting ban	3	36
MU	PD	Mandatory disclosure and microtargeting ban	3	36
MU	MU	Status quo microtargeting and undisclosed interests	3	36
PD	PU	Obfuscation of interests	3	36
PD	MD	Microtargeting implementation	3	36
PD	MU	Obfuscation of interests and microtargeting implementation	3	36
PD	PD	Status quo public communication and disclosed interests	3	36
MD	MU		3	36
MD	PD		3	36
PU	MU		3	36
PU	PD		3	36
		Total	36	432

3.7 Experimental procedures

The experiment was programmed in zTree (Fischbacher, 2007) and was conducted at the computer 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, risk elicitation task and survey. Detailed instructions for one of our treatments (MU-PU) can be found in Appendix A.11. Subject earnings were 42.15 euros on average.

4 Analysis individual games

We first present the results of the individual games and address Hypothesis 1. In the next section, we analyze how interventions in the social media environment influence voter decision-making and test Hypothesis 2. In Section 6, we shift our attention to the study of departures from the traditional media environment and test Hypothesis 3.

For our study of the individual games, we estimate a linear random effects model for each of our measures.[22] This approach accounts for the dependency in our data which arises because we have repeated observations on each of 432 subjects.[23] All our models have the following form.

$$\begin{gathered} y_{it} = \alpha + \beta Type_{it} + u_i + \epsilon_{it}, \\ i = 1..., 432, \quad t = 21, ..., 40. \end{gathered} \tag{1}$$

In this model, an observation for one of our binary outcome variables is given by y_{it}, with subject indicator i and round indicator t. We consider the outcome measures truth, election efficiency and election inefficiency for subjects in the role of interest group and trust, vote efficiency and vote inefficiency for subjects in the role of voter. The intercept α 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

[22] For ease of interpretation of the coefficients, we present a linear probability model despite considering solely binary outcome variables. Our estimations of a subject-level random effects probit model and a three-level mixed effects probit model produce qualitatively similar results.

[23] 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 respondents 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.

otherwise. Furthermore, (1) includes a subject-specific random effect u_i and error term ϵ_{it}.

Our coefficient of interest is β, which quantifies the difference in the outcome rate between a Minority and a Majority (interest group or voter) type. To estimate our models, we use data from Part-I games only because we cannot rule out that behavior in Part-II games is influenced by the play in the preceding Part-I games. We restrict our analysis to the second half (rounds 21-40) of the Part-I games to purge the data of learning effects.[24]

Table 3 displays our results regarding interest group behavior (truth) in each of the four games.[25] 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.[26] This marginal effect corresponds to the coefficient β in Equation 1. We present our results about individual voter behavior (Tables 4 and 5) and aggregate election outcomes (Table 6) in the same way.[27]

Table 3: Summary statistics and marginal effects: truth-telling by interest groups - Part I

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

[24]Behavior in the first half of the experiment is more volatile. Qualitatively, however, our core results are the same if we include all rounds.

[25]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 20 in the Appendix.

[26]In addition to this table, we present the observed interest group strategy profiles in Table 16.

[27]The distribution of individual voting behavior profiles is presented in Table 17.

Table 4: Summary statistics and marginal effects: trust of voters - Part I

	Game	Measure	Interest group	Majority voter Mean	Majority voter St. Dev.	Minority voter Mean	Minority voter St. Dev.	Marginal effect Coeff.	Marginal effect S. E.	Marginal effect Sig.	Marginal effect 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 1 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 5: Summary statistics and marginal effects: efficiency of individual voter decision-making - Part I

	Game	Measure	Interest group	Majority voter Mean	Majority voter St. Dev.	Minority voter Mean	Minority voter St. Dev.	Marginal effect Coeff.	Marginal effect S. E.	Marginal effect Sig.	Marginal effect Pred.
1	MD	Vote efficiency	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		Vote inefficiency	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	0.02	0.26	0.08	0.13	0.03	***	+
7	MU	Vote efficiency	Majority	0.60	0.13	0.39	0.07	-0.21	0.04	***	−
8			Minority	0.36	0.15	0.46	0.14	0.10	0.04	***	+
9			All	0.52	0.09	0.41	0.07	-0.11	0.03	***	−
10		Vote inefficiency	Majority	0.13	0.07	0.25	0.06	0.12	0.03	***	+
11			Minority	0.39	0.16	0.19	0.07	-0.21	0.03	***	−
12			All	0.21	0.05	0.23	0.05	0.02	0.02		+
13	PD	Vote efficiency	Majority	0.78	0.10	0.69	0.09	-0.09	0.03	***	=
14			Minority	0.36	0.12	0.45	0.16	0.08	0.05	*	=
15			All	0.64	0.09	0.61	0.09	-0.03	0.03		=
16		Vote inefficiency	Majority	0.10	0.05	0.15	0.05	0.05	0.02	**	=
17			Minority	0.25	0.11	0.27	0.09	0.01	0.04		=
18			All	0.15	0.06	0.19	0.04	0.04	0.02	**	=
19	PU	Vote efficiency	Majority	0.55	0.20	0.49	0.25	-0.06	0.04		=
20			Minority	0.45	0.19	0.42	0.26	-0.03	0.07		=
21			All	0.52	0.19	0.47	0.24	-0.05	0.04		=
22		Vote inefficiency	Majority	0.17	0.08	0.16	0.08	-0.01	0.04		=
23			Minority	0.30	0.17	0.31	0.18	0.01	0.06		=
24			All	0.21	0.10	0.21	0.10	0.00	0.03		=

Notes. This table displays summary statistics and marginal effects of vote efficiency and vote inefficiency 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 1 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 6: Summary statistics and marginal effects: election efficiency and election inefficiency - Part I

	Game	Measure	Majority interest group Mean	St. Dev.	Minority interest group Mean	St. Dev.	Marginal effect Coeff.	S. E.	Sig.	Pred.
1	MD	Election efficiency	0.86	0.06	0.27	0.10	-0.60	0.05	***	−
2		Election inefficiency	0.07	0.03	0.47	0.10	0.40	0.06	***	+
3	MU	Election efficiency	0.65	0.10	0.35	0.18	-0.30	0.05	***	−
4		Election inefficiency	0.17	0.09	0.48	0.17	0.31	0.05	***	+
5	PD	Election efficiency	0.69	0.12	0.36	0.15	-0.33	0.04	***	−
6		Election inefficiency	0.13	0.05	0.31	0.12	0.19	0.04	***	+
7	PU	Election efficiency	0.54	0.18	0.45	0.13	-0.09	0.07		−
8		Election inefficiency	0.20	0.10	0.31	0.16	0.10	0.07		+

Notes. This table displays summary statistics and marginal effects of election efficiency and election inefficiency 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 1 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 3) and significantly *more* truthful to a Minority voter (estimated coefficient of 0.23; row 2 of Table 3). Considering voters' actions (Table 4), 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 5) and average vote inefficiency is significantly *higher* (row 6). As can be seen from Table 6, 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 3). 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.[28] 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 a Minority voter than a Majority interest group (row 4 of Table 3).[29] Trust is significantly lower for Minority voters than for Majority voters (rows 3 and 4 of Table 4). 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

[28]See footnote 19 for an explanation.

[29]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 16, 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.

voters best respond by trusting a message received, conditional on non-abstention.

Compared to Majority voters, a Minority voter has a significantly higher average vote efficiency (row 6 of Table 5), which is predicted by our theory. There is no statistically significant difference in average vote inefficiency between Majority and Minority voters. 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 7 and 8 of Table 5). Despite the higher ex ante probability that the interest group has type Majority, the latter effect (estimate of -0.21) completely undoes the first effect (estimate of 0.12) because of the higher abstention rate of Minority voters.[30] Aggregate effects are as expected: election efficiency (inefficiency) is significantly lower (higher) with a Minority interest group than with a Majority interest group (rows 3 and 4 of Table 6).

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 3). In contrast to our prediction, Minority voters have significantly lower trust than Majority voters if the interest group has type Majority and vice versa with a Minority interest group (rows 5 and 6 of Table 4). 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 15 of Table 5). For the average vote inefficiency, the difference between Minority and Majority voters is statistically significant but small (estimated difference of 0.04; row 18 of Table 4). 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 6).

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 3). 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

[30] Since voting action A is neither efficient nor inefficient, abstention can be calculated as $1 -$ mean of vote efficiency $-$ mean of vote inefficiency. The average abstention rate in the MU-game is $1 - 0.52 - 0.21 = 0.27$ for a Majority voter and $1 - 0.41 - 0.23$ for a Minority voter.

voter (estimated difference of -0.05; $p < 0.10$).[31] 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 21 and 24 of Table 5). Surprisingly, we find that election efficiency and inefficiency are not significantly different for Minority and Majority interest groups (rows 7 and 8 of Table 6). 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 1.[32] We derive the following result based on the discussion of the individual games above.

Result 1.

a. ***Vote efficiency.*** *In Public games, the average efficiency of voter decision-making is similar for Minority and Majority voters. In Microtargeting games, Minority voters have a lower average individual efficiency 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 1a and Result 1a. 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 1a and Result 1a). These

[31] 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 4. 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.)

[32] We measure the efficiency of individual voter decision-making by two variables (vote efficiency and vote inefficiency) and, hence, test two hypotheses regarding efficiency on the individual level. Analogously, we test two hypotheses regarding the efficiency of aggregate voter decision-making. To take into account that we have tested multiple hypotheses, we adjust the significance threshold levels by applying the Bonferroni correction (Bonferroni, 1935). This correction implies a division of the significance threshold level by the number of tests. We report our marginal effects with corrected significance levels in Tables 12 and 13 in the Appendix. Qualitatively, our key results do not change as a result of the Bonferroni correction.

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 performed reasonably 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 1b and Result 1b).

The observed behavior *within* the four games is fairly similar to our theoretical predictions (see Section 3.3). For our study of interventions to prevent election manipulation through social media, however, we need to know how interest groups and voters behave when the communication environments (games) change. That is why we turn our attention to behavior *across* games in the next section.

5 Analysis across games: interventions in the social media environment

In this section, we study whether the efficiency of voting actions in a social media environment with microtargeted 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 aim is to analyze the effects of the three interventions compared to the status quo (MU-game) *after* subjects have been locked into the social media environment. That is why we 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 estimate the following random effects model for each of our measures.[33]

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

In (2), y_{ith} is the binary outcome of one of our six measures in round t of part h for

[33]For ease of interpretation, we present a linear probability model. We have also estimated a subject-level random effects probit model, which produces very similar estimates and gives rise to the same qualitative results.

subject i.[34] The intercept α 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. Table 7 reports the coefficient estimates (marginal effects) of β_1, β_2 and β_3.[35] These estimates capture the marginal effects of the interventions (PU, MD and PD) compared to (the effect of) the status quo (MU-game in Part II). 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 ϵ_{ith}.

Microtargeting ban. We first study the effects of a shift from the MU-game to the PU-game, which is represented by β_1 in (2). Recall that, in equilibrium, the public message coincides with the microtargeted message directed at Majority voters (see Section 3.3). Indeed, we do not observe any significant differences in truth to Majority voters (rows 1 and 3 of Table 7). At odds with this reporting behavior, Majority voters display a lower level of trust after the introduction of the microtargeting ban (rows 5 and 7 of Table 7).[36] This drop in trust, however, does not lead to a significant change in the efficiency of voting behavior of a Majority voter: on average, the microtargeting ban has no statistically significant effect on vote efficiency and inefficiency (rows 13 and 19 of Table 7).

Minority voters are expected to benefit from the discipline effect of the public communication technology, which forces the interest group to report the same message to a Minority voter as to Majority voters (Hypothesis 2a). This prediction is not borne out by the data. As predicted, the Majority interest group becomes significantly more truthful to Minority voters(row 2 of Table 7).[37] Minority voters, however, do not trust the message from the Majority interest group significantly more after the implementation of the ban (row 6 of

[34]As before, our dependent variables are truth, election efficiency and election inefficiency for subjects in the role of interest group and trust, vote efficiency and vote inefficiency for subjects in the role of voter. We have observations on 144 subjects divided into twelve sessions. As in Section 4, we only consider rounds 21-40 to account for learning effects.

[35]Summary statistics for Part II of the experiment can be found in Section A.6. In this Section, we only present the estimated marginal effects and omit the regressions. We present the regression output for one of our measures (trust) in Table 21 in the Appendix.

[36]This effect is only significant if the interest group has type Minority.

[37]Looking at individual strategy profiles in Table 17, it becomes apparent that a quarter of the subjects in the role of Majority interest group in the PU-game follows a mixing strategy (i.e., these subjects are sometimes truthful and sometimes report a lie), which is at odds with our prediction. In the PD-game, this mixing behavior is much less prevalent: only 7% of the subjects in the Majority interest group role follow this strategy (Table 16). This difference between the two public games suggests that the lack of disclosure causes the high frequency of mixing. Some subjects in the role of majority interest group may anticipate that voters cannot distinguish them from the untrustworthy Minority interest group type but fail to take into account that, ex ante, an interest group is more likely to have type Majority than type Minority, which makes truth-telling the optimal strategy.

Table 7). As a result, the microtargeting ban does not make Minority voters better off: the estimated average effects of the ban on our individual efficiency measures are not significant (rows 14 and 20 of Table 16).

In line with our prediction (Hypothesis 2b), 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. Since the desired outcome for the most prevalent interest group type (Majority) is the efficient election outcome, this intervention has a significantly negative average effect on election efficiency (row 23 of Table 7). We do not observe a similar (significant) effect on election inefficiency (row 26 of Table 7).

Mandatory disclosure of interests. The second intervention under analysis is a shift from the MU-game to the MD-game. According to theory, both voter groups benefit from this change to the social media environment if the interest group has type Minority. Majority voters can no longer be led astray and Minority voters can now rely on the truthfulness of the message received. In line with this prediction, we observe a higher average truth of the messages sent by the Minority interest group (rows 3 and 4 of Table 7, less trust by a Majority voter (row 7) and more trust by a Minority voter (row 8).

Our theory predicts that mandatory disclosure of interest does not alter equilibrium behavior if the interest group has type Majority (see Section 3.3). Indeed, truth by the interest group and trust by the Minority voter do not significantly change after the intervention (rows 1, 2 and 6 of Table 7). Majority voters, however, display a significantly and sizeably higher level of trust due to disclosure (row 5 of Table 7).[38]

The changes in interest group and voter behavior lead to significant improvements in voter decision-making by the Majority voters (rows 13 and 19 of Table 7). We also find an increase in vote efficiency (row 14) and a decrease in vote inefficiency (row 20) for the Minority voters but these effects are not significantly different from zero.

On aggregate, we find that mandatory disclosure of interests does neither lead to a significant increase in efficient election outcomes nor a significant decrease in inefficient election outcomes (rows 23 and 26 of Table 7). These results are consistent with our theory, which predicts an ambiguous aggregate effect (Hypothesis 2).

Microtargeting ban and mandatory disclosure of interests. The third intervention implements a shift from the MU-game to the PD-game. Our predictions about the effects of this intervention on interest group behavior are all borne out by the data. The disciplining

[38]Looking at individual voter strategy profiles (Table 17), we find that 38% of the subjects in the role of Majority voter follow a mixing strategy in the MU-game, whereas only 6% of the subjects mix in the MD-game. This suggests that a sizeable fraction of the subjects takes insufficiently into account that the interest group is ex-ante most likely to have type Majority, which makes it a best response to trust.

effect of the public communication technology causes the Majority interest group to be more truthful to a Minority voter (row 2 of Table 7) but has no effect on its reporting behavior to Majority voters (row 1). In line with this reporting behavior, we see that trust increases for a Minority voter (row 6). At odds with our prediction, we observe that Majority voters also have a higher level of trust due to the intervention (row 5).[39]

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 (see Section 3.3). Both of these predictions are confirmed (rows 3 and 4 of Table 7). 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).

Based on Hypothesis 2, the changes in interest group and voter behavior due to the intervention should unambiguously improve the efficiency of voter decision-making for (i) Majority voters, (ii) Minority voters and (iii) the electorate as a whole. Our experimental data confirm this prediction. On the individual level, vote efficiency significantly increases (rows 13 and 14 of Table 7) and vote inefficiency significantly decreases for both voter types (rows 19 and 20). On the aggregate level, we do not observe a significant rise in efficient election outcomes (row 23). However, the estimated probability of having an inefficient election outcome decreases significantly (row 26).

Summarizing our analysis in this section, we state the following result.[40]

Result 2.

a. ***Vote efficiency.*** *A microtargeting ban has no significant effect on the average efficiency of voter decision-making for Majority and Minority voters. Mandatory disclosure of interests improves the efficiency of voting for Majority voters but not Minority voters. A microtargeting ban in combination with mandatory disclosure of interests has a positive effect on the efficiency of voting for both voter groups.*

b. ***Election efficiency.*** *A microtargeting ban decreases the average efficiency of aggregate voter decision-making. Mandatory disclosure of interests, in combination with*

[39] Considering individual voter strategy profiles (Table 17), we notice that 38% of subjects follow a mixing strategy in the MU-game: they sometimes trust and sometimes do not trust a message received. These subjects may insufficiently take into account that the Majority interest group, which has an incentive to report truthfully, is the most prevalent type. In the PD-game, the mixing strategy is less common. Since we observed a similar pattern in the MU-MD treatment, it is likely that this change in behavior is driven by disclosure of the interest group type.

[40] We report our marginal effects with Bonferroni corrected significance levels to account for multiple hypothesis testing in Table 14 in the Appendix. Qualitatively, our key results do not change as a result of the Bonferroni correction.

or without a microtargeting ban, improves the efficiency of aggregate voter decision-making.

Result 2 contains three key takeaways. First and foremost, the implementation of mandatory disclosure of interests is necessary to improve the efficiency of voter decision-making on the aggregate level (Hypothesis 2b and Result 2b). Our experimental findings show that this intervention, with or without a microtargeting ban, substantially reduces the frequency of inefficient election outcomes (Table 7). Second, only the implementation of mandatory disclosure of interests in combination with a microtargeting ban improves the efficiency of voting actions for both Majority and Minority voters (Result 2a). In contrast to our prediction, mandatory disclosure of interests alone is not sufficient to improve voter decision-making of Minority voters (Hypothesis 2a and Result 2a). Third, the implementation of a microtargeting ban without mandatory disclosure of interests is not beneficial to any voter group and has an adverse aggregate effect (Hypothesis 2 and Result 2). In our experiment, the frequency of efficient election outcomes decreased due to the ban (Table 7).

Table 7: Marginal effects of microtargeting ban and disclosure of interests

		Interaction		Intervention											
	Measure	Interest group	Voters	Microtargeting ban				Disclosure				Combination			
				Coeff.	St. dev.	Sig.	Pred.	Coeff.	St. dev.	Sig.	Pred.	Coeff.	St. dev.	Sig.	Pred.
1	Truth	Majority	Majority	-0.05	0.07		=	0.07	0.05		=	0.07	0.05		=
2			Minority	0.25	0.11	**	+	-0.06	0.12		=	0.34	0.09	***	+
3		Minority	Majority	-0.02	0.14		=	0.23	0.14		+	0.30	0.13	**	+
4			Minority	-0.30	0.14	**	−	0.23	0.11	**	+	0.02	0.13		=
5	Trust	Majority	Majority	-0.09	0.06		=	0.17	0.06	***	=	0.11	0.06	*	=
6			Minority	0.14	0.09		+	0.00	0.10		=	0.44	0.09	***	+
7		Minority	Majority	-0.17	0.08	**	=	-0.35	0.09	***	−	-0.45	0.08	***	−
8			Minority	0.25	0.11	**	+	0.43	0.12	***	+	0.16	0.12		=
9	Vote efficiency	Majority	Majority	-0.08	0.07		=	0.24	0.07	***	+	0.17	0.06	***	+
10			Minority	0.16	0.09	*	+	-0.07	0.08		=	0.51	0.07	***	+
11		Minority	Majority	0.00	0.08		=	0.00	0.08		+	-0.03	0.08		+
12			Minority	-0.05	0.11		−	0.41	0.14	***	+	0.01	0.12		=
13		All	Majority	-0.05	0.06		=	0.15	0.06	**	+	0.09	0.05	*	+
14			Minority	0.08	0.07		+	0.08	0.08		+	0.33	0.07	***	+
15	Vote inefficiency	Majority	Majority	0.00	0.03		=	-0.13	0.03	***	−	-0.13	0.03	***	−
16			Minority	-0.13	0.06	**	−	-0.07	0.07		=	-0.28	0.05	***	−
17		Minority	Majority	0.00	0.08		=	-0.22	0.08	***	−	-0.17	0.08	**	−
18			Minority	0.23	0.10	**	+	-0.07	0.07		−	0.12	0.08		=
19		All	Majority	0.00	0.03		=	-0.13	0.03	***	−	-0.13	0.03	***	−
20			Minority	-0.01	0.05		−	-0.07	0.05		−	-0.15	0.04	***	−
21	Election efficiency	Majority	All	-0.20	0.07	***	−	0.16	0.07	**	+	0.08	0.07		+
22		Minority	All	0.05	0.12		+	-0.08	0.14		−	-0.10	0.11		+
23		All	All	-0.12	0.07	*	−	0.08	0.09		?	0.01	0.09		+
24	Election inefficiency	Majority	All	0.02	0.06		+	-0.09	0.06		−	-0.11	0.05	**	−
25		Minority	All	-0.07	0.11		−	-0.03	0.12		+	-0.23	0.11	**	−
26		All	All	-0.02	0.07		+	-0.07	0.08		?	-0.16	0.07	**	−

Notes. This table reports 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 2 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.

6 Analysis across games: interventions in the traditional media environment

The previous section has shown that the implementation of a microtargeting ban in combination with mandatory disclosure of interests within a social media environment makes all voters better off. In this section, we explore the impact of changes to the traditional media environment (PD-game) within our framework and test Hypothesis 3. Specifically, we analyze the effects of (i) the implementation of microtargeted communication technology (MD-game), (ii) obfuscation of interests (PU-game) and (iii) a combination of these two changes (MU-game) on the efficiency of voter decision-making on the individual and aggregate level compared to the status quo (PD-game).

We adopt a similar approach as in the previous section by only using data from treatments in which subjects start in the traditional media environment (PD-game; treatments five to eight in Table 2). We estimate a random effects model that is represented by the following equation:[41]

$$y_{ith} = \alpha + \beta_1 MD + \beta_2 PU + \beta_3 MU + \beta_4 Part + u_i + \epsilon_{ith}, \quad (3)$$
$$i = 1..., 144, \quad t = 21, ..., 40, \quad h = 1, 2.$$

As in (2), y_{ith} in (3) represents the binary outcome of one of our six measures in round t of part h for subject i.

The intercept α 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 (β_1, β_2 and β_3) compared to the status quo (α) are reported in Table 8.[42] 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 ϵ_{ith}.

Microtargeting implementation. We first consider the efficiency effects of a shift from the PD-game to the MD-game. As expected (Hypothesis 2a), the acquired capacity of the interest group to send separate messages to the two voter groups only affects the Minority voters: average vote efficiency significantly decreases (row 14 of Table 8). This drop in efficiency is caused by the changed reporting behavior of the Majority interest group: being

[41] Our estimates of a subject-level random effects probit model generates similar estimates of the treatment effects but are harder to interpret than the (estimated) linear probability model.

[42] We only report marginal effects and leave out the regressions. The regression output for one of our measures (trust) can be found in Table 22 in the Appendix.

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 8). 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 21, 22 and 23 in Table 8). 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 24, 25 and 26 of Table 8).

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. There are slight drops in vote efficiency and increases in vote inefficiency for both voter types but the estimated coefficient is only significant regarding vote inefficiency for Majority voters (rows 13, 14, 19 and 20 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 18).[43] On the aggregate level, the estimated coefficients regarding election efficiency have the predicted signs but only the coefficient for average election inefficiency is significant (Hypothesis 2b and rows 23 and 26 of Table 8).

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 3). On the individual level, the experimental data are in line with Hypothesis 3a: average vote efficiency and average vote inefficiency significantly increase for both voter types (rows 13, 14, 19 and 20 of Table 8). On the aggregate level, the support for our theoretical prediction (Hypothesis 3b) is milder. Table 8 (rows 23 and 26) reports that the estimated effects of the intervention on election efficiency and inefficiency have the expected signs but that only the latter coefficient is significant (at the 10%-level).

[43] 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.

Based on the experimental findings reported in this section, we state the following key result.[44]

Result 3.

a. ***Vote efficiency.*** *Microtargeting implementation has a negative effect on the average efficiency of voter decision-making for Minority voters but not for Majority voters. Obfuscation of interests has no significant effect on the efficiency of voting for Majority and Minority voters. Obfuscation of interests in combination with microtargeting implementation decreases the average efficiency of voter decision-making for both voter groups.*

b. ***Election efficiency.*** *Obfuscation of interests in combination with microtargeting implementation has a significant negative effect on the efficiency of aggregate voter decision-making at the* 10%*-level only. On their own, the interventions do not have a significant effect.*

Most importantly, Result 3 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 8), which confirms our theoretical prediction (Hypothesis 3a).

As expected (Hypothesis 3a), 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 3a and Result 3a). 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 3b). We only find that microtargeting implementation in combination with obfuscation of interests increases the frequency of inefficient election outcomes (at the 10%-level of significance).

[44] We report our marginal effects with Bonferroni corrected significance levels to account for multiple hypothesis testing in Table 15 in the Appendix. Qualitatively, our key results are the same as without the Bonferroni correction.

Table 8: Marginal effects of microtargeting implementation and obfuscation of interests

		Interaction		Intervention											
	Measure	Interest group	Voters	Microtargeting implementation				Obfuscation of interests				Combination			
				Coeff.	St. dev.	Sig.	Pred.	Coeff.	St. dev.	Sig.	Pred.	Coeff.	St. dev.	Sig.	Pred.
1	Truth	Majority	Majority	0.00	0.04		=	-0.03	0.05		=	0.01	0.04		=
2			Minority	-0.37	0.08	***	−	-0.03	0.05		=	-0.37	0.08	***	−
3		Minority	Majority	-0.15	0.14		=	-0.12	0.14		−	-0.27	0.13	**	−
4			Minority	0.23	0.10	**	+	-0.12	0.14		−	0.02	0.13		=
5	Trust	Majority	Majority	0.09	0.04	**	=	-0.15	0.06	**	=	-0.28	0.06	***	=
6			Minority	-0.51	0.08	***	−	-0.09	0.08		=	-0.44	0.07	***	−
7		Minority	Majority	-0.19	0.09	**	=	0.33	0.09	***	+	0.26	0.10	***	+
8			Minority	0.36	0.11	***	+	0.09	0.10		+	-0.10	0.12		=
9	Vote efficiency	Majority	Majority	0.08	0.05	*	=	-0.13	0.07	*	=	-0.25	0.06	***	=
10			Minority	-0.49	0.07	***	−	-0.08	0.09		=	-0.50	0.07	***	−
11		Minority	Majority	-0.05	0.07		=	0.19	0.07	***	−	0.16	0.07	**	−
12			Minority	0.43	0.12	***	+	0.06	0.11		−	0.02	0.12		=
13		All	Majority	0.05	0.04		=	-0.03	0.06		−	-0.11	0.06	*	−
14			Minority	-0.18	0.07	**	−	-0.04	0.08		−	-0.31	0.07	***	−
15	Vote inefficiency	Majority	Majority	-0.01	0.03		=	0.07	0.04	**	=	0.09	0.03	***	=
16			Minority	0.16	0.05	***	+	0.04	0.04		=	0.27	0.05	***	+
17		Minority	Majority	-0.05	0.06		=	0.14	0.07	**	+	0.12	0.07	*	+
18			Minority	-0.11	0.10		−	0.00	0.10		+	-0.05	0.09		=
19		All	Majority	-0.01	0.03		=	0.07	0.04	**	+	0.09	0.03	***	+
20			Minority	0.06	0.04		+	0.02	0.04		+	0.15	0.04	***	+
21	Election efficiency	Majority	All	0.18	0.05	***	=	-0.14	0.07	**	−	-0.14	0.06	**	−
22		Minority	All	-0.22	0.09	**	−	0.05	0.12		−	0.02	0.11		−
23		All	All	0.04	0.09		−	-0.07	0.07		−	-0.07	0.07		−
24	Election inefficiency	Majority	All	-0.02	0.04		=	0.10	0.04	**	+	0.12	0.05	***	+
25		Minority	All	0.28	0.09	***	+	0.15	0.10		+	0.12	0.08		+
26		All	All	0.07	0.06		+	0.11	0.05	**	+	0.10	0.05	**	+

Notes. This table reports 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 3 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.

7 Conclusion

In this paper, we have developed and experimentally tested a game of interest group-voter communication and voter decision-making in different media environments. In our first set of treatments, we have analyzed the effects of the implementation of a microtargeting ban and mandatory disclosure of interests within a social media environment on the efficiency of voter decision-making. Our experimental results show that mandatory disclosure of interests is an effective measure to prevent inefficient election outcomes. With this intervention in place, the frequency of election outcomes won by a party that was not supported by a majority of the electorate reduced significantly and substantially. The implementation of a microtargeting ban without disclosure of interests, on the other hand, made inefficient election outcomes more common. Only the joint implementation of a microtargeting ban and disclosure of interests was successful in making both voter groups within the electorate better off: our results show that Minority voters only benefit from transparency of the interest group's motives if the disciplining effect of the public communication technology is in place. Going beyond our framework, our results suggest that disclosure requirements can prevent election manipulation by malevolent interest groups by allowing voters to take into account the motives of the sender when evaluating the credibility of the political message received. This measure could be implemented in combination with a microtargeting ban as a way to promote equal (access to) political information across voter groups.

In a second set of treatments, we have shown that a transition from a traditional to a social media environment makes both Majority and Minority voters worse off. This finding suggests that allowing interest groups to hide their motives and microtarget their messages is harmful to all voter groups, regardless of their size.

Lastly, we have compared the efficiency of voter decision-making between the smaller Minority voter group and the larger group of Majority voters within different media environments. Whereas Minority and Majority voters are roughly equally efficient in public communication environments, Minority voters stay behind in environments with microtargeting. To the extent that our experimental results have meaning outside the laboratory, we conjecture the following to hold more generally. First, political advertising through a public media technology, such as television commercials, will affect all voters in the same way: all voter groups will make equally informed voting decisions. Second, microtargeted political advertising, such as personalized social media advertising, will create inequality among voter groups: voters with 'mainstream' political ideologies will choose better-informed voting actions than voters belonging to 'niche' voter groups.[45]

[45] At the same time, the advent of social media platforms as news distribution channels may benefit smaller voter groups. Strömberg (2004) argues that profit-maximizing mass media outlets predominantly cover policy issues that concern voter groups due to the necessity to recoup the high fixed costs of news provision. He finds that large voter groups are better-informed and enjoy more favorable policy outcomes as a result of this cost structure. Since news outlets that operate on social media platforms tend to have lower fixed costs, they might also find it worthwhile to cover issues that concern smaller audiences, which may improve the policy outcomes of minority voter groups. A key difference between our model and Strömberg (2004) is that we consider purely ideologically motivated suppliers of political information, whereas Strömberg (2004) focuses on profit-maximizing news outlets.

References

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

Battaglini, M. and Makarov, U. (2014). Cheap talk with multiple audiences: An experimental analysis. *Games and Economic Behavior*, 83:147–164.

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D., Marlow, C., Settle, J. E., and Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature*, 489(7415):295–298.

Bonferroni, C. E. (1935). Il calcolo delle assicurazioni su gruppi di teste. In *Studi in onore del professore Salvatore Ortu Carboni*, pages 13–60.

Cai, H. and Wang, J. T.-Y. (2006). Overcommunication in strategic information transmission games. *Games and Economic Behavior*, 56(1):7–36.

Cain, D. M., Loewenstein, G., and Moore, D. A. (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 Harbaugh, R. (2019). Biased recommendations from biased and unbiased experts. *Journal of Economics & Management Strategy*, 28(3):520–540.

Drugov, M., Hernan Gonzalez, R., Kujal, P., and Troya-Martinez, M. (2017). Cheap talk with two audiences: An experiment. Working paper, Available at http://dx.doi.org/10.2139/ssrn.3086405.

Farrell, J. and Gibbons, R. (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 Tirole, J. (1991). Perfect bayesian equilibrium and sequential equilibrium. *Journal of Economic Theory*, 53(2):236–260.

Fujiwara, T., Müller, K., and Schwarz, C. (2020). 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 Capra, C. M., Croson, R. T., Rigdon, M. L., and Rosenblat, T. S., editors, *Handbook of Experimental Game Theory*, pages 376–398. Edward Elgar Publishing.

Guess, A. M. and Lyons, B. A. (2020). Misinformation, disinformation, and online propaganda. In Persily, N. and Tucker, J. A., editors, *Social Media and Democracy: The State of the Field, Prospects for Reform*, pages 10–33. Cambridge University Press.

Guess, A. M., Nyhan, B., and Reifler, J. (2020). Exposure to untrustworthy websites in the 2016 US election. *Nature human behaviour*, 4(5):472–480.

Holt, C. A. and Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, 92(5):1644–1655.

Jones, J. J., Bond, R. M., Bakshy, E., Eckles, D., and Fowler, J. H. (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 Newman, N. (2017). I saw the news on Facebook. Reuters Institute for the Study of Journalism.

Kang, H., Bae, K., Zhang, S., and Sundar, S. S. (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 Tremewan, J. (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 Tyran, J.-R. (2020). Fake news, voter overconfidence, and the quality of democratic choice. Working paper, Available at http://dx.doi.org/10.2139/ssrn.3637866.

Kim, D. (2020). Clustering standard errors at the "session" level. Working Paper, Available at http://dx.doi.org/10.2139/ssrn.3635181.

Liberini, F., Russo, A., Cuevas, Á., and Cuevas, R. (2020). Politics in the Facebook era. Evidence from the 2016 US presidential elections. Working Paper, Available at https://ssrn.com/abstract=3584086.

Newman, N., Fletcher, R., Schulz, A., Andi, S., Robertson, C. T., and Nielsen, R. K. (2021). Reuters institute digital news report 2021. Reuters Institute for the Study of Journalism.

Nyarko, Y. and Schotter, A. (2002). An experimental study of belief learning using elicited beliefs. *Econometrica*, 70(3):971–1005.

Persily, N. and Tucker, J. A. (2020). *Social Media and Democracy: The State of the Field, Prospects for Reform*. Cambridge University Press.

Pogorelskiy, K. and Shum, M. (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.

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.

Strömberg, D. (2004). Radio's impact on public spending. *The Quarterly Journal of Economics*, 119(1):189–221.

Sun, J., Schram, A., and Sloof, R. (2021). Elections under biased candidate endorsements — an experimental study. *Games and Economic Behavior*, 125:141–158.

Tucker, J. A., Theocharis, Y., Roberts, M. E., and Barberá, P. (2017). From liberation to turmoil: Social media and democracy. *Journal of Democracy*, 28(4):46–59.

Van Gils, F., Müller, W., and Prüfer, J. (2020). Big data and democracy. TILEC Discussion Paper No. DP 2020-003, Available at http://dx.doi.org/10.2139/ssrn.3556512.

Zhuravskaya, E., Petrova, M., and Enikolopov, R. (2020). Political effects of the internet and social media. *Annual Review of Economics*, 12:415–438.

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

A Appendix

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.

The timing of our four games is as follows. First, nature determines θ and s. The interest group observes θ and s. Voters do not observe θ 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.

The solution concept for our games is the Perfect Bayesian Equilibrium (Fudenberg and Tirole, 1991). We limit ourselves to the voter payoff-maximizing most informative equilibrium, in which the most information is transmitted from the interest group to voters and in which voters' expected payoffs are maximized. 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 $\mu^*(m = One) \geq \mu^*(m = Two)$.[46] 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. To facilitate the derivation of the most informative equilibria, we assume that a voter who is indifferent between voting (for party

[46]Equilibria formed by a permutation of the messages $m = One$ and $m = Two$ also exist.

X or party Y) and abstaining (voting action $\emptyset$) chooses to abstain in equilibrium.

A.2 Equilibrium MU-game

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

$$m^*(s,\theta) = \begin{cases} \left.\begin{array}{ll} m_{Maj} = One \ \ and \ m_{Min} = \tilde{m} & if \ \theta = One \\ m_{Maj} = Two \ \ and \ m_{Min} = \tilde{m} & if \ \theta = Two \end{array}\right. & for \ \ s = Majority \\ \\ \left.\begin{array}{ll} m_{Maj} = Two \ \ and \ m_{Min} = \tilde{m} & if \ \theta = One \\ m_{Maj} = One \ \ and \ m_{Min} = \tilde{m} & if \ \theta = Two \end{array}\right. & for \ \ s = Minority \end{cases} \tag{4}$$

$$d^*(r,m_r) = \begin{cases} \left.\begin{array}{ll} X & if \ m_{Maj} = One \\ Y & if \ m_{Maj} = Two \end{array}\right. & for \ \ r = Majority \\ \\ \emptyset & for \ r = Minority \end{cases} \tag{5}$$

$$\mu^*(r,m_r) = \begin{cases} \left.\begin{array}{ll} \frac{2}{3} & if \ m_{Maj} = One \\ \frac{1}{3} & if \ m_{Maj} = Two \end{array}\right. & for \ \ r = Majority \\ \\ \frac{1}{2} & for \ r = Minority \end{cases} \tag{6}$$

Proof. Given (4), it follows from Bayes' rule that voters' posterior beliefs are as specified in (6). 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 (6) and Table 1b, we find that conditions (i) and (ii) are fulfilled if a voter follows the voting rule stated in (5). Using (5) and Table 1a, we find that an interest group cannot increase its payoff by deviating from the strategy stated in (4). Thus, the game has an equilibrium in which the interest group follows the strategy in (4) and voters vote according to (5) and update their beliefs according to (6). Since $\mu^*\left(r = Majority, m_r = One\right) > \mu^*\left(r = Majority, m_r = Two\right)$, the equilibrium is informative. In any equlibrium in which $\mu^*\left(r = Majority, m_r = One\right) > \mu^*\left(r = Majority, m_r = Two\right)$, the interest group has an incentive to follow the reporting strategy specified in (4) (see (5) and Table 1a). It then follows from Bayes' rule that the beliefs of a Majority voter must be as stated in (6). 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 (6). Suppose this is not the case (i.e., $\mu^*\left(r = Minority, m_r = One\right) > \mu^*\left(r = Minority, m_r = Two\right)$). 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 (4)-(6) is also the most informative equilibrium. $\square$

A.3 Equilibrium MD-game

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

$$m^*(s,\theta) = \begin{cases} \left.\begin{array}{ll} m_{Maj} = One \text{ and } m_{Min} = \tilde{m} & \text{if } \theta = One \\ m_{Maj} = Two \text{ and } m_{Min} = \tilde{m} & \text{if } \theta = Two \end{array}\right. & \text{for } s = Majority \\ \\ \left.\begin{array}{ll} m_{Maj} = \tilde{m} \text{ and } m_{Min} = One & \text{if } \theta = One \\ m_{Maj} = \tilde{m} \text{ and } m_{Min} = Two & \text{if } \theta = Two \end{array}\right. & \text{for } s = Minority \end{cases} \tag{7}$$

$$d^*(r,s,m_r) = \begin{cases} \left.\begin{array}{ll} X & \text{if } m_{Maj} = One \text{ and } s = Majority \\ Y & \text{if } m_{Maj} = Two \text{ and } s = Majority \\ \emptyset & \text{if } s = Minority \end{array}\right. & \text{for } r = Majority \\ \\ \left.\begin{array}{ll} Y & \text{if } m_{Min} = One \text{ and } s = Minority \\ X & \text{if } m_{Min} = Two \text{ and } s = Minority \\ \emptyset & \text{if } s = Majority \end{array}\right. & \text{for } r = Minority \end{cases} \tag{8}$$

$$\mu^*(r,s,m_r) = \begin{cases} \left.\begin{array}{ll} 1 & \text{if } m_{Maj} = One \text{ and } s = Majority \\ 0 & \text{if } m_{Maj} = Two \text{ and } s = Majority \\ \frac{1}{2} & \text{if } s = Minority \end{array}\right. & \text{for } r = Majority \\ \\ \left.\begin{array}{ll} 1 & \text{if } m_{Min} = One \text{ and } s = Minority \\ 0 & \text{if } m_{Min} = Two \text{ and } s = Minority \\ \frac{1}{2} & \text{if } s = Majority \end{array}\right. & \text{for } r = Minority \end{cases} \tag{9}$$

Proof. Given (7), it follows from Bayes' rule that voter beliefs are as specified in (9). Given (9) and using Table 1b, the maximization of the expected voter payoff implies that voters follow the voting rule in (8). Given (8) and using Table 1a, we find that no interest group can profitably deviate by changing the reporting strategy stated in (7). Thus, the game has an equilibrium in which the interest group acts according to (7) and the voters follows (8) and (9). 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, (8) and (7)), which would be inconsistent with equilibrium voter beliefs. Thus, the equilibrium described in (7)-(9) is the most informative equilibrium of the MD-game. □

A.4 Equilibrium PU-game

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

$$m^*(s, \theta) = \begin{cases} \left.\begin{array}{ll} One & \text{if } \theta = One \\ Two & \text{if } \theta = Two \end{array}\right. & \text{for } \; s = Majority \\ \\ \left.\begin{array}{ll} Two & \text{if } \theta = One \\ One & \text{if } \theta = Two \end{array}\right. & \text{for } \; s = Minority \end{cases} \tag{10}$$

$$d^*(r, m) = \begin{cases} \left.\begin{array}{ll} X & \text{if } m = One \\ Y & \text{if } m = Two \end{array}\right. & \text{for } \; r = Majority \\ \\ \left.\begin{array}{ll} Y & \text{if } m = One \\ X & \text{if } m = Two \end{array}\right. & \text{for } \; r = Minority \end{cases} \tag{11}$$

$$\mu^*(m) = \begin{cases} \frac{2}{3} & \text{if } m = One \\ \frac{1}{3} & \text{if } m = Two \end{cases} \tag{12}$$

Proof. Given (10), it follows from Bayes' rule that voter beliefs are as specified in (12). Given (10) and using Table 1b, the maximization of the expected payoffs implies that voters follow the voting rule given in (11). Given (11) and using Table 1a, we find that no interest group can profitably deviate by changing the reporting strategy in (10). Thus, there is an equilibrium described by (10)-(12). Since $\mu^*(m = One) > \mu^*(m = Two)$, the equilibrium is informative. The interest group has an incentive to report as stated in (10) in any equilibrium in which $\mu^*(m = One) > \mu^*(m = Two)$. It then follows from Bayes' rule that the beliefs must be as given in (12). 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 voter payoff-maximizing most informative perfect Bayesian equilibrium of the PD-game:*

$$m^*(s,\theta) = \begin{cases} \begin{rcases} m = One & \text{if } \theta = One \\ m = Two & \text{if } \theta = Two \end{rcases} & \text{for } s = Majority \\ \\ \tilde{m} & \text{for } s = Minority \end{cases} \tag{13}$$

$$d^*(r,s,m) = \begin{cases} \begin{rcases} X & \text{if } m = One \text{ and } s = Majority \\ Y & \text{if } m = Two \text{ and } s = Majority \\ \emptyset & \text{if } s = Minority \end{rcases} & \text{for } r = Majority \\ \\ \begin{rcases} Y & \text{if } m = One \text{ and } s = Majority \\ X & \text{if } m = Two \text{ and } s = Majority \\ \emptyset & \text{if } s = Minority \end{rcases} & \text{for } r = Minority \end{cases} \tag{14}$$

$$\mu^*(s,m) = \begin{cases} 1 & \text{if } m = One \text{ and } s = Majority \\ 0 & \text{if } m = Two \text{ and } s = Majority \\ \frac{1}{2} & \text{if } s = Minority \end{cases} \tag{15}$$

Proof. Given (13), it follows from Bayes' rule that voter beliefs are as specified in (15). Given (13) and using Table 1b, we find that voters follow the voting rule in (14) to maximize their expected payoff. Given (14) and using Table 1a, we find that no interest group can profitably deviate by changing the reporting strategy stated in (13). Thus, the strategy profiles and beliefs described in (13)-(15) 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, (13) and (14)), which would be inconsistent with equilibrium voter beliefs. Thus, the equilibrium described in (13)-(15) is the most informative equilibrium of the PD-game. $\square$

A.6 Summary statistics Part II

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

Table 9: Summary statistics: truth-telling by interest groups - Part II

		Majority interest group		Minority interest group	
Game	Measure	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 10: Summary statistics: trust, vote efficiency and vote inefficiency by voters - Part II

			Majority voter		Minority voter	
Game	Measure	Interest group	Mean	St. Dev.	Mean	St. Dev.
MD	Trust	Majority	0.95	0.05	0.36	0.11
		Minority	0.34	0.13	0.80	0.10
	Vote efficiency	Majority	0.90	0.09	0.27	0.07
		Minority	0.27	0.10	0.73	0.14
		All	0.69	0.08	0.42	0.06
	Vote inefficiency	Majority	0.06	0.04	0.23	0.03
		Minority	0.20	0.06	0.12	0.12
		All	0.11	0.04	0.19	0.05
MU	Trust	Majority	0.64	0.11	0.40	0.11
		Minority	0.65	0.09	0.40	0.13
	Vote efficiency	Majority	0.57	0.12	0.31	0.09
		Minority	0.34	0.09	0.40	0.15
		All	0.49	0.10	0.34	0.09
	Vote inefficiency	Majority	0.15	0.08	0.27	0.09
		Minority	0.38	0.09	0.15	0.05
		All	0.22	0.07	0.23	0.07
PD	Trust	Majority	0.87	0.10	0.81	0.12
		Minority	0.42	0.19	0.52	0.21
	Vote efficiency	Majority	0.83	0.13	0.77	0.15
		Minority	0.34	0.10	0.38	0.19
		All	0.67	0.11	0.64	0.12
	Vote inefficiency	Majority	0.05	0.04	0.06	0.04
		Minority	0.23	0.06	0.24	0.11
		All	0.11	0.02	0.12	0.04
PU	Trust	Majority	0.67	0.20	0.62	0.19
		Minority	0.69	0.14	0.56	0.23
	Vote efficiency	Majority	0.60	0.21	0.56	0.22
		Minority	0.37	0.17	0.33	0.19
		All	0.52	0.18	0.49	0.19
	Vote inefficiency	Majority	0.12	0.06	0.15	0.08
		Minority	0.41	0.13	0.32	0.13
		All	0.22	0.06	0.21	0.07

Notes. This table displays summary statistics of trust, vote efficiency and vote inefficiency 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.

Table 11: Summary statistics: election efficiency and election inefficiency - Part II

		Majority interest group		Minority interest group	
Game	Measure	Mean	St. Dev.	Mean	St. Dev.
MD	Election efficiency	0.90	0.08	0.18	0.15
	Election inefficiency	0.07	0.04	0.53	0.11
MU	Election efficiency	0.64	0.12	0.33	0.13
	Election inefficiency	0.19	0.11	0.47	0.15
PD	Election efficiency	0.78	0.12	0.33	0.16
	Election inefficiency	0.08	0.05	0.28	0.10
PU	Election efficiency	0.57	0.14	0.38	0.14
	Election inefficiency	0.19	0.11	0.44	0.10

Notes. This table displays summary statistics and marginal effects of election efficiency and election inefficiency 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.7 Marginal effects tables with Bonferroni correction

Table 12: Summary statistics and marginal effects: efficiency of individual voter decision-making - Part I (Bonferroni correction)

				Majority voter		Minority voter		Marginal effect			
	Game	Measure	Interest group	Mean	St. Dev.	Mean	St. Dev.	Coeff.	S. E.	Sig.	Pred.
1	MD	Vote efficiency	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		Vote inefficiency	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	0.02	0.26	0.08	0.13	0.03	***	+
7	MU	Vote efficiency	Majority	0.60	0.13	0.39	0.07	-0.21	0.04	***	−
8			Minority	0.36	0.15	0.46	0.14	0.10	0.04	***	+
9			All	0.52	0.09	0.41	0.07	-0.11	0.03	***	−
10		Vote inefficiency	Majority	0.13	0.07	0.25	0.06	0.12	0.03	***	+
11			Minority	0.39	0.16	0.19	0.07	-0.21	0.03	***	−
12			All	0.21	0.05	0.23	0.05	0.02	0.02		+
13	PD	Vote efficiency	Majority	0.78	0.10	0.69	0.09	-0.09	0.03	***	=
14			Minority	0.36	0.12	0.45	0.16	0.08	0.05		=
15			All	0.64	0.09	0.61	0.09	-0.03	0.03		=
16		Vote inefficiency	Majority	0.10	0.05	0.15	0.05	0.05	0.02	**	=
17			Minority	0.25	0.11	0.27	0.09	0.01	0.04		=
18			All	0.15	0.06	0.19	0.04	0.04	0.02	*	=
19	PU	Vote efficiency	Majority	0.55	0.20	0.49	0.25	-0.06	0.04		=
20			Minority	0.45	0.19	0.42	0.26	-0.03	0.07		=
21			All	0.52	0.19	0.47	0.24	-0.05	0.04		=
22		Vote inefficiency	Majority	0.17	0.08	0.16	0.08	-0.01	0.04		=
23			Minority	0.30	0.17	0.31	0.18	0.01	0.06		=
24			All	0.21	0.10	0.21	0.10	0.00	0.03		=

Notes. This table displays summary statistics and marginal effects of vote efficiency and vote inefficiency 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 1 with cluster-robust standard errors, * $p < 0.05$, ** $p < 0.025$, *** $p < 0.005$. The symbols '+', '−', and '=' in the column Pred. denote that the predicted marginal effect is, respectively, positive, negative, and equal to zero.

Table 13: Summary statistics and marginal effects: election efficiency and election inefficiency - Part I (Bonferroni correction)

			Majority interest group		Minority interest group		Marginal effect			
	Game	Measure	Mean	St. Dev.	Mean	St. Dev.	Coeff.	S. E.	Sig.	Pred.
1	MD	Election efficiency	0.86	0.06	0.27	0.10	-0.60	0.05	***	−
2		Election inefficiency	0.07	0.03	0.47	0.10	0.40	0.06	***	+
3	MU	Election efficiency	0.65	0.10	0.35	0.18	-0.30	0.05	***	−
4		Election inefficiency	0.17	0.09	0.48	0.17	0.31	0.05	***	+
5	PD	Election efficiency	0.69	0.12	0.36	0.15	-0.33	0.04	**	−
6		Election inefficiency	0.13	0.05	0.31	0.12	0.19	0.04	***	+
7	PU	Election efficiency	0.54	0.18	0.45	0.13	-0.09	0.07		−
8		Election inefficiency	0.20	0.10	0.31	0.16	0.10	0.07		+

Notes. This table displays summary statistics and marginal effects of election efficiency and election inefficiency 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 1 with cluster-robust standard errors, * $p < 0.05$, ** $p < 0.025$, *** $p < 0.005$. The symbols '+', '−', and '=' in the column Pred. denote that the predicted marginal effect is, respectively, positive, negative, and equal to zero.

Table 14: Marginal effects of microtargeting ban and disclosure of interests (Bonferroni correction)

	Measure	Interaction		Intervention											
		Interest group	Voters	Microtargeting ban				Disclosure				Combination			
				Coeff.	St. dev.	Sig.	Pred.	Coeff.	St. dev.	Sig.	Pred.	Coeff.	St. dev.	Sig.	Pred.
1	Vote efficiency	Majority	Majority	-0.08	0.07		$=$	0.24	0.07	***	$+$	0.17	0.06	***	$+$
2			Minority	0.16	0.09		$+$	-0.07	0.08		$=$	0.51	0.07	***	$+$
3		Minority	Majority	0.00	0.08		$=$	0.00	0.08		$+$	-0.03	0.08		$+$
4			Minority	-0.05	0.11		$-$	0.41	0.14	***	$+$	0.01	0.12		$=$
5		All	Majority	-0.05	0.06		$=$	0.15	0.06	**	$+$	0.09	0.05		$+$
6			Minority	0.08	0.07		$+$	0.08	0.08		$+$	0.33	0.07	***	$+$
7	Vote inefficiency	Majority	Majority	0.00	0.03		$=$	-0.13	0.03	***	$-$	-0.13	0.03	***	$-$
8			Minority	-0.13	0.06	*	$-$	-0.07	0.07		$=$	-0.28	0.05	***	$-$
9		Minority	Majority	0.00	0.08		$=$	-0.22	0.08	***	$-$	-0.17	0.08	**	$-$
10			Minority	0.23	0.10	*	$+$	-0.07	0.07		$-$	0.12	0.08		$=$
11		All	Majority	0.00	0.03		$=$	-0.13	0.03	***	$-$	-0.13	0.03	***	$-$
12			Minority	-0.01	0.05		$-$	-0.07	0.05		$-$	-0.15	0.04	***	$-$
13	Election efficiency	Majority	All	-0.20	0.07	***	$-$	0.16	0.07	*	$+$	0.08	0.07		$+$
14		Minority	All	0.05	0.12		$+$	-0.08	0.14		$-$	-0.10	0.11		$+$
15		All	All	-0.12	0.07		$-$	0.08	0.09		?	0.01	0.09		$+$
16	Election inefficiency	Majority	All	0.02	0.06		$+$	-0.09	0.06		$-$	-0.11	0.05	*	$-$
17		Minority	All	-0.07	0.11		$-$	-0.03	0.12		$+$	-0.23	0.11	*	$-$
18		All	All	-0.02	0.07		$+$	-0.07	0.08		?	-0.16	0.07	**	$-$

Notes. This table reports 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 2 with robust standard errors, * $p < 0.05$, ** $p < 0.025$, *** $p < 0.005$. The symbols '+', '−','=', and '?' in the column Pred. denote that the predicted marginal effect is, respectively, positive, negative, equal to zero, and ambiguous.

Table 15: Marginal effects of microtargeting implementation and obfuscation of interests (Bonferroni correction)

		Interaction		Intervention											
	Measure	Interest group	Voters	Microtargeting implementation				Obfuscation of interests				Combination			
				Coeff.	St. dev.	Sig.	Pred.	Coeff.	St. dev.	Sig.	Pred.	Coeff.	St. dev.	Sig.	Pred.
1	Vote efficiency	Majority	Majority	0.08	0.05		=	-0.13	0.07		=	-0.25	0.06	***	=
2			Minority	-0.49	0.07	***	−	-0.08	0.09		=	-0.50	0.07	***	−
3		Minority	Majority	-0.05	0.07		=	0.19	0.07	**	−	0.16	0.07	*	−
4			Minority	0.43	0.12	***	+	0.06	0.11		−	0.02	0.12		=
5		All	Majority	0.05	0.04		=	-0.03	0.06		−	-0.11	0.06		−
6			Minority	-0.18	0.07	*	−	-0.04	0.08		−	-0.31	0.07	***	−
7	Vote inefficiency	Majority	Majority	-0.01	0.03		=	0.07	0.04	*	=	0.09	0.03	***	=
8			Minority	0.16	0.05	***	+	0.04	0.04		=	0.27	0.05	***	+
9		Minority	Majority	-0.05	0.06		=	0.14	0.07	*	+	0.12	0.07		+
10			Minority	-0.11	0.10		−	0.00	0.10		+	-0.05	0.09		=
11		All	Majority	-0.01	0.03		=	0.07	0.04	*	+	0.09	0.03	***	+
12			Minority	0.06	0.04		+	0.02	0.04		+	0.15	0.04	***	+
13	Election efficiency	Majority	All	0.18	0.05	***	=	-0.14	0.07	*	−	-0.14	0.06	**	−
14		Minority	All	-0.22	0.09	**	−	0.05	0.12		−	0.02	0.11		−
15		All	All	0.04	0.09		−	-0.07	0.07		−	-0.07	0.07		−
16	Election inefficiency	Majority	All	-0.02	0.04		=	0.10	0.04	**	+	0.12	0.05	***	+
17		Minority	All	0.28	0.09	***	+	0.15	0.10		+	0.12	0.08		+
18		All	All	0.07	0.06		+	0.11	0.05	*	+	0.10	0.05	*	+

Notes. This table reports 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 3 with cluster-robust standard errors, * $p < 0.05$, ** $p < 0.025$, *** $p < 0.005$. The symbols '+', '−', and '=' in the column Pred. denote that the predicted marginal effect is, respectively, positive, negative, and equal to zero.

A.8 Individual strategy and lying task profiles

In this section, we present the distributions of subjects' strategy profiles in the role of interest group (Table 16) and voter (Table 17) in Part I of the experiment. In addition, we provide an overview of the lying task profile distributions for each of the four games in Part I (Table 18).

Table 16: Distribution of interest group strategy profiles

		Majority interest group			Minority interest group		
Game	Measure	Always	Mix	Never	Always	Mix	Never
MD	Truth to Majority voter	0.94	0.06	0.00	0.38	0.46	0.17
	Truth to Minority voter	0.38	0.35	0.27	0.63	0.29	0.08
MU	Truth to Majority voter	0.85	0.11	0.03	0.33	0.29	0.38
	Truth to Minority voter	0.51	0.27	0.22	0.71	0.27	0.02
PD	Truth	0.92	0.07	0.01	0.54	0.27	0.19
PU	Truth	0.71	0.25	0.04	0.38	0.21	0.42

Notes. This table reports the distribution of interest group strategy profiles ("Always", "Mix" or "Never") by interest group type for each of the four games. Following the convention in the literature, starting with Cai and Wang (2006), we classify a subject's strategy in an interest group role as "Always" if 80% or more of the subject's messages are truthful, "Mix" if less than 80% but more than 20% of the messages are truthful, and "Never" if 20% or less of the messages are truthful. We use data from the second half (rounds 21-40) of all Part-I games.

Table 17: Distribution of voter strategy profiles

			Majority voter			Minority voter		
Game	Measure	Interest group	Always	Mix	Never	Always	Mix	Never
MD	Trust	Majority	0.88	0.06	0.06	0.19	0.29	0.52
		Minority	0.17	0.42	0.42	0.83	0.08	0.08
MU	Trust	Undisclosed	0.49	0.38	0.13	0.36	0.37	0.27
PD	Trust	Majority	0.80	0.14	0.06	0.56	0.28	0.16
		Minority	0.23	0.31	0.46	0.73	0.13	0.15
PU	Trust	Undisclosed	0.42	0.44	0.14	0.45	0.31	0.23

Notes. We classify a subject's strategy in a voter role as "Always" if 80% or more of the subject's voting actions are trust, "Mix" if less than 80% but more than 20% of the voting actions are trust, and "Never" if 20% or less of the voting actions are trust. We use data from the second half (rounds 21-40) of all Part-I games.

Table 18: 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.

A.9 Risk elicitation task

In Section A.11.4, 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.11.4). 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.[47] 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 19.

Table 19: 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).

A.10 Regressions

In this section, we present the random effects panel regressions of trust in the MD-game, the social media environment and the traditional media environment. Regressions for our other variables of interest are available upon request.

[47] We only allow subjects to switch from Option A to Option B once.

Table 20: Random effects panel regressions of trust in the MD-game

	Majority interest group	Minority interest group
Type	-0.525***	0.485***
	(0.056)	(0.067)
Intercept	0.875***	0.304***
	(0.030)	(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 21: 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***	-0.002	-0.350***	0.432***
	(0.064)	(0.101)	(0.088)	(0.123)
PD	0.105*	0.445***	-0.449***	0.160
	(0.061)	(0.088)	(0.076)	(0.116)
PU	-0.089	0.141	-0.174**	0.252**
	(0.063)	(0.091)	(0.076)	(0.108)
Part	-0.077	0.076	-0.070	0.121
	(0.045)	(0.069)	(0.049)	(0.080)
Intercept	0.738***	0.411***	0.774***	0.385***
	(0.047)	(0.067)	(0.051)	(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 22: 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**	-0.509***	-0.186**	0.361***
	(0.039)	(0.084)	(0.086)	(0.108)
MU	-0.281***	-0.444***	0.260***	-0.105
	(0.060)	(0.074)	(0.097)	(0.122)
PU	-0.149**	-0.092	0.334***	0.092
	(0.059)	(0.080)	(0.086)	(0.104)
Part I	-0.019	-0.055	-0.003	0.056
	(0.023)	(0.036)	(0.065)	(0.071)
Intercept	0.858***	0.801***	0.415***	0.462***
	(0.031)	(0.044)	(0.064)	(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.11 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.11.1), instructions for Part II (Section A.11.2), instructions for the lying aversion task (Section A.11.3), instructions for the risk elicitation task (Section A.11.4), tutorial questions for Part I (Section A.11.4), tutorial questions for Part II (Section A.11.4) and the questionnaire (Section A.11.4).[48]

The description of the 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.11.1 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.

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.

[48]The set of instructions for the other treatments are available upon request.

(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.

The payoffs of an A1-participant

	Decision of a B-participant (B1 or B2)		
State	X	Y	Z
Heads	30	10	20
Tails	10	30	20

The payoffs of an A2-participant

	Decision of a B-participant (B1 or B2)		
State	X	Y	Z
Heads	10	30	20
Tails	30	10	20

The payoffs of a B1-participant

	Decision of a **B1**-participant		
State	X	Y	Z
Heads	90	27	60
Tails	27	90	60

The payoffs of a B2-participant

	Decision of the **B2**-participant		
State	X	Y	Z
Heads	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

Which state do you think was selected? (The two numbers need to sum up to 100.)	
Chance (in %) that the state is **Heads** [ ]	Chance (in %) that the state is **Tails** [ ]
Which type do you think the A-participant has? (The two numbers need to sum up to 100.)	
Chance (in %) that the A-participant is of **Type 1** [ ]	Chance (in %) that the A-participant is of **Type 2** [ ]

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 1." 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.11.2 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.

The payoffs of an A1-participant

	Decision of a B-participant (B1 or B2)		
State	X	Y	Z
Heads	30	10	20
Tails	10	30	20

The payoffs of an A2-participant

	Decision of a B-participant (B1 or B2)		
State	X	Y	Z
Heads	10	30	20
Tails	30	10	20

The payoffs of a B1-participant

	Decision of a **B1**-participant		
State	X	Y	Z
Heads	90	27	60
Tails	27	90	60

The payoffs of a B2-participant

	Decision of the **B2**-participant		
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.11.3 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.11.4 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.11.5 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?

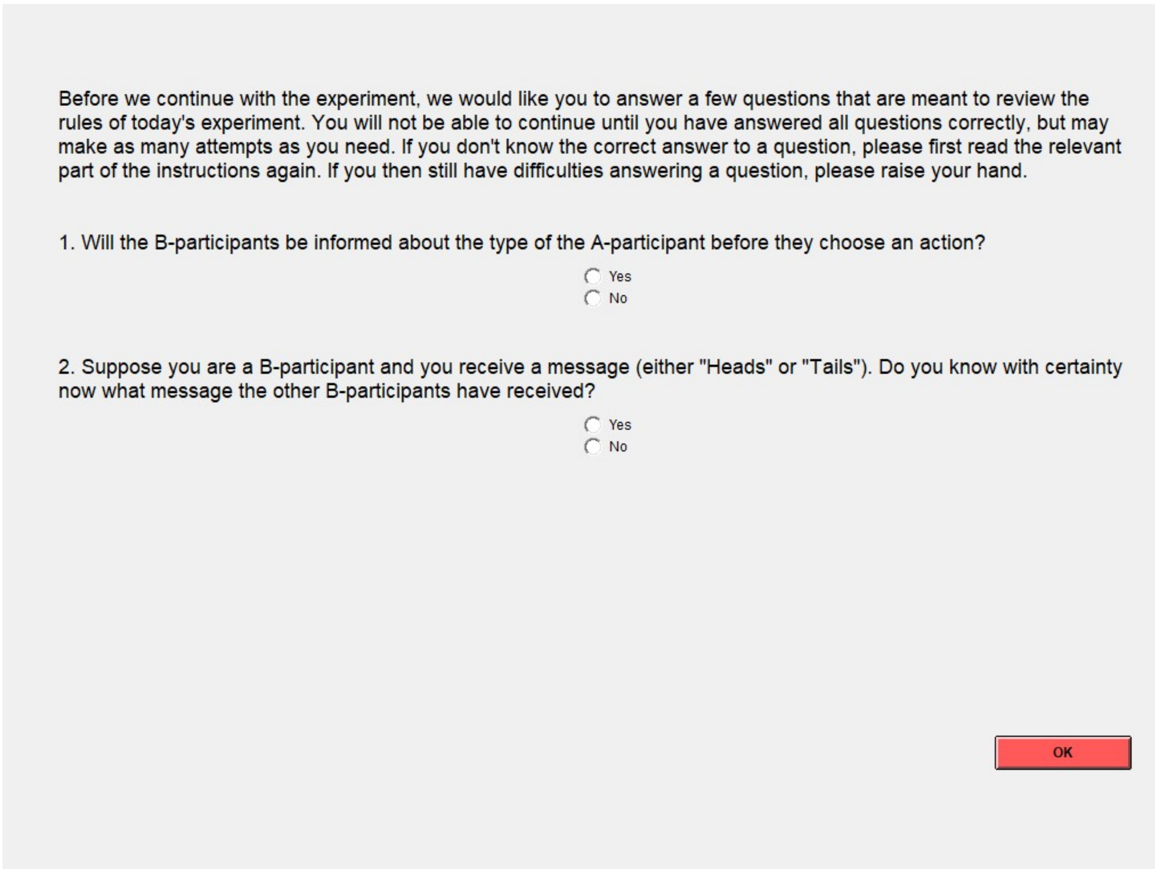
- Yes
- No

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.11.6 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

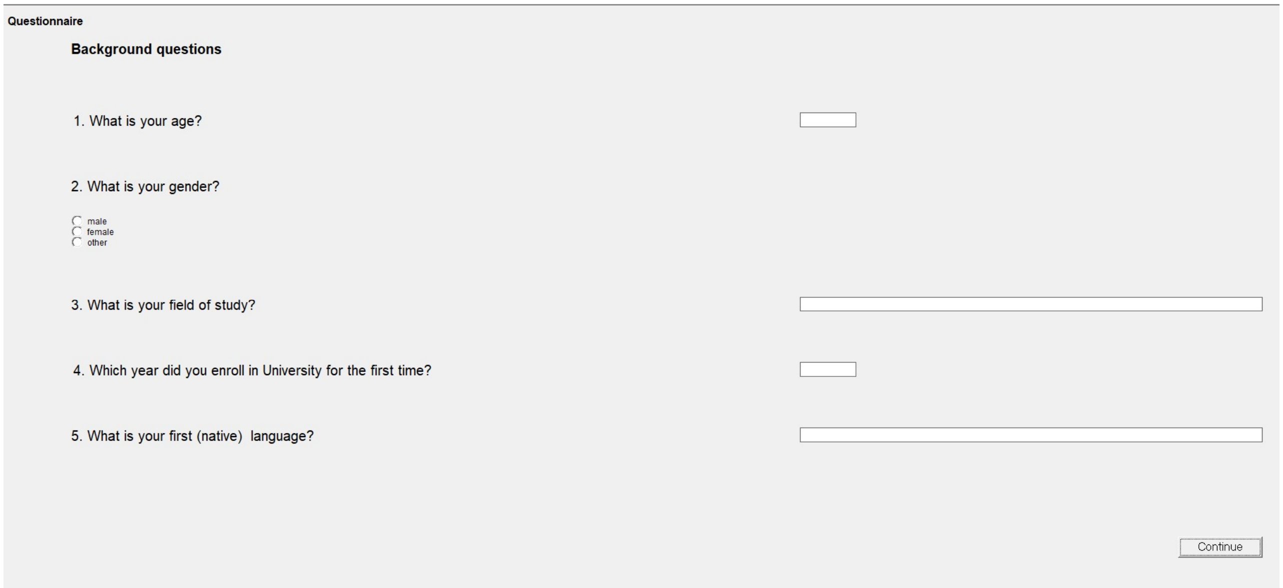
2. Suppose you are a B-participant and you receive a message (either "Heads" or "Tails"). Do you know with certainty now what message the other B-participants have received?

Yes

No

OK

A.11.7 Questionnaire



Questionnaire

Background questions

1. What is your age?

2. What is your gender?

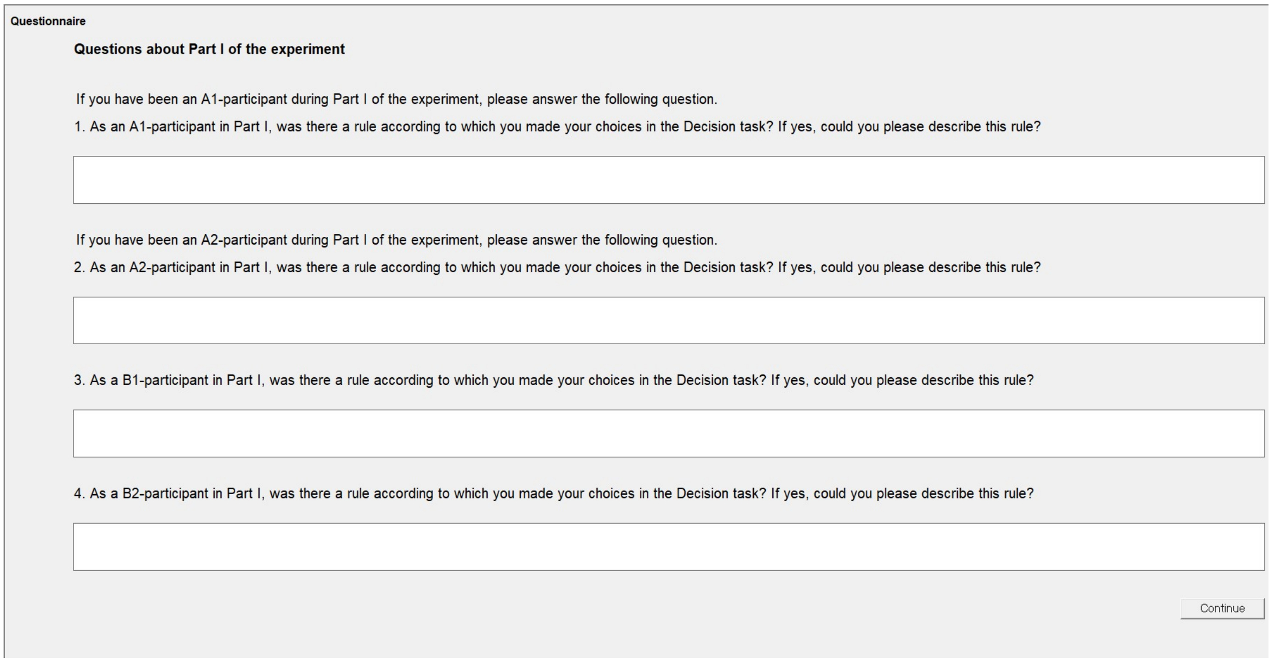
- male
- female
- other

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?

Continue



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