Experimental Economics (2022) 25:680–705
https://doi.org/10.1007/s10683-021-09729-4

ORIGINAL PAPER

Deliberative structures and their impact on voting under economic conflict

Jordi Brandts[1] · **Leonie Gerhards**[2] · **Lydia Mechtenberg**[1]

Received: 29 October 2019 / Revised: 15 August 2021 / Accepted: 18 August 2021 /
Published online: 15 September 2021

Abstract
We conduct a laboratory experiment to investigate how different deliberative structures of varying inclusiveness affect collective decisions in the presence of economic conflict. An electorate consists of two groups, one informed and one uninformed about an uncertain state of the economy. This state affects payoffs differently for the two groups. We study three deliberative structures that vary in how the uninformed are included in pre-vote communication. Compared with a setting without any communication, we find that communication in all three deliberation treatments leads to more frequent votes for the efficient policies. The most inclusive deliberative structure motivates more truthfulness, more trust, more cooperativeness (i.e. refraining from protest votes), and more votes for the efficient policies, than the least inclusive structure. However, comparison among the deliberation treatments reveals that the most inclusive deliberative structure is not the one that generates the highest degree of truthfulness. The dynamics of communication lead to a general deterioration of truth-telling and cooperativeness, reinforced by the use of disrespectful and uncooperative language.

Keywords Communication · Economic conflict · Experiments

JEL Classifications C92 · D9

✉ Jordi Brandts
jordi.brandts@iae.csic.es

1 Institut d'Anàlisi Econòmica (CSIC) and Barcelona GSE, Campus UAB, 08193 Bellaterra (Barcelona), Spain

2 Universität Hamburg, Hamburg, Germany

1 Introduction

Modern societies in the West are under strain. Populist politicians and parties are becoming more successful in campaigning against government policies that tended to appear self-evident in the past decades. Past policies were based on efficiency-enhancing values such as international integration and globalization, individual freedom and meritocracy (Sandel, 2020). Nowadays, a political divide opens up between those social groups that have internalized these values and continue supporting the respective policies and those other social groups that oppose these policies. To better understand these developments, it can be helpful to consider different facets of the social divide: differences in economic opportunities, differences in access to relevant (economic) information, as well as different levels of inclusiveness characterizing public debates. A sound analysis of how these differences interact is essential for a better understanding of some of the current tensions in modern democracies.

To this purpose, we develop an experimental model (Mäki, 2005) that represents in a simplified way the situation of modern societies in the West. Although we do not frame our experiment, its main application is to a situation in which societies have to choose between alternative policies, facing a potential conflict between two social groups. The state of the economy is uncertain, and while the members of one of the groups (the whites) have some information about it, the members of the other group (the blues) are uninformed.[1] In both states of the economy the same set of policies can be implemented, which lead to different distributions of material payoffs between the groups. In one state of the world the two groups have conflicting material interests, in the other state their material interests are aligned. Hence, only the whites have some information on whether interests are conflicting. The collective choice of policy is determined through a vote in which all individuals from both groups can participate. Before the vote, the groups may communicate.

This simple setting captures modern democratic societies' critical issue of dealing with potential economic conflicts between social groups. In most of these societies there is arguably a social class with little wealth, whose incomes are highly dependent on the thriving of the national industrial sector (the "dependent class"). In contrast, leading politicians and opinion leaders in the media, cultural and educational professions are mainly recruited from another social class that is less dependent on the national industries and more entrenched in the multi-national economic sectors (the "entrenched class"). Individuals in the entrenched class have higher education and somewhat more reliable information on economic conditions than those in the dependent class. Hence, they have an informed idea about whether their material interests align or clash with those of the dependent class, depending on the current state of the economy. By contrast, the dependent class only knows that their interests may or may not clash with the entrenched class' interests, depending on the unknown state. How much can such a society turn their collective choice toward efficient policies? And how do truthfulness, trust, and cooperativeness between

[1] We ran the experiment in German, where the colors white and blue do not connote "white-collar" or "blue-collar" classes (or any other classes), unlike in English.

these classes develop, depending on the degree of inclusiveness that the communication channels in the society provide, e.g., the degree to which discussion panels and media are representative of both classes?

Across experimental treatments, we exogenously vary inclusiveness as follows. Pre-vote communication is determined by three distinct protocols or *deliberative structures* that differ in how the communication between the whites and the blues is designed. In the *FullyPublic* deliberative structure, we implement a stylized ideal of deliberation that is inspired by the normative deliberation literature (Dawes et al., 1990, Orbell et al., 1988, Dryzek & List, 2003): All individuals, regardless of their group affiliation, have the same possibilities of sending to and receiving messages from all other individuals. This ideal deliberative structure is the most inclusive.[2] We then decrease inclusiveness stepwise: In the deliberative structure that we refer to as *TopDown*, only the whites have the possibility of sending messages; the blues are reduced to being receivers. However, at least in their role as receivers, whites and blues are on the same footing since each message sent is simultaneously received by all. We further reduce inclusiveness in the deliberative structure that we refer to as *TopDownClosed*. There, the whites first communicate among each other ("behind closed doors") before sending messages to all.[3] For completeness, we compare these three deliberative structures to the benchmark voting game without any communication (*NoChat*).

Our main interest is in comparing the three implemented deliberative structures with respect to how they contribute to solving a potential economic conflict. For this to happen, in the experimental model at hand, the whites need to truthfully reveal their information to which the blues need to respond with a cooperative vote for the efficient policy, even if this is against their material interest. Building on the deliberation literature cited above, which posits that inclusiveness is essential for truthfulness and a sense of cooperation in society, we propose that different deliberative structures trigger distinct preferences in the spirit of Bowles and Polanía-Reyes (2012). Our assumption is that a deliberative structure's inclusiveness triggers efficiency preferences, translating into increased truthfulness and cooperativeness.[4] In a nutshell, we hypothesize that truthfulness and cooperativeness will be lowest in the setting without communication and increase with increasing inclusiveness of the deliberative structures.

[2] Researchers in political science have devoted much attention to issues of deliberation, see in particular Cohen (1989), Gutmann and Thompson (1996), Habermas (2015) and Landwehr (2010). Myers and Mendelberg (2013) give an overview of work on political deliberation and Karpowitz and Mendelberg (2011) survey the experimental literature in political science on the topic.

[3] Separate communication between groups has been hypothesized to lead to polarization in opinions (Sunstein, 2009; Benoît & Juan, 2016).

[4] In Bowles and Polanía-Reyes (2012), actions are considered to be motivated by a heterogeneous repertoire of preferences whose salience depends on the nature of the decision situation. The general idea is that preferences often depend on some specific features surrounding the act of choice which are salient to the decision-makers involved. Bowles and Polanía-Reyes (2012) focus on how the presence of monetary incentives triggers different preferences. In an industrial organization setting, Apffelstaedt and Mechtenberg (2020) analyze context-dependent consumer preferences in a competitive market. In our case we propose that different deliberative structures affect players' preferences.

Compared with the setting without any communication, we find that communication in all three deliberation treatments leads to more frequent choices of the efficient policies. Comparing our three deliberative structures with each other, we find that the blues are more trustful and more cooperative in *FullyPublic* than in *TopDownClosed*, which is reflected in their higher propensity to vote for the efficient policies in the former than in the latter treatment. Hence, when comparing those two extreme degrees of inclusiveness, we do find that more inclusiveness contributes to solving the economic conflict between the two groups. However, this is less clear when we compare *FullyPublic* and *TopDown* as well as *TopDown* and *TopDownClosed*. And although the whites are more truthful in *TopDown* than in *TopDownClosed*, their truthfulness does not differ significantly between *FullyPublic* and *TopDownClosed*. We then study the dynamics of the chat and voting behavior in the three deliberation treatments. We find that the dynamics of communication and behavior exhibit a general deterioration of blues' cooperativeness and trust and whites' truthfulness in all deliberation treatments. Moreover, in *FullyPublic* there is an additional factor that enforces deterioration, namely disrespectful language of the whites and uncooperative talk of the blues.

2 Related literature

Pre-vote communication has already been studied in the extensive literature on voting, recently surveyed in Palfrey (2016). For instance, the results in Guarnaschelli et al. (2000) and Goeree and Yariv (2011) document that pre-play communication in the form of either a straw-vote or unrestricted chat leads to an increase in the efficiency of the voting outcome. By contrast, Buechel and Mechtenberg (2019) show that pre-vote communication in social networks that is restricted to information aggregation can lower efficiency even in a common-interest setting.

Andreoni and Rao (2011) study how pre-distribution communication affects distribution. They find that when pairs of individuals play a dictator game the distribution of the endowment varies depending on whether the allocator can send a message to the receiver or the receiver can send a message to the allocator. The receiver obtains a smaller share in the first case than in the second.

Moreover, previous experimental work has found evidence in favor of communication affecting group identity (see Akerlof & Kranton, 2000, 2010), and hence preferences. Chen and Li (2009) report on an experiment in which they study the effects of induced group identity in an environment with an ingroup and an outgroup. They find that participants are more altruistic towards members of an ingroup and that chat communication within the ingroup leads to stronger ingroup favoritism. In the related experiment of Chen and Chen (2011) participants play a coordination game with either an ingroup or an outgroup. In one of the treatments the coordination game is preceded by a chat. They find that stronger communication—more words, more content—has a positive effect on the ingroup and a negative effect on the outgroup. Robalo et al. (2017) also induce ingroup bias in an experiment related to political issues without using communication. They group people according to the results of a personality questionnaire and find that political participation is higher

Table 1 Blue and white players' payoffs, conditional on the state of the world and implemented policies

Policy	State X		Policy	State Y	
	Whites	Blues		Whites	Blues
A	20	20	A	10	0
B	0	0	B	20	10
C	0	10	C	0	20

when ingroup bias is stronger. In our case, groups are distinguished by asymmetric payoffs and access to information.

Like in our study, Palfrey and Pogorelskiy (2017) investigate the effects of two different communication structures on voting: public communication (all voters exchange messages through a computer chat) and party communication (messages are only exchanged within each party). However, they focus on voter turnout in an experiment with costly voting. The issue of voter turnout is quite different from the research question that we address. In our environment voting is costless and, hence, is it not suited to studying turnout; indeed, we observe very little abstention in all our treatments.

Pronin and Woon (2017) study how the economic benefits of deliberation can be robust to the existence of private communication between parts of the society, prior to a public discussion. In a setting in which a group of players has to allocate a fixed budget between themselves and a public good they find that allowing for private messages before the public discussion leads to the under-provision of the public good. Again, the particular issue they study is very different from ours, but the communication structure they study is related to our *TopDownClosed* treatment.

Although our main interest is in communication on the societal level, our analysis can also be related to the effects of *institutionalized communication structures* in organizational economics (Ambrus et al., 2013). For example, Brandts and Cooper (2007) compare the effects on coordination of various communication structures between a manager and workers.

Our novel contribution to the literature reviewed above is that we simultaneously study (1) how two groups solve a state-dependent conflict of interest, (2) how efficiently they aggregate information on that state held by one of the groups, and (3) how the conflict solution is affected by communication structures.

3 Experimental design

We consider the following voting game: Six players form a voting group, consisting of three white players and three blue players. These players vote on a policy from a set of three alternatives (A, B, and C). The implemented policy determines state-dependent payoffs that may differ by the players' colors, see Table 1. At the beginning of the game, nature draws the state of the world, which is either X or Y with equal probability. Then, nature randomly draws an informative private signal on the state of the world for each white player. These signals are conditionally independent and true with probability $p = 0.7$. Blue players do not receive any signal.

Next, a communication stage starts. We consider three deliberative structures that vary in their inclusiveness. In particular, they gradually differ in the extent of the whites' control over the communication process. In treatment *FullyPublic*, whites and blues can publicly communicate with each other. In *TopDown* the whites, but not the blues can send (public) messages and in *TopDownClosed* the whites can first communicate with each other unobserved by the blues and then send public messages that are also received by the blues. Messages are sent simultaneously, and remaining silent is possible for all senders.

Communication is implemented as computerized free-form chat.[5] In *FullyPublic* and *TopDown*, the chat lasted for two minutes. In *TopDownClosed*, both the first (private) chat among the whites and the second (public) chat lasted for one minute each.[6] Benchmark treatment *NoChat* implements our game without any communication stage. After the information stage, subjects in *NoChat* have the opportunity to take private notes in a computer window that looks exactly like the chat window in the other treatments. We decided to exogenously restrict the duration of the chat (note taking) stage in order to keep the total duration of the experimental sessions comparable within and across treatments. We thus tightly control the task- and time-structure of all treatments. We asked our subjects to focus their communication (in *NoChat* their notes) on the voting decision at hand. Apart from that, we did not impose any restrictions on their writing.[7]

In the following voting stage, each individual chooses whether to vote for one of the three policies *A*, *B*, and *C*, or to abstain. Voting is costless. The final policy is elected according to the plurality rule (i.e., the final policy is the one that got most votes); and ties are resolved randomly, with equal probabilities.

The state of the world interacts with the implemented policy in generating final payoffs, as displayed in Table 1. In state *X*, whites and blues would agree on the most preferred policy: Both would like to implement policy *A*. This is, however, not true in state *Y*: While the whites would prefer *B* to be chosen, selfish blues would prefer *C* instead. Hence, the two groups have a state-dependent conflict. This

[5] Deliberative democracy literature typically considers that communication involves reason giving and persuasion. For this reason we considered free-form communication more appropriate than structured communication. Also, the survey by Brandts et al. (2019) documents that free-form has a different effect than structured communication. As an example Charness and Dufwenberg (2006, 2010) compare the effects of a full written page of free-form communication with that of a protocol in which agents could only choose between sending a pre-formulated promise and not sending any message. They find that 'bare promises' have substantially smaller effects than richer communication.

[6] From the post-experimental feedback that we received from the subjects and the analysis of the chat contents, we are confident that our time constraint on the chat is not binding. Moreover, in a comparable experimental setup, Goeree and Yariv (2011) observe that an unconstrained pre-vote chat between privately informed voters lasted only for 26 ± 11 seconds on average. We hence conjecture that a chat duration of two minutes gives our subjects sufficient time to share the whites' information (or lies) as well as to deliberate on the policy to be chosen.

[7] Translated instructions to all treatments are included in Appendix D of the Electronic Supplementary Material.

conflict in state Y is particularly sharp since, for selfish whites, C is the worst of all options. The state-dependent efficient policy choice is A in state X and B in state Y. It is hence in line with the preferences of the whites.[8]

The state-dependent conflict gives the white players an incentive to lie about the state of the world in the communication stage, if, given their signals, they expect state Y. In this case, truthfully reporting the majority signal (i.e., the signal received by the majority of whites) would lead selfish blue players to vote for C. The whites would vote for B, which ultimately generates a tie between policies B and C yielding each white player an expected payoff of $\frac{1}{2} \times 20 + \frac{1}{2} \times 0 = 10$ if state Y prevails. If, however, the whites successfully lied about the state of the world such that the blues expected state X and hence voted for A, the whites would expect to earn $\frac{1}{2} \times 10 + \frac{1}{2} \times 20 = 15$ if they themselves chose policy B. Obviously, and as shown in Appendix A in the Electronic Supplementary Material, successful lies cannot be part of an equilibrium here – instead, communication would become meaningless ("babbling").

Our main underlying assumption is that deliberation works in a sense consistent with normative deliberation theory (e.g., Dawes et al., 1990, Orbell et al., 1988, Dryzek & List, 2003). Central to this theory is the idea that the more inclusive the deliberation protocol, the more inclusive become the preferences of those participating in it.[9] Interpreting deliberation theory, we predict that a player has selfish preferences, unless the structure of the deliberative process leads him or her to internalize the interests of others. For the blues, preferences only impact their voting. For the whites they affect both their voting and the revelation of the information they have. Based on this, we can characterize the white and blue players' optimal behavior in our four treatments.

First, in the benchmark treatment without deliberation, *NoChat*, players from both groups have selfish preferences. The blues vote for C, which is the policy that benefits them most in expectation if information about the true state is absent. The whites vote for A or B.[10]

Consider next the deliberation treatments, starting with the one with the lowest degree of inclusiveness, *TopDownClosed*, which gives the blues a passive role in the communication process and allows the whites to communicate exclusively with the other whites in the first (private) chat stage, like in the party communication treatment of Palfrey and Pogorelskiy (2017). Supposedly, the exclusion of the blues and the in-group bias of the whites triggered by the private chat lead both groups to have

[8] Given the chosen payoffs, state X can be considered the good state, Y the bad state: On the one hand, the efficient policy in X, policy A, yields a larger total payoff than the efficient policy B in Y ($3 \times 20 + 3 \times 20 = 120$ vs. $3 \times 20 + 3 \times 10 = 90$); on the other hand, the efficient policy in X leads to a fair allocation of payoffs (both white and blue players earn 20), while the efficient policy in Y generates a payoff inequity (20 for white players, 10 for blue players).

[9] Put differently, the more heterogeneous the people a person speaks to before making a decision, the more heterogeneous are the interests that this person will take into account when finally reaching a decision.

[10] The whites' votes depend on their private signal or on a coordination strategy. For equilibrium selection, see our Theoretical Appendix A in the Electronic Supplementary Material.

selfish preferences. Then the interaction between them involves the state-dependent conflict described above, and the whites will lie to the blues if they receive majority signal *Y* (i.e. only babbling equilibria exist). Therefore, the blues will always vote for *C*, which maximizes their expected payoffs. The whites, on the other hand, will vote according to their shared information, i.e., for *A* in case the majority signal is *X*, and for *B* in case the majority signal is *Y*.[11] Hence, voting behavior in *TopDownClosed* can be outcome-equivalent to what we predict for *NoChat*.

Consider next *TopDown*. In this treatment's deliberation protocol, the blues have a passive role in the communication process as in *TopDownClosed* and, hence, only care for their own interests. However, the whites now address the society as a whole in public, without their in-group bias being triggered by the private chat as in *TopDownClosed*. Thus, the hypothesis based on deliberation theory is that the whites become empathetic with the blues and hence develop efficiency preferences. This makes them truthful toward both colors. As a consequence, the conflict described above is, in total, ameliorated. The whites do no longer lie to the blues about the signals they have received. Hence, in equilibrium the blues now obtain information about the true state. Following their material interests, they vote for *A* along with the whites when they are told that the state is *X*, but for *C* (rather than *B* along with the whites) when they are told that the state is *Y*.

Finally, consider *FullyPublic*. Here, both whites and blues are senders as well as receivers, as required by normative deliberation theory for an ideal deliberative structure. Our main underlying assumption implies that now both whites and blues care for both colors; i.e., they all have efficiency preferences. Therefore, the equilibrium strategies of the two colors are the following: The whites truthfully report their signals to all other players, and players vote in such a way that a plurality of votes is for *A* if the majority of signals indicate that the state is *X* and for *B* if the majority of signals indicate that the state is *Y*. Such strategy profiles implement a compromise: If state *Y* is more likely than *X*, the blues refrain from voting for their best choice *C* and support their second-best choice *B* instead, which is efficient. Given this behavior of the blues, not even selfish whites would have an incentive to lie to them about the state.

Table 2 summarizes for each of the four treatments (1) if players may read and/or write messages and (2) how the different treatments affect their efficiency preferences, (3) the incentive of the whites to lie to the blues and (4) the predicted voting decisions. In the analysis of the results we will mainly focus on the comparative statics of behavior (see Schotter, 2015), given by the last five columns of Table 2.

We conducted 20 sessions with 468 subjects from various study backgrounds at the WISO-laboratory of Hamburg University.[12] Subjects kept their roles throughout the 20 periods of the experiment. Half of them randomly assumed the roles of white, the other half the roles of blue players. Groups were randomly re-composed

[11] Note that in the main text, we slightly abuse notation and refer to states, signals, and messages by the same (capital) letters, *X* and *Y*.

[12] In *NoChat* and *FullyPublic* we ran five sessions, each of them comprising 24 subjects. In both *TopDown* and *TopDownClosed* we ran four sessions with 24 subjects and one session with 18 subjects.

Table 2 Preferences for efficiency and their impact on information aggregation and voting decisions in equilibrium

	Whites		Blues		Efficiency preferences of		Whites' incentive to lie	Voting given majority signal			
								X		Y	
	Write	Read	Write	Read	Whites	Blues		Whites	Blues	Whites	Blues
NoChat	–	–	–	–	No	No	–	*A* or *B*	*C*	*A* or *B*	*C*
FullyPublic	✓	✓	✓	✓	✓	✓	No	*A*	*A*	*B*	*B*
TopDown	✓	✓	–	✓	✓	No	No	*A*	*A*	*B*	*C*
TopDownClosed	✓/✓*	✓/✓*	–/–	–/✓	No	No	✓	*A*	*C*	*B*	*C*

*The first entry refers to the private chat among the whites, the second entry relates to the subsequent public chat. In columns 5 – 7, ✓ indicates for each of the treatments if in equilibrium, (i) the whites and blues have efficiency preferences and (ii) if the whites have an incentive to lie to the blues and subjects' chat IDs were randomly reassigned in the beginning of every period such that subjects were not able to track individuals throughout the different periods (stranger matching).

We used hroot (Bock et al., 2014) to recruit subjects and coded the experiment in z-tree (Fischbacher, 2007). During the sessions payments were expressed in experimental currency units (points) which were exchanged to Euros at a rate of 1 Euro = 3 points at the end of the experiment. Average earnings for the 120 minutes sessions amounted to 23.28 Euro (s.d. 4.73), including a 10 Euro show-up fee (minimum earnings = 10 Euro, maximum earnings = 30 Euro).

4 Results

4.1 Whites' truth-telling and blues' trustfulness

We first investigate how the inclusiveness of the deliberative structure affects truth-telling – or conversely, lying – of the whites. We define lying as white players reporting majority message *X*, that is, at least two of the three whites report message *X*, if, in fact, the majority signal was *Y*. That means that we consider only pivotal lies. Analogously, we define truth-telling as the white players reporting majority message *Y* if their majority signal was *Y*. Since the whites have no incentive to lie if they receive majority signal *X*, we restrict the analysis of lying/truth-telling to periods in which the whites receive majority signal *Y*. Comparing across treatments, 22.16% of the white subgroups lie in *FullyPublic*, 11.44% of the white subgroups lie in *TopDown* and 35.08% in *TopDownClosed*.[13]

To econometrically test for the treatment differences in truth-telling, we run a linear probability model, in which we regress the dummy variable for a white subgroup reporting majority message *Y* if in fact the majority signal is *Y* (truth-telling) on treatment dummies *FullyPublic* and *TopDownClosed*. We hence treat *TopDown* as baseline treatment which differs from each of the other two deliberation treatments by only one design feature.[14] Additionally, we control for the period of play.

Other then expected, whites lying does not (weakly) increase with decreasing inclusiveness across the three deliberation treatments. Although the respective treatment coefficient in Table 3, specification (1), reveals that the whites are less truthful in *TopDownClosed* than in *TopDown*, truth-telling is significantly less prevalent in *FullyPublic* than in *TopDown*, too. Also, truth-telling does not differ significantly between *FullyPublic* and *TopDownClosed*, see the respective Wald test result reported in the lower part of the table. We summarize our findings as follows:

13 In addition, we observe 12.04% of silent white subgroups in the public chat of *TopDownClosed* (none in the other two deliberation treatments).

14 The blues may talk in the chat in *FullyPublic*, but not in *TopDown*. The whites have the opportunity to chat privately in *TopDwonClosed*, but not in *TopDown*.

Table 3 Truth-telling if the majority signal is *Y*

	Majority message: Y	
	(1)	(2)
FullyPublic (FP)	−0.111**	−0.109*
	(0.042)	(0.092)
TopDownClosed (TDC)	−0.277**	−0.316**
	(0.012)	(0.010)
Period	−0.012***	−0.013***
	(0.000)	(0.000)
FullyPublic × Period		−0.000
		(0.957)
TopDownClosed × Period		0.004
		(0.585)
Constant	1.008***	1.017***
	(0.000)	(0.000)
Wald test result for comparison of treatment coefficients (p-value):		
FP vs. TDC	0.110	
R^2	0.099	0.099
Number of clusters	15	15
Observations	554	554

Linear probability models. Dependent variable: Reported majority message: Y. Robust standard errors are clustered at the session level and p-values are given in parentheses: *p < 0.10, **p < 0.05, ***p < 0.01. TopDown serves as baseline treatment

Finding 1 (Whites' truth-telling) Increasing inclusiveness of deliberation does not lead to more truthfulness across the three deliberation treatments. The whites are more truthful in *TopDown* than in *FullyPublic* and *TopDownClosed*.

Moreover, the significant *Period* coefficient in specification (1) of Table 3 shows that the whites frequently report majority signal *Y* truthfully in the first periods, but lie increasingly more often in later periods. In the extended specification (2) we add interaction terms between treatments and period. There is no evidence that the period of play has a significant effect on the observed treatment differences.

Next we turn to the question of how the inclusiveness of the deliberative structure affects the blue players' trustfulness; the analysis is presented in specifications (3) and (4) in Table 4. We quantify trustfulness as the blues' propensity to vote for the efficient policy *A* after having received majority message *X*. In a linear probability model we regress the individual blues' votes for the respective expected efficient policy *A* on treatment dummies *FullyPublic* and *TopDownClosed*, using *TopDown* as baseline treatment, and control for the period of play. We restrict the sample of analysis to periods in which the whites report majority message *X*. Table 4, specification (3), reveals that the blues are significantly more trustful in the most inclusive treatment *FullyPublic* than in the least inclusive treatment *TopDownClosed*, see the

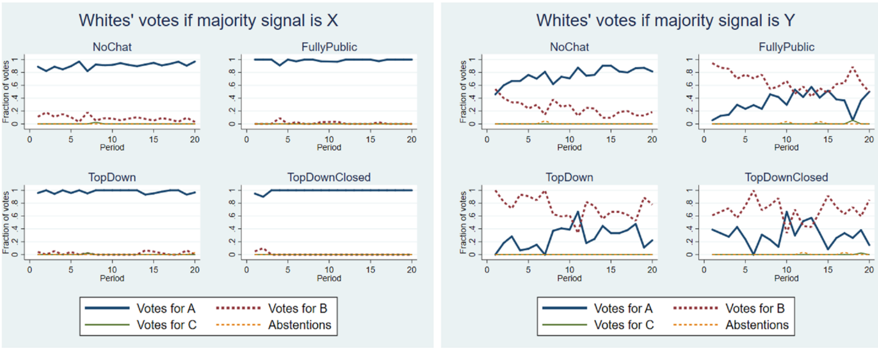

Fig. 1 Whites' individual voting decisions

reported Wald test result in the lower part of the table. There is neither a significant difference in the blues' trustfulness between *FullyPublic* and *TopDown* nor between *TopDown* and *TopDownClosed*.

Finding 2 (Blues' trustfulness) The blues' trustfulness (weakly) decreases with decreasing inclusiveness of the deliberative structure, though not all differences turn out significant. Blue players are more trustful in *FullyPublic* than in *TopDownClosed*.

Considering the two extremes of the implemented deliberative structures, treatments *FullyPublic* and *TopDownClosed*, the blue players' greater beliefs in the white players' truth-telling in the former compared to the latter treatment seem justified to some extent. The actual difference in truth-telling between these treatments is insignificant, though ($p = 0.110$). Interestingly, however, the blues seem unaware of the whites' greater truthfulness in *TopDown* compared to *FullyPublic*.

Finally, the significantly negative *Period* coefficient in specification (3) of Table 4 suggests that the blues' trustfulness in the whites' reported majority message X decreases significantly by on average 2.3 percentage points per period, independently of treatment. Adding interaction terms between treatments and period of play in specification (4) does not reveal significant changes in treatment differences over periods. Except for the significant treatment difference between *FullyPublic* and *TopDownClosed*, the blues mistrust the whites' reported majority signal to similar extents in all three deliberation treatments and in all periods of play.

4.2 Whites' voting decisions

We now investigate how voting behavior of the whites varies with the inclusiveness of the deliberative structure. Note that we do not expect any variation here: At the voting stage, the material interests of the informed whites overlap with efficiency preferences. Hence, even if whites' efficiency preferences were triggered by inclusive deliberation, we would not able to observe this in the whites' votes.

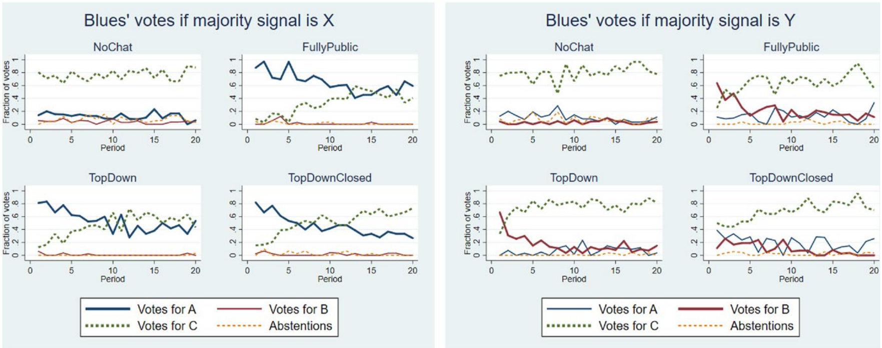

Fig. 2 Blues' individual voting decisions

Figure 1 displays the white players' voting decisions over periods. If the majority signal is *X*, whites predominantly vote for policy *A* in all four treatments, see the four graphs on the left. If the majority signal is *Y*, the majority of whites vote for policy *B* in all three deliberation treatments, in particular in the first periods and for *A* in *NoChat*, see the four graphs on the right.

The linear probability models in Table 7, specifications (1) and (5), complement this graphical presentation and confirm that the varying inclusiveness does not significantly change the whites' voting behavior across the deliberation treatments (see Appendix).[15]

Finding 3 (Whites' voting decisions) Given majority signal *X* (*Y*), the whites' propensity to vote for the respective efficient policy *A* (*B*) does not differ across *FullyPublic*, *TopDown* and *TopDownClosed*. However, implementation of the efficient policy is more likely in those deliberation treatments than in *NoChat*.

4.3 Blues' voting decisions

4.3.1 Blues' votes for the efficient policy

We now move on to investigating how the inclusiveness of the deliberative structure affects the voting behavior of the blues. Note that here, we do expect effects, namely that, with increasing inclusiveness, the blues vote weakly more often for the state-dependent efficient policy (and weakly less often for *C*).

[15] Independently of treatment, the whites' propensity to vote for the efficient policy is not too strongly affected by the period of play. Specifications (2) and (6) reveal that, with the exception of the treatment difference between *TopDownClosed* and *NoChat* in case of majority signal *Y*, the treatment differences are virtually unaffected by periods of play (even if interaction terms turn out significant, their coefficients are negligibly small). Conditioning on sent majority messages, instead, we again do not find noteworthy treatment differences of period effects in the whites' propensity to vote for the efficient policies, see specifications (3) and (4) and (7) and (8), respectively.

Table 4 Blues' votes for the efficient policies

	Majority signal: X		Majority message: X		Majority signal: Y		Majority message: Y	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	A vote	A vote	A vote	A vote	B vote	B vote	B vote	B vote
FullyPublic (FP)	0.529***	0.703***	0.096	0.095	0.190***	0.390***	0.122**	0.193
	(0.000)	(0.000)	(0.151)	(0.274)	(0.000)	(0.001)	(0.021)	(0.111)
TopDown (TD)	0.426***	0.601***			0.113***	0.254***		
	(0.000)	(0.000)			(0.000)	(0.000)		
TopDownClosed (TDC)	0.347***	0.571***	−0.068	−0.048	0.058**	0.192***	0.000	0.101
	(0.000)	(0.000)	(0.198)	(0.606)	(0.013)	(0.001)	(0.992)	(0.190)
Period	−0.017***	−0.003	−0.023***	−0.022***	−0.011***	0.000	−0.016***	−0.011***
	(0.000)	(0.205)	(0.000)	(0.000)	(0.001)	(0.886)	(0.001)	(0.000)
FP × Period		−0.017**		0.000		−0.019**		−0.007
		(0.017)		(0.995)		(0.025)		(0.420)
TD ×Period		−0.017***				−0.013***		
		(0.000)				(0.000)		
TDC ×Period		−0.022***		−0.002		−0.012***		−0.010
		(0.002)		(0.729)		(0.002)		(0.121)
Constant	0.297***	0.161***	0.776***	0.769***	0.153***	0.028**	0.331***	0.283***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.013)	(0.000)	(0.000)
Wald test results for comparison of treatment coefficients (p-values)								
FP vs. TDC	0.001		0.008		0.010		0.032	
FP vs. TD	0.097				0.093			
TD vs. TDC	0.128				0.014			
R^2	0.206	0.217	0.086	0.086	0.082	0.094	0.072	0.076
Number of clusters	20	20	15	15	20	20	15	15
Observations	2451	2451	2112	2112	2229	2229	1278	1278

Linear probability models. Dependent variable: Decision to vote for the respective policy. Robust standard errors are clustered at the session level and p-values are given in parentheses: *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$. NoChat serves as baseline treatment in regressions (1), (2), (5) and (6). TopDown is the baseline treatment in the remaining regressions

"FP" refers to FullyPublic; "TD" refers to TopDown,and "TDC" refers to TopDownClosed

Figure 2 displays the blue players' voting decisions over periods for the cases of majority signal *X* (see the four graphs on the left) and majority signal *Y* (see the four graphs on the right). As evident, the blues predominantly vote for policy *C* in treatment *NoChat*. In all other treatments, a non-negligible fraction of blues votes for the respective efficient policy at least in the beginning of the experiment. This is particularly true if the majority signal is *X*. However, the choice of *C* increases over time. For majority signal *X*, the choice of *C* becomes as frequent as that of *A* in *FullyPublic* and *TopDown* and more frequent than *A* in *TopDownClosed*. For majority signal *Y*, *C* quickly becomes the most frequent choice in all treatments.

In the linear probability models presented in Table 4 we study the blue players' voting decisions in more detail. Given majority signal *X* (see specification (1)), the blues vote significantly more often for the efficient policy *A* in *FullyPublic* than in *TopDown*, *TopDownClosed* and *NoChat*, and more often in *TopDown* than in *NoChat*, see the treatment coefficients and corresponding Wald test results which are at least significant at the 10% level. However, the blues' propensity to vote for *A* is not significantly higher in *TopDown* than in *TopDownClosed*. Focusing on those periods in which the majority signal is *Y* (see specification (5)), the blues' propensity to vote for the efficient policy *B* is significantly higher in *FullyPublic* than in *TopDown*, *TopDownClosed* and *NoChat* (at least at the 10% significance level). This specification also reveals significant treatment differences between *TopDown*, *TopDownClosed* and *NoChat*. In sum, increasing inclusiveness of the deliberative structure does indeed direct the blues' votes more toward efficiency.

Finding 4 (Blues' voting decisions: efficiency) Given both majority signals, the blues' propensity to vote for the respective efficient policy is higher in all deliberation treatments than in *NoChat* and their propensity to vote for the efficient policy is highest in *FullyPublic* compared to the other deliberation treatments. Given majority signal Y, the respective difference between *TopDown* and *TopDownClosed* is significant, too.

At the same time, Fig. 2 as well as the significantly negative *Period* coefficients in regression specifications (1) and (5) in Table 4 reveal a general, treatment-independent decline in the blues' propensity to vote for the efficient policy over periods, for both majority signals *X* and *Y*. It amounts to on average 1.7 percentage points per period in case of majority signal *X* and 1.1 percentage points per period in case of majority signal *Y*. Furthermore, the extended specifications (2) and (6) reveal that all treatment differences between *NoChat* and the deliberation treatments are particularly large in the beginning of the session and then decrease significantly over the 20 periods of play, given both majority signals.

4.3.2 Blues' cooperativeness

We define blues' cooperativeness as voting for *B* rather than *C* after receiving majority message *Y*. Given the material incentives, the whites would prefer the blues to believe that the state is *X* when it is *Y*. Hence, the whites sending majority message

Y is believable and the blues then voting *B* is consistent with them rewarding whites' truthfulness. Considering specifications (7) and (8) of Table 4, blues' response to the whites sending majority message *Y* leads to significantly more *B* votes in *FullyPublic* than in the other two deliberation treatments, with no difference between the latter two.

Moreover, we observe a general, but small negative period effect on voting for the efficient policy.[16] Overall, we conclude that the blues' cooperativeness improves with full inclusiveness.

4.4 The dynamics reveal deterioration of deliberation

Reconsidering our previous analyses, the significant *Period* coefficients in the regressions from Tables 3 and 4 show that the incidences of conflict, manifested by lying, low degrees of trust and not voting for the state-specific efficient policies, increase in all treatments over periods. After an initial phase of high cooperation and low conflict, the opportunity to deliberate does not lead to *sustainable* coordination on the efficient outcomes in the deliberation treatments. In particular, the results for *FullyPublic* suggest that the effectiveness of deliberative democracy deteriorates over periods.

To better understand the changes in behavior over periods, we study the interactions between blues' voting decisions, whites' lying behavior and the content of the chat conversations. We hired two research assistants to code the chat messages independently from each other (we refer to them as Coder #1 and Coder #2). We provided them with a list of chat dimensions that we considered potentially relevant a priori to classify the individual statements made by whites and blues.[17] In a second step, we then aggregated these individual classifications to chat message indicator variables at the color group level.

In the regressions presented in Table 5 we analyze the interactions between blues' voting decisions in the deliberation treatments. In our analysis, we focus on chat classifications that we observed in more than 15% of all chat messages. Consider first specification (1), in which we focus on those periods in which the whites report majority message *X*. These are the periods in which the whites either truthfully reveal their majority signal or lie to make the blues believe that situation *X* prevails. Like in Sect. 4.1 above, we take the blues' propensity to vote for the efficient

[16] Further empirical analyses on voting outcomes and efficiency are presented in Appendix B in the Electronic Supplementary Material.

[17] This list includes classifications (indicator variables) of whether a white lies or tells the truth about their individual signal. Moreover, coders classified voting recommendations, noted if someone stresses the public spirit, suspects lying and recorded who is addressed in a statement (own or other color group, all), the tone of the message (respectful, disrespectful, neutral), and whether fairness and/or efficiency was mentioned. A full list of the dimensions in which the chat messages were coded can be found in Appendix C of the Electronic Supplementary Material. Note that in the regressions presented in the main part of this paper, we rely on the work done by Coder #1. With the exception of one coefficient, all significant results presented in Table 5 and Table 6 are similarly found when relying on the codings of Coder #2 instead, see Tables 8 and 9 in the Appendix in the back of the paper.

Table 5 Deliberation treatments: Blues' votes for the efficient policies

	Majority message: X	Majority message: Y
	(1)	(2)
	A vote	B vote
FullyPublic (FP)	0.120*	0.142**
	(0.076)	(0.019)
TopDownClosed (TDC)	−0.070	−0.007
	(0.169)	(0.779)
Potential lie in previous period	−0.052**	−0.016
	(0.044)	(0.642)
Respectful whites	0.017	0.083**
	(0.575)	(0.035)
Disrespectful whites	−0.133***	−0.145**
	(0.001)	(0.010)
Whites mention the experimental environment as information	0.008	−0.013
	(0.652)	(0.732)
Whites mention the experimental environment to justify their behavior	0.023	−0.026
	(0.603)	(0.518)
Whites mention the public spirit	0.018	0.007
	(0.635)	(0.786)
Whites mention whites' and blues' joint payoffs	0.016	0.029
	(0.669)	(0.452)
Period	−0.020***	−0.012***
	(0.000)	(0.005)
Constant	0.758***	0.304***
	(0.000)	(0.000)
Wald test results for comparison of treatment coefficients (*p*-values)		
FP vs. TDC	0.003	0.032
R^2	0.082	0.075
Number of clusters	15	15
Observations	2001	1230

Linear probability models. Dependent variable: Decision to vote for the respective efficient policy. Robust standard errors are clustered at the session level and *p*-values are given in parentheses: $*p < 0.10$, $**p < 0.05$, $***p < 0.01$. TopDown serves as baseline treatment in both regressions. All chat content categories that were recorded in at least 15% of the whites' chat messages (except specific voting recommendations) are included as explanatory variables

policy *A* as a measure of trust. In a linear probability model we regress the blues' propensity to vote for policy *A* on a dummy variable that indicates if the reported majority message in the previous period was inconsistent with the actual state of the world ("Potential lie"), and two further dummies that capture the tone of the whites' messages (respectful language, that is, rather appreciating statements and

Table 6 Deliberation treatments: Whites' decision to lie

	Only FullyPublic treatment (1)	All deliberation treatments (2)
FullyPublic (FP)		0.108**
		(0.014)
TopDownClosed (TDC)		0.104*
		(0.081)
Suspicious blue in previous period	−0.050	
	(0.353)	
Blue recommended voting for A in previous period	0.052	
	(0.294)	
Blue recommended voting for B in previous period	−0.065	
	(0.189)	
Blue recommended voting for C in previous period	0.093**	
	(0.045)	
Disrespectful blue in previous period	−0.014	
	(0.778)	
All blues voted for C in previous period	0.061***	0.040*
	(0.007)	(0.100)
# convinced blues in previous lie	0.068**	0.065***
	(0.044)	(0.003)
Period	0.005	0.010***
	(0.146)	(0.000)
Constant	0.088	−0.026
	(0.178)	(0.331)
Wald test result for comparison of treatment coefficients (p-values)		
FP vs. TDC		0.954
R^2	0.051	0.055
Number of clusters	5	15
Observations	545	1555

Linear probability models. Dependent variable: Decision to lie, conditional on receiving signal Y. Robust standard errors are clustered at the session level and p-values are given in parentheses: *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$. In regression (1) all chat content categories that were recorded in at least 15% of the blues' chat messages are included as explanatory variables. The variable *# convinced blues in previous lie* only takes into account falsely stated majority messages (=lies) that happened in the preceding period

disrespectful language, that is, rather impolite statements).[18] Moreover, we include four additional dummy variables that control for whether the whites mention the

[18] Examples for respectful language include statements like "you are totally right" and "well noted!", while "white capitalist gang" and "you probably have no friends" are typical examples for statements that were classified as disrespectful.

experimental environment as justification of their behavior ("our signals are not 100% correct" and similar statements) and attempt to appeal to the blues' public spirit. Lastly, we control for treatment by including dummies for *FullyPublic* and *TopDownClosed* (*TopDown* serves as baseline treatment) and the period of play.

Voting for *A* after majority message *X* is significantly less likely if the reported majority message in the previous period was inconsistent with the actual state of the world ("Potential lie") and if the whites treated the blues disrespectfully.[19]

Next we turn to those periods in which the whites report majority message *Y*. Since the whites have no incentive to make the blues believe that *Y* prevails if in fact they believe that it is *X*, what we study here is not the effects on the blues' trustfulness, but rather their cooperative response to having received a truthful message. For this, we consider their votes for the efficient policy *B*, given majority message *Y*. Specification (2) regresses the blues' propensity to vote for *B* on the same explanatory variables that we used in specification (1). We find that voting for *B* again is more likely in *FullyPublic* than in *TopDown* and less likely if the whites treat the blues disrespectfully. Moreover, treating the blues respectfully now has a significantly positive effect, potentially reinforcing the general positive effect of telling the truth.[20]

And lastly, similar to what we found in Table 4, the *Period* coefficient is significantly negative in both specifications in Table 5, too, indicating a general, treatment-unspecific tendency of the blues to vote for policy *C* increasingly often over periods.

Considering the found negative long-term consequences of whites' lies on blues' cooperativeness, the question arises why the whites lie to the blues and—considering that they do so also in the public chat in *FullyPublic* and *TopDown*—why they do so even at the expense of lying to their fellow whites. The regression specifications in Table 6 attempt to shed light on this question. In the reported linear probability models we regress the individual white players' decisions to lie (to report *X* instead of *Y*) on all chat content categories that were recorded in at least 15% of the blues' chat messages in *FullyPublic*. Moreover, we include a dummy that takes value 1 if all blue players that a white player was matched to in the previous period voted for *C* in that period. We also include a variable capturing the number of convinced blues (that is, the number of blue players who voted for *A* following a lie) in the previous period. Lastly, we control for the period of play and, in column (2), for treatment.

Specification (1) considers only the *FullyPublic* treatment. As evident, the whites' propensity to report a wrong state of the world increases if they encountered at least one blue player who recommended voting for *C* and if all blue players voted for *C*

[19] Whites' justifying themselves by referring to the experimental environment (e.g., stating that wrong messages can occur due to wrong signals) or whites' mentioning the group's "joint welfare" to appeal to the blues' cooperativeness have no significant effects on the blues' voting for the efficient policy after majority signal *X*.

[20] Whites' referring to the experimental environment in order to justify their behavior or mentioning the joint welfare to appeal to blues' cooperativeness, again, have no significant effects on the blues' voting decisions. Also, voting for *B* does not depend on the perceived correctness of the previous message about the state of the world (see the insignificant coefficient of "Potential lie").

in the previous period. Also, the more successful a lie was in the previous period (measured as number of convinced blues), the higher is a whites' propensity to lie again.

When considering all deliberation treatments (see specification (2)), we can only condition on blue players' voting decisions in previous periods since they have no opportunity to participate in the chat in *TopDown* and *TopDownClosed*. Again, the whites' propensity to lie increases if all blue players voted for *C* in the previous period and in the number of blue players who voted for *A* following a lie in the previous period. Similar to what we found in Table 3, the whites lie increasingly often in later periods of play. Also the general treatment differences reported in Table 3 remain significant after including the chat content categories.

Finding 5 (Deterioration of deliberation) The dynamics of communication reveal a general deterioration of blues' cooperativeness and trust and whites' truthfulness in all deliberation treatments. Moreover, in *FullyPublic* there is an additional factor that enforces deterioration, namely disrespectful language of the whites' and blues' recommendation and choice of policy *C*.

5 Discussion

We use a laboratory experiment to shed light on an important socio-economic issue: the difficulty of reaching an efficient collective policy choice in a democratic environment in which two social groups with different material interests have also different information and different access to communication channels. In our setting, one group has more information than the other on whether their material interests align or clash. In three deliberation treatments, we vary the inclusiveness of the deliberative structure that shapes the pre-vote communication of the two groups.

Compared with a benchmark setting without any communication, we find that communication leads to a higher propensity to vote for the efficient policies. More interestingly, we learn from our experiment that fully inclusive deliberation, as modelled in *FullyPublic*, has both advantages and disadvantages. The main advantage is that with full inclusiveness, the blues are more cooperative in their voting behavior: they reward truthfulness of the whites (i.e. revelation of majority message *Y*) more than in less inclusive settings. The main disadvantage, however, is that over time the whites become less truthful in the fully inclusive setting. The reason is the existence of a vicious circle: The more whites lie, the less blues vote for efficiency; the more blues recommend voting for *C*, the more whites lie. In addition, the emotional connotation of communication content is also relevant. In particular, whites' use of disrespectful language increases conflict. Our results here point to a phenomenon that we may call the curse of unrestricted communication: In an adversarial situation, the unrestricted back and forth of communication that is possible in the *FullyPublic* treatment may lead to an escalation in animosity. Our findings challenge the conception that fully inclusive deliberative environments will lead to the highest level of truthfulness. The reality of what is actually said in the communication can go against what a priori may appear to be an environment conducive to truth-telling.

Naturally, the particular deliberative structures we focus on are not the only interesting ones. In particular, it would be interesting to study a deliberative structure in which both whites and blues could separately communicate among each other before communicating with both groups. Such a structure would reflect the fact that in democratic societies the dependent classes also have access to restricted communication channels. It is not easy to gauge what the results would be in such an environment, but one may conjecture that conflict would be higher than in *FullyPublic*.

We believe that the phenomena we observe are relevant beyond our experiment. First, in modern democracies the advice pertaining to policy options given by experts and the more educated to the society at large is increasingly often ignored by the less informed members of society. This occurs out of a combination of distrust vis-à-vis those who are seen as privileged and the experience that expert knowledge is often less than perfect, so that expert advice that is ex post incorrect is not infrequent. Second, the immediacy and anonymity of inclusive communication that is now possible through digital media often leads to aggressiveness and disrespect between groups, which can make it difficult to reach a large societal consensus on important issues. If, instead, the informed group controls the communication process things can be even worse, because a group with a purely passive role in public communication loses sight of society's general interests and becomes particularistic.

Appendix

Additional tables

Table 7 Whites' votes for the efficient policies

	Majority signal: X		Majority message: X		Majority signal: Y		Majority message: Y	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	A vote	A vote	A vote	A vote	B vote	B vote	B vote	B vote
FullyPublic (FP)	0.078***	0.115***	0.013	0.001	0.418***	0.428***	0.044	0.094
	(0.000)	(0.001)	(0.428)	(0.910)	(0.000)	(0.000)	(0.375)	(0.128)
TopDown (TD)	0.067***	0.117***			0.472***	0.430***		
	(0.007)	(0.000)			(0.000)	(0.000)		
TopDownClosed (TDC)	0.083***	0.108***	−0.093	−0.034	0.446***	0.291***	0.063	0.104*
	(0.000)	(0.002)	(0.114)	(0.391)	(0.000)	(0.001)	(0.403)	(0.066)
Period	0.002**	0.005**	−0.002	−0.000	−0.010***	−0.015***	−0.003	−0.001
	(0.014)	(0.013)	(0.162)	(0.796)	(0.000)	(0.005)	(0.110)	(0.599)
FP ×Period		−0.004*		0.001		−0.002		−0.005
		(0.076)		(0.302)		(0.810)		(0.210)
TD × Period		−0.005**				0.004		
		(0.030)				(0.476)		
TDC ×Period		−0.003		−0.005**		0.014**		−0.004
		(0.228)		(0.044)		(0.039)		(0.447)
Constant	0.887***	0.860***	0.997***	0.979***	0.354***	0.401***	0.838***	0.810***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Wald test results for comparison of treatment coefficients (p-values):								
FP vs. TDC	0.549		0.068		0.664		0.792	
FP vs. TD	0.546				0.347			
TD vs. TDC	0.364				0.711			
R^2	0.039	0.043	0.049	0.055	0.164	0.169	0.008	0.009
Number of clusters	20	20	15	15	20	20	15	15
Observations	2451	2451	2112	2112	2229	2229	1278	1278

Linear probability models. Dependent variable: Decision to vote for the respective policy. Robust standard errors are clustered at the session level and p-values are given in parentheses: *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$. NoChat serves as baseline treatment in regressions (1), (2), (5) and (6). TopDown is the baseline treatment in the remaining regressions

"FP" refers to FullyPublic; "TD" refers to TopDown,and "TDC" refers to TopDownClosed

Table 8 Deliberation treatments: Blues' votes for the efficient policies—Chat messages coded by Coder #2

	Majority message: X (1) A vote	Majority message: Y (2) B vote
FullyPublic (FP)	0.151**	0.171***
	(0.041)	(0.005)
TopDownClosed (TDC)	−0.065	−0.018
	(0.224)	(0.526)
Potential lie in previous period	−0.049**	−0.014
	(0.045)	(0.651)
Respectful whites	0.053	0.266***
	(0.351)	(0.004)
Disrespectful whites	−0.166**	−0.202***
	(0.016)	(0.003)
Whites mention the experimental environment as information	0.034	0.006
	(0.274)	(0.848)
Whites mention the experimental environment to justify their behavior	−0.006	−0.076*
	(0.835)	(0.053)
Whites mention the public spirit	−0.072*	−0.068**
	(0.094)	(0.018)
Whites mention whites' and blues' joint payoffs	0.061	0.124***
	(0.104)	(0.001)
Period	−0.020***	−0.013***
	(0.000)	(0.003)
Constant	0.758***	0.296***
	(0.000)	(0.000)
Wald test results for comparison of treatment coefficients (*p*-values)		
FP vs. TDC	0.001	0.005
R^2	0.086	0.095
Number of clusters	15	15
Observations	2001	1230

Linear probability models. Dependent variable: Decision to vote for the respective efficient policy. Robust standard errors are clustered at the session level and p-values are given in parentheses: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. TopDown serves as baseline treatment in both regressions. All chat content categories that were recorded in at least 15% of the whites' chat messages (except specific voting recommendations) are included as explanatory variables

Table 9 FullyPublic treatment: Whites' decision to lie—Chat messages coded by Coder #2

	Only FullyPublic treatment
	(1)
Suspicious blue in previous period	−0.001
	(0.983)
Blue recommended voting for A in previous period	0.034
	(0.300)
Blue recommended voting for B in previous period	−0.071
	(0.171)
Blue recommended voting for C in previous period	0.077
	(0.102)
Disrespectful blue in previous period	0.081
	(0.138)
All blues voted for C in previous period	0.044*
	(0.063)
# convinced blues in previous lie	0.069**
	(0.038)
Period	0.004
	(0.101)
Constant	0.081
	(0.233)
R^2	0.056
Number of clusters	5
Observations	545

Linear probability model. Dependent variable: Decision to lie, conditional on receiving signal *Y*. Robust standard errors are clustered at the session level and *p*-values are given in parentheses: $*p < 0.10$, $**p < 0.05$, $***p < 0.01$. All chat content categories that were recorded in at least 15% of the blues' chat messages are included as explanatory variables. The variable *# convinced blues in previous lie* only takes into account falsely stated majority messages (=lies) that happened in the preceding period

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10683-021-09729-4.

Acknowledgements We would like to thank seminar participants at Caltech, University of Konstanz, University of Marburg, GATE Lyon and Erasmus University Rotterdam as well as participants of the Workshop on Microeconomics at Leuphana University Lüneburg 2016, the 10th Maastricht Behavioral and Experimental Economics Symposium, the EWEBE in Bologna 2017, the TIBER 2017 Symposium on Psychology and Economics at Tilburg University and the UECE Game Theory Lisbon Meetings 2017 for their helpful comments and useful suggestions. Lydia Mechtenberg gratefully acknowledges the hospitality of Caltech in spring 2018. The authors gratefully acknowledge financial support from the Spanish Ministry of Economics and Competitiveness through Grant: ECO2017-88130 and through the Severo Ochoa Program for Centers of Excellence in R&D (SEV2015-0563), the Generalitat de Catalunya (Grant: 2017 SGR 1136) and the Antoni Serra Ramoneda (UAB – Catalunya Caixa) Research Chair as well as from the Graduate School of the Faculty of Business, Economics and Social Sciences, Universität Hamburg.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Data availability All data, analysis codes and experimental materials are included in the Electronic Supplementary Materials. The data and analysis codes have also been deposited in the openICPSR data repository (https://doi.org/10.3886/E147521V1) .

References

Akerlof, George A., & Kranton, Rachel E. (2000). Economics and identity. *The Quarterly Journal of Economics, 115*(3), 715–753.

Akerlof, George A., & Kranton, Rachel E. (2010). *Identity Economics*. How Our Identities Shape Our Work: Wages, and Well-Being, Princeton University Press.

Ambrus, Attila, Azevedo, Eduardo M., & Kamada, Yuichiro. (2013). Hierarchical cheap talk. *Theoretical Economics, 8*(1), 233–261.

Andreoni, James, & Rao, Justin M. (2011). The power of asking: How communication affects selfishness, empathy, and altruism. *Journal of Public Economics, 95*(7), 513–520.

Apffelstaedt, Arno & Mechtenberg, Lydia (2020). Competition for Context-Sensitive Consumers. *Management Science, 67*(5), 2828–2844.

Benoît, J.-P., & Juan, D. (2016). A theory of rational attitude polarization, *Working Paper, Social Science Research Network*.

Bock, Olaf, Baetge, Ingmar, & Nicklisch, Andreas. (2014). hroot: Hamburg registration and organization online tool. *European Economic Review, 71*, 117–120.

Bowles, Samuel, & Polanía-Reyes, Sandra. (2012). Economic incentives and social preferences: Substitutes or complements? *Journal of Economic Literature, 50*(2), 368–425.

Brandts, Jordi, & Cooper, David J. (2007). Its what you say, not what you pay: An experimental study of manager-employee relationships in overcoming coordination failure. *Journal of the European Economic Association, 5*(6), 1223–1268.

Brandts, Jordi, David J. Cooper, and Christina Rott, Communication in laboratory experiments, in Handbook of research methods and applications in experimental economics, Edward Elgar Publishing, 2019.

Buechel, Berno, & Mechtenberg, Lydia. (2019). The swing voters curse in social networks. *Games and Economic Behavior, 118*, 241–268.

Charness, Gary, & Dufwenberg, Martin. (2006). Promises and partnership. *Econometrica, 74*(6), 1579–1601.

Charness, Gary, & Dufwenberg, Martin. (2010). Bare promises: An experiment. *Economics Letters, 107*(2), 281–283.

Chen, Roy, & Chen, Yan. (2011). The potential of social identity for equilibrium selection. *American Economic Review, 101*(6), 2562–2589.

Chen, Y., & Sherry, X. L. (2009). Group identity and social preferences. *The American Economic Review, 99*(1), 431–457.

Cohen, Joshua. (1989). Deliberation and democratic legitimacy. In A. Hamlin & P. Pettit (Eds.), *The good polity: Normative analysis of the state* (pp. 67–92). Oxford: Blackwell.

Dawes, Robyn M, Alphons JC Van de Kragt, and John M Orbell, Cooperation for the benefit of us: Not me, or my conscience, in J. Mansbridge, ed., *Beyond Self-Interest*, The University of Chicago Press, 1990, pp. 97–110.

Dryzek, John S., & Christian, List. (2003). Social choice theory and deliberative democracy: A reconciliation. *British Journal of Political Science, 33*(1), 1–28.

Fischbacher, Urs. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics, 10*(2), 171–178.

Goeree, Jacob K., & Leeat, Yariv. (2011). An experimental study of collective deliberation. *Econometrica, 79*(3), 893–921.

Guarnaschelli, Serena Richard, McKelvey, D., & Palfrey, Thomas R. (2000). An experimental study of jury decision rules. *American Political Science Review, 94*(2), 407–423.

Gutmann, Amy, & Thompson, Dennis. (1996). *Democracy and disagreement: Why moral conflict cannot be avoided in politics, and what can be done about it, Cambridge*. MA: Harvard University Press.

Habermas, Jürgen. (2015). *Between facts and norms: Contributions to a discourse theory of law and democracy, Cambridge*. MA: MIT Press.

Karpowitz, Christopher F and Tali Mendelberg, An experimental approach to citizen deliberation, in James N. Druckman, Donald P. Green, James H. Kuklinski, and Arthur Lupia, eds., *Cambridge Handbook of Experimental Political Science*, Cambridge University Press Cambridge, 2011, pp. 258–272.

Landwehr, Claudia. (2010). Discourse and coordination: Modes of interaction and their roles in political decision-making. *Journal of Political Philosophy, 18*(1), 101–122.

Myers, C Daniel and Tali Mendelberg, Political deliberation, in David Sears Leonie Huddy and Jack Levy, eds., *Oxford Handbook of Political Psychology*, Oxford University Press, 2013, pp. 699–734.

Mäki, Uskali. (2005). Models are experiments, experiments are models. *Journal of Economic Methodology, 12*(2), 303–315.

Orbell JM, Alphons J Van de Kragt, Robyn MD, Explaining discussion-induced cooperation.,*Journal of Personality and Social Psychology*, 1988, *54* (5): 811–819.

Palfrey, Thomas R, Experiments in political economy, in A.Roth and John H.Kagel, eds., *Handbook of Experimental Economics*, Vol. 2, Princeton University Press, 2016, pp. 347–434.

Palfrey, Thomas R., & Pogorelskiy, Kirill. (2017). Communication among voters benefits the majority party. *Economic Journal, 129*(618), 961–990.

Pronin, Kira and Jonathan Woon, Public deliberation, private communication, and collective choice, *Working Paper, New York University*, 2017.

Robalo, P., Schram, A., & Sonnemans, J. (2017). Other-regarding preferences, in-group bias and political participation: An experiment. *Journal of Economic Psychology, 62*, 130–154.

Sandel, Michael J., *The tyranny of merit: What's become of the common good* 2020.

Schotter, Andrew, On the relationship between economic theory and experiments, in G.R. Frechette and Andreas Schotter, eds., *The Handbook of Experimental Economic Methodology*, Oxford University Press, 2015, pp. 58–85.

Sunstein, Cass R. (2009). *R, Going to extremes: How like minds unite and divide*. Oxford: Oxford University Press.