Deliberative Structures and their Impact on Voting under Economic Conflict

Jordi Brandts
Leonie Gerhards
Lydia Mechtenberg

This version: May 2019

February 2018

Barcelona GSE Working Paper Series

Working Paper nº 1022

Deliberative structures and their impact on voting under economic conflict*

Jordi Brandts † Leonie Gerhards‡ Lydia Mechtenberg‡

May 20, 2019

Abstract

Inequalities in democracies not only involve economic differences, but also differences in access to information and social influence. We identify the *tragedy of the informed*: Privileged access to information about economic conditions can create lying incentives. In a laboratory experiment, we study an electorate that consists of two groups, one informed and one uninformed about an uncertain state of the economy. Incentives depend on this state. Before voting the two groups can communicate. In addition to a treatment without communication, we study three different *deliberative structures* that vary in how much the uninformed can partake, i.e., in inclusiveness. We hypothesize that these deliberative structures affect preferences and voting and that their efficiency-enhancing effect on voting outcomes increases with increasing inclusiveness. This predicted efficiency ranking is confirmed by the data, but the differences in total expected earnings are not statistically significant, despite significant differences in voting behavior. We find three reasons for this unpredicted flatness of the efficiency ranking: First, the uninformed do not anticipate how lying behavior of the informed varies with the deliberative structure. Second, compared to the other deliberative structures, fully inclusive deliberation better allows the uninformed to coordinate – not only alongside the informed, but also against them. Third, the back-and-forth of communication and votes leads to growing animosity between the informed and the uninformed and hence to a deterioration of economic consensus.

Keywords: Communication, Economic Conflict, Inequality, Experiments
JEL codes: C92, D9

*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.

†Corresponding author: Institut d'Anàlisi Econòmica (CSIC) and Barcelona GSE, Campus UAB, 08193 Bellaterra (Barcelona), Spain. Fax: +34 93 580 1452. Tel.: +34 93 580 6612. Email: jordi.brandts@iae.csic.es

‡Universität Hamburg

1 Introduction

Inequality in democracies does not only involve differences in economic opportunities, but also differences in access to information about economic conditions, as well as different levels of social influence through communication channels. A sound analysis of how these inequalities interact and potentially even reinforce each other is essential for a better understanding of some of the current tensions in modern societies. As a first step, we present the results from a laboratory experiment based on a voting game that sheds light on these interactions.

We study an environment which represents a society split into two distinct groups. The state of the world is uncertain and while the members of one of the groups have some information about the state of the world, the members of the other group are uninformed. In both states of the world 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, whereas in the other state their material interests are aligned. Hence, there is a state in which a consensus should be easy to reach and another state which leads to potential conflict. The collective choice of policy is determined through a vote in which all individuals from both groups can participate. Before the vote takes place the individuals of both groups can communicate with each other under protocols or *deliberative structures* that differ in how the hierarchy between the informed and the uninformed is designed. Our focus is on how such different deliberative structures affect preferences, information aggregation and transmission, voting behavior, and outcomes.

Our motivation stems primarily from democracies in which the information about prevailing economic conditions is unequally distributed between social groups (Borgonovi, d'Hombres, and Hoskins, 2010, Feddersen and Pesendorfer, 1996, Morton and Tyran, 2011, Pande, 2011). In particular, in cases where the better informed and the less-well informed parts of the society have different economic interests, efficient outcomes may be hard to reach, because the problems of information aggregation and transmission interact with the conflict between social groups. Information about the economic fundamentals transmitted by the informed to the uninformed may not be truthful or may be suspected not to be. All this may lead to society's inefficient use of information and to economic losses.

Thus, democracy at its best involves more than just decision making through voting. It also requires a process of interaction between members of a society, without exclusion, in which different economic options are discussed. When this process takes place under ideal conditions it is often referred to as deliberation. Through a fully inclusive deliberation process people can become more willing to care about the interest of the society as a whole

(Dawes, Van de Kragt, and Orbell, 1990, Orbell, Van de Kragt, and Dawes, 1988, Dryzek and List, 2003).[1] However, if communication before voting takes place under restrictive conditions then the outcome may be that the members of one or more of the social groups involved in the process simply defend the material interest of their own group.

The first circumstance that may matter for how deliberative structures affect group behavior is whether a group has access to addressing the other group or is prevented from doing so and, hence, has a mere passive role. If a group is subordinated to others in such a way it may be less inclined to take the interests of the society as a whole into account. Second, it may also be of consequence whether members of a group have to directly address those of the other group or can first communicate among each other in a private secluded way. Separate communication between groups has been hypothesized to lead to polarization in opinions (Sunstein, 2009, Benoît and Dubra, 2014).

We propose that different deliberative structures trigger distinct preferences of the members of the two groups in the spirit of the notion of state-dependent preferences introduced by Bowles and Polanía-Reyes (2012). State-dependence arises because actions are motivated by a heterogeneous repertoire of preferences the salience of which depends on the nature of the decision situation (p. 373). 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.[2] In our case we propose that different deliberative structures affect group identities and hence players' preferences.

In our experiments the two groups have different payoffs and receive different information, so that there are two fundamental asymmetries between them. We use different colors to refer to the two groups, *white* for the informed and *blue* for the uninformed.[3] Given these differences we study how group identity is affected by the different deliberative structures. We vary the audience that players of each color can address at the communication stage. Our hypothesis is that these variations lead to two different group identities of the two groups, being either color-group identity or voting-group identity, i.e., the group that includes both color groups.

We study three deliberative structures in four distinct experimental treatments. In our baseline treatment, *NoChat*, there is no communication between participants. The three deliberative structures model different degrees of openness or inclusiveness of a society.

[1] 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.

[2] Bowles and Polanía-Reyes (2012) focus on how the presence of monetary incentives triggers different preferences. In an industrial-organization setting, Apffelstaedt and Mechtenberg (2018) analyze context-dependent consumer preferences in a competitive market.

[3] The experiment was conducted in Germany, where "white" and "blue" (collar) do not have the same social connotation as in the English-speaking world.

In the first of the deliberative structures we study, called *Deliberation*, the two groups are on equal foot and communication is unrestricted. All members of both groups can freely chat with each other. Specifically, each individual can write messages that are seen by all other participants and read the messages written by all other participants. With this structure we want to represent the ideal situation of an open society where all members of a society can participate under equal conditions in the exchange of ideas that takes place before voting.

In the two other structures we incorporate into the modeling the unequal access of different social groups to the public communication process in actual democracies. In the deliberative structure called *TopDown* only the informed group has access to public communication channels. In this case all members of the informed group can write messages that are seen by all other participants, whereas all the members of the uninformed group can read all the messages written by all participants in the informed group, but can themselves not write any messages. Hence, those who are better informed are also those who dominate the communication process. Nevertheless in *TopDown* the content of whites' communication is transparent. The society is integrated and all its members know what is being said at all times. By contrast, in *TopDownClosed*, there is an additional element of segregation in the communication process. In this structure, there are two stages. First, all members of the informed group can freely communicate with each other without the uninformed being able to read these messages. The second stage is like in *TopDown* above, i.e. all members of the informed group can write messages that are seen by all other participants, whereas all the members of the uninformed group can read all the messages written by all participants in the informed group, but can themselves not write any messages.

Given the idea that different deliberative structures affect group identities and hence players' preferences, we take to the extreme the claims of normative deliberation theory and propose the following: In *NoChat*, the treatment without any communication, both groups simply have the identity of their own group (color-group identity). In *Deliberation* both groups have the identity of the society as a whole (voting-group identity), since in the communication process they both address the society, i.e. the voting group as a whole. In *TopDown* the uninformed group has a color-group identity again, since it has merely a passive role in communication. By contrast the informed group has a voting-group identity, since its members are still in a position in which they directly address the society as a whole and take the interests of all members of society into account. Finally, in *TopDownClosed* both groups have a color-group identity for the following reasons: For the uninformed the situation is the same as in *TopDown*: they are fully passive and hence have their default identity. For the informed we conjecture that they also adopt the identity of their own group, since they communicate privately among themselves before

addressing the society as a whole, a circumstance that makes them focus exclusively on the interests of their own group.

We predict that communication and voting behavior is based on the adopted group identity. This implies that in a group with voting-group identity, individuals have efficiency preferences and maximize the expected material payoffs of the entire society. In a group with color-group identity, individuals only maximize the expected material payoffs of their own group. Therefore, according to our prediction efficiency is highest in *Deliberation*, second-highest in *TopDown* and lowest in *TopDownClosed* and *NoChat*. We formalize these ideas in the Theoretical Appendix C. Equilibrium analysis of the resulting game provides us with comparative-static predictions guiding the empirical analysis of our experimental data (see Schotter, 2015).

The rest of the paper is organized as follows. Section 2 reviews the relevant literature, Section 3 presents the experimental design and procedures. Section 4 contains our hypotheses. Section 5 reports on our results. Section 6 presents what we can learn from the deviations; and in Section 7 we relate the phenomena we observe in the laboratory to social and economic issues that we observe in reality.

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, McKelvey, and Palfrey (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 (2017) show that pre-vote communication in social networks that is restricted to information aggregation can lower efficiency even in a common-interest setting.

Moreover, previous experimental work has found evidence in favor of communication affecting group identity (see Akerlof and Kranton, 2000, 2010). 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, Schram, and Sonnemans (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 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. However, they focus on voter turnout in an experiment with individuals belonging to two competing parties and costly voting. The issue of voter turnout is quite different from the research question that we address. Nonetheless, the distinction between public communication (all voters exchange messages through a computer chat) and party communication (messages are only exchanged within each party) is related to our distinction between different deliberative structures. Their result is that both types of communication favor the majority party.

Pronin and Woon (2017) study how the economic benefits of deliberation are 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, Azevedo, and Kamada, 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 both conflict solution and economic efficiency are affected by communication structures.

3 Experimental design

The game 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 color. 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.

Subsequently, a communication stage starts. We consider three alternative deliberative

structures: (i) Whites and blues can publicly communicate with each other; (ii) the whites but not the blues can send (public) messages; and (iii) the whites can first communicate with each other unobserved by the blues and then send public messages that are also received by the blues. Hence, moving from (i) to (iii), the whites get gradually more control over the communication process. On all communication stages, messages are sent simultaneously, and sending an empty message is possible for all senders.

Table 1: Payoffs for blue and white players, conditional of the state of the world and implemented policies

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

Finally, the voting stage starts. 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 implemented 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. Given the chosen payoffs, state X can be considered the good state of the world, 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).

In the good 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 the bad state Y: While the whites would prefer B to be chosen, the blues would prefer C instead. Hence, the two color groups have a state-dependent conflict. This conflict in state Y is particularly sharp since, in the eyes of the whites, C is the worst of all options. The state-dependent efficient policy choice would be A in state X and B in state Y and is hence in line with the preferences of the whites.

We chose this design for two reasons. First, we wanted to model an understudied informational asymmetry that is often observed in reality: Only one group (the whites) has information on whether or not their material interests conflict with those of the other group (the blues). Second, we wanted to rule out a trade-off between efficient information

aggregation and efficient policy-choice. Though interesting in itself, such a trade-off is not what we want to study in this paper. Our game is designed to study a combined information-transmission and collective-choice problem if only one part of the collective has information about whether the choice is to be made in a common-interest situation or in the presence of group conflict.

The state-dependent conflict gives the white players an incentive to lie about the state of the world, if, given their signals, they expect the bad state of the world 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 the good 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 their optimal policy B in state Y. Obviously, and as shown in Appendix C, successful lies cannot be part of an equilibrium here – instead, communication would become meaningless ("babbling").

Stretching the interpretation of the game, this dilemma could be called the *tragedy of the informed*: If those who do not belong to the better-informed group do not internalize its interests but only care for their own, the better informed have an incentive to use their informational advantage to manipulate the less-well informed away from the economic conflict. But if they do so, trust and hence information aggregation break down and the conflict sharpens.

Experimental treatments We conducted four experimental treatments as depicted in Table 2.[4] The treatments *Deliberation*, *TopDown*, and *TopDownClosed* implement the above game with communication stage (i), (ii), and (iii), respectively. Communication is implemented as computerized free-form chat. In *Deliberation* 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.[5] We decided to exogenously restrict the duration of the chat stage in order to keep the total duration of the experimental sessions comparable within and across treatments.

Treatment *NoChat* implements the above game without the communication stage. However, directly after the information stage, our subjects in *NoChat* have the opportu-

[4]Translated instructions to all treatments are included in the Supplementary online material (SOM)

[5]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.

nity to take private notes in a computer window that looks exactly like the chat window in the other treatments. We thus tightly control the task- and time-structure of all treatments.

Table 2: Implemented deliberative structures across treatments

	White players		Blue players	
	write	read	write	read
NoChat	–	–	–	–
Deliberation	✓	✓	✓	✓
TopDown	✓	✓	–	✓
TopDownClosed	✓/✓	✓/✓	–	–/✓

In *TopDownClosed* the first entry refers to the private chat among the whites, the second entry relates to the subsequent public chat.

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. All subjects received IDs that indicated their color type (white or blue) and a number between 1 and 3 (for instance, "Blue 2"). These IDs were randomly assigned in the beginning of every period such that subjects were not able to recognize and track individuals throughout the different periods.

Procedures Overall, we conducted 20 sessions with 468 subjects in total, half of them assumed the roles of white, the other half the roles of blue players. In *NoChat* and *Deliberation* we ran five sessions each, all of them comprising 24 subjects. In *TopDown* and *TopDownClosed* we ran four sessions with 24 subjects and one session with 18 subjects. Sessions lasted for 20 periods. Subjects were randomly assigned their color (white or blue) at the beginning of the session and kept it throughout the 20 periods of the experiment. The groups, however, were randomly re-matched at the beginning of each period (stranger matching).

We used z-tree developed by Fischbacher (2007) to computerize our treatments and the recruiting software hroot developed by Bock, Baetge, and Nicklisch (2014) to randomly assign subjects to treatments. The experiment was run with student subjects from various study backgrounds at the WISO-laboratory of Hamburg University. During the sessions payments were expressed in experimental currency 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).

For the three communication treatments we analyzed the chat content following the procedures of Cooper and Kagel (2005) and Brandts and Cooper (2007).

4 Hypotheses

Our hypotheses below are derived from our equilibrium analysis in Appendix C.[6] The basic intuition of how the equilibria depend on the deliberative structures can be presented in terms of the *color-group* and *voting-group identities* of whites and blues and whites' lying incentives.

Consider first *TopDownClosed*, where both groups have a *color-group identity*, i.e., their preferences are not efficiency-oriented towards the total voting group but restricted to their own material interests. Then the interaction between the groups involves a dilemma that leads to inefficiency in both states of the world: The blues prefer C over B if state Y is more likely than X. Since C is the worst choice for the whites under any signal distribution, the whites have the incentive to lie to the blues, making them believe that the state is X rather than Y and that, therefore, A is the blues' best choice, rather than C. But if the whites lie, their messages will not be believed in equilibrium. Hence, only babbling equilibria exist. Therefore, the blues will be even more motivated to vote for C, which is the policy that benefits them most in expectation if information about the true state is absent.

Consider now *TopDown*. There the dilemma described above is ameliorated since the whites have efficiency preferences: They do not lie to the blues about the signals they have even if the blues vote for C when they learn that the majority signal is Y. Hence, in equilibrium the blues now obtain information about the true state. Following their material interests, they vote for A when they are told that the state is X, and for C when they are told that it is Y. Thus the voting outcome is efficient in X, but inefficient in Y.

Finally, consider *Deliberation*. Here, both whites and blues have efficiency preferences. Therefore, the following strategies of the two colors are efficient and part of an equilibrium: 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. Note that 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 vote for their second-best choice B instead, which is efficient. Note that given this behavior of the blues, the whites have no incentive to lie to them about the state.[7]

[6] We use equilibrium selection criteria to ensure that (1) messages are used for information transmission and not as non-informative coordination devices, (2) the most informative equilibrium is played, (3) players with the same color who are in the same information set use the same strategy, and (4) this strategy maximizes their utility, given their information. Note that Appendix C also contains standard economic predictions (which predict no treatment differences between *NoChat*, *Deliberation* and *TopDown* and only one minor difference between these treatments and *TopDownClosed*).

[7] Since our formal modeling of the idea that deliberation makes participants more cooperative is an extreme interpretation of what normative deliberation theorists have in mind, any deviation from the resulting predictions must not be read as a falsification of these theories.

Table 3 summarizes how our four different treatments affect (1) the preferences of the blues, (2) the possibility of information aggregation, and (3) the incentive of the whites to lie to the blues.

Table 3: Preferences for efficiency and their impact on information aggregation

	Efficiency preferences of the...		Equilibrium incentives of the whites to lie to the blues	Information aggregation is possible
	Whites	Blues		
NoChat	no*	no	–	no
Deliberation	✓	✓	no	✓
TopDown	✓	no	no	✓
TopDownClosed	no	no	✓	✓**

✓indicates for each of the treatments if, in equilibrium, (i) the whites and blues assume a voting-group identity and hence have efficiency preferences, (ii) if the whites have an incentive to lie to the blues and (iii) whether information aggregation is possible. *In equilibrium in NoChat, the whites act as if they had efficiency preferences. **Note that in TopDownClosed, *in equilibrium* information aggregation is only possible in the private chat, but not in the public chat.

Based on our equilibrium analysis in Appendix C, our hypotheses below consider the following possible voting outcomes: A/A (all votes placed by whites/blues are for policy A), A/C (the whites vote for A and the blues for C), B/B (all votes placed by whites/blues are for policy B), and B/C (the whites vote for B and the blues for C). We call the voting outcomes in which the blues vote for C a *conflict outcome*.[8] Moreover, we also consider what we call the *split-whites* equilibrium: The whites vote in line with their individual signals, i.e., either for A or for B, while the blues vote for C. Finally, while in Appendix C we also report equilibria with abstention, we do not consider them here, for two reasons. First, they are outcome-equivalent (with respect to the probability of A, B, or C being implemented) to equilibria without abstention. Second, we find extremely few instances of abstention in our data.

Based on Propositions 1, 2, 3, and 4 (see Appendix C) that pertain to behavior in *NoChat (NoC)*, *Deliberation (D)*, *TopDown (TD)*, and *TopDownClosed (TDC)*, we predict to observe the following voting outcomes: Voting strategies per treatment are as presented in Table 4; and the comparative statics across treatments are as summarized in hypotheses 1-4. For readability, in the hypotheses we only refer to predicted treatment differences, and not to predicted absences of differences. Hence, if in the results section we find a

[8]Other voting outcomes that are not part of any equilibrium might empirically occur, too. One salient example would be B/A under majority signal Y: The whites, though informed that Y is likely to pertain, successfully convince the blues that X pertains, so that, while they themselves vote for B, the blues vote for A, which is better than C for the whites. This is no equilibrium (since in equilibrium, lies would not be believed), but though not predicting such outcomes, we do not exclude them from the empirical analysis.

Table 4: Whites' and blues' predicted voting decisions, by treatment and majority signal

	X		Y	
	Whites	Blues	Whites	Blues
NoChat	A or B	C	A or B	C
Deliberation	A	A	B	B
TopDown	A	A	B	C
TopDownClosed	A	C	B	C

treatment difference not specified by our hypotheses below this difference is to be treated as a deviation from a hypothesized equality.

Hypothesis 1a (Voting outcomes given majority signal X) *The frequency ordering of A/A (A/C) is: D, TD > TDC, NoC (TDC ≥ NoC > D, TD).*

Hypothesis 1b (Voting outcomes given majority signal Y) *The frequency ordering of B/B (B/C) is: D > NoC, TD, TDC (TD, TDC ≥ NoC > D).*

Hypothesis 2a (Whites' voting decisions) *Given majority signal X (Y), the frequency ordering of whites' votes for A (B) is: D, TD, TDC ≥ NoC. Whites never vote for C.*

Hypothesis 2b (Blues' voting decisions) *Given majority signal X (Y), the frequency ordering of blues' votes for A (B) is: D, TD > NoC, TDC (D > NoC, TD, TDC). The frequency orderings of blues' votes for C are reversed, compared to the above frequency orderings.*

Hypothesis 3a (Whites' truth-telling) *The frequency ordering of instances where the majority message equals the majority signal is: D, TD > TDC.*

Hypothesis 3b (Blues' trustfulness) *The frequency ordering of blues' votes for A (C) after majority message X is: D, TD > TDC (D, TD < TDC).*

Hypothesis 4a (Efficiency ranking) *The frequency ordering of total expected joint earnings is: D > TD > TDC > NoC.*

Hypothesis 4b (Earnings' rankings) *The frequency ordering of whites' (blues') expected earnings is: D > TD > TDC > NoC (TD > TDC > NoC > D).*

5 Results

5.1 Summary of main findings

In the subsections below, we present the results that correspond to our hypotheses above. Together, these results confirm the following predictions: Communication leads to higher total expected earnings than *NoChat*. In *Deliberation* the blues vote more often for the efficient policy even if this is against their material interest than in the other treatments. Hence, compared to the other two deliberative structures, *Deliberation* yields more voter coordination on the efficient policy in the bad state Y despite conflicting interests of colors. Conversely, in *TopDownClosed* the whites lie more and the blues vote less often for the efficient policy than in *Deliberation*, which results in less voter coordination on the efficient policies in either state of the world.

However, we also find deviations with respect to almost all our hypotheses. These deviations aggregate to two main findings: First, *Deliberation* is less efficient in terms of total (and whites') earnings than predicted, and second, *TopDownClosed* is more efficient than predicted. These two results lead to insignificant differences in total expected joint earnings between the three communication treatments, although qualitatively, the ranking of deliberative structures according to total expected earnings is as predicted.

As we will show in subsections 6.2 and 6.3, the deviations from our predictions have important and interesting reasons: First, we show in subsection 6.2 that having a voice in *Deliberation* enables the blues to better coordinate among themselves than in the other treatments. As a consequence, they not only coordinate more often on the efficient policy (A or B), but also not less often on the least efficient policy C, compared to the other communication treatments. Second, in subsection 6.3 we analyze the dynamics and show that the back-and-forth of communication between the whites and blues in *Deliberation* leads to increasing animosity and dishonesty and a resulting decrease of efficiency over time. Third, we find that the blues do not correctly anticipate how whites' lying behavior depends on the deliberative structure: Instead of trusting the whites (i.e., voting for A) more in *TopDown* than in *TopDownClosed* when the whites send the doubtful message X, the blues exhibit what we call *flat (dis-)trust* in these two treatments. In addition, there is weak evidence that they trust the whites even more in *Deliberation* than in the two other communication treatments in which they have no say, i.e., they do not anticipate the deterioration of whites' truthfulness in *Deliberation*.

Hence, giving the uninformed a voice has important consequences for the functioning of deliberation that partly counteract each other: When given a voice, the uninformed become more cooperative but also more coordinated if uncooperative; and they do not anticipate how much their coordinating their uncooperative votes puts whites' truthfulness

under pressure.

5.2 Voting outcomes at the group level

We start by testing our hypotheses on realized voting outcomes conditional on the majority signal, that is, the signal received by the majority of whites (Hypotheses 1a and 1b). In Table 5 we present the frequencies of the following voting outcomes that we introduced in Section 4 above: *A/A*, *B/B*, *A/C*, and *B/C*. As a reminder: *A/A* and *B/B* refer to voting outcomes in which all six voting-group members either vote for A or B, respectively. *A/C* and *B/C* describe voting outcomes in which the three blues vote for C and the three whites either for A or B, respectively. We expect *A/A* and *A/C* to only occur if the majority signal is X, and *B/B* and *B/C* only if the majority signal is Y. Besides these voting outcomes, we also consider the split-whites outcome in which all blues vote for C and all whites follow their individual signal (vote for A if their own signal is X, vote for B if it is Y).[9]

Table 5: Frequencies of voting outcomes at the group level – Conditional on the received majority signal

	Majority signal: X				Majority signal: Y			
	NoChat	Deliberation	TopDown	TopDownClosed	NoChat	Deliberation	TopDown	TopDownClosed
A/A	0	**0.386**	**0.201**	0.143	0	0.059	0	0.021
Almost A/A	0.038	0.284	0.358	0.328	0.018	0.059	0.055	0.105
(Almost) A/A	0.038	0.670	0.559	0.471	0.018	0.119	0.055	0.126
A/C outcome	**0.338**	0.121	0.106	**0.159**	**0.295**	0.103	0.114	0.052
of this: Split-whites	0.184	0.023	0.011	0.037	-	-	-	-
Almost A/C outcome	0.513	0.181	0.307	0.339	0.476	0.086	0.095	0.126
(Almost) A/C outcome	0.850	0.302	0.413	0.497	0.771	0.189	0.209	0.178
B/B	0	0	0	0	0	0.086	0.005	0.005
Almost B/B	0	0.005	0	0.005	0	0.124	0.065	0.058
(Almost) B/B	0	0.005	0	0.005	0	0.211	0.070	0.063
B/C outcome	0	0	0	0	0.012	**0.216**	**0.313**	**0.283**
of this: Split-whites	-	-	-	-	0.006	0.092	0.164	0.152
Almost B/C outcome	0.013	0	0.006	0	0.090	0.238	0.328	0.257
(Almost) B/C outcome	0.013	0	0.006	0	0.102	0.454	0.642	0.539
Split-whites	0.269	0.023	0.017	0.037	0.030	0.108	0.174	0.168
Other	0.098	0.023	0.022	0.026	0.108	0.027	0.025	0.094
Observations	234	215	179	189	166	185	201	191

In "Almost" outcomes at most one player per color group deviates from the respective outcome. "(Almost)" outcomes comprise both the "Almost" outcomes and the outcomes itself, without deviation. Grey cells indicate predicted equilibrium outcomes, figures printed in bold highlight the observed modal voting outcomes for the respective treatment and signal combination.

Grey cells in Table 5 indicate the predicted equilibrium outcomes as derived in Appendix C, figures printed in bold highlight observed modal voting outcomes. Besides the precise equilibrium outcomes, we also present information about voting outcomes in which

[9]See Appendix C for a more detailed description of the voting outcomes that we expect to result from equilibrium play.

at most one of the blue and/or one of the white players deviates from the equilibrium strategy. We call these realizations "almost" realizations. In Table 6, we regress the voting outcomes that are part of predicted equilibria on treatment dummies and additionally control for period effects. In all regressions, *NoChat* serves as baseline treatment. Additional results from Wald tests on treatment differences are presented in the bottom part of the table.

Result 1a (Voting outcomes given majority signal X) *The frequency ordering of A/A (A/C) is: D > TD, TDC > NoC (NoC > D, TD, TDC).*

As predicted in Hypothesis 1a, we find significantly more A/A-outcomes and significantly fewer A/C-outcomes after majority signal X in *Deliberation* and *TopDown* than in *NoChat*. This becomes already visible in the left part of Table 5 and is corroborated by the logit regressions (1) and (2) in Table 6. However, these regressions also reveal deviations from Hypothesis 1a: After majority signal X, there are significantly more A/A-outcomes and significantly fewer A/C-outcomes in *TopDownClosed* than in *NoChat*, where the predictions were no differences. Relatedly, *TopDownClosed* does not lead to significantly less (more) A/A -outcomes (A/C-outcomes) than *TopDown*, other than predicted. Moreover, there are significantly more A/A-outcomes in *Deliberation* than in *TopDown*, where the prediction, again, was no difference.

Result 1b (Voting outcomes given majority signal Y) *The frequency ordering of B/B (B/C) is: D > TD, TDC > NoC (D, TD, TDC > NoC).*

As predicted in Hypothesis 1b, after majority message Y, we find significantly more B/B-outcomes in *Deliberation* than in any other treatment. This can be seen in the right part of Table 5 and in Model (3) of Table 6. We also find that B/C occurs significantly more often in *TopDown* and *TopDownClosed* than in *NoChat*, again as predicted (see Model (4)). However, our regressions also reveal deviations, namely that after majority signal Y, B/B occurs significantly more often in *TopDown* and *TopDownCosed* than predicted, compared to *NoChat* (see Model (3)), and B/C occurs more often in *Deliberation* than predicted, namely not significantly less than in the other communication treatments (see Model (4)).

5.3 Whites' individual votes

Result 2a (Whites' voting decisions) *Given majority signal X (Y), the frequency ordering of whites' votes for A (B) is: D, TD^w, TDC > NoC (D, TD, TDC > NoC), where by the superscript ^{w} we denote (here and hereafter) weak statistical significance of treatment difference (0.5< p <0.1).*

Table 6: Voting outcomes

	Majority signal: X		Majority signal: Y	
	(1) A/A	(2) A/C	(3) B/B	(4) B/C
Deliberation	18.966***	−1.417***	16.835***	3.182***
	(0.492)	(0.448)	(0.737)	(0.944)
TopDown	17.791***	−1.554***	13.909***	3.678***
	(0.618)	(0.489)	(1.030)	(0.924)
TopDownClosed	17.262***	−1.048***	14.029***	3.512***
	(0.477)	(0.300)	(1.038)	(0.904)
Period	−0.178***	0.076***	−0.480***	0.044**
	(0.033)	(0.020)	(0.083)	(0.018)
Constant	−17.606***	−1.460***	−16.519***	−4.930***
	(0.339)	(0.324)	(0.555)	(0.949)
Wald test results for comparison of treatment coefficients (p values):				
D vs. TD	0.000	0.797	0.002	0.142
TD vs. TDC	0.389	0.231	0.922	0.403
D vs. TDC	0.000	0.326	0.006	0.248
Pseudo R^2	0.305	0.086	0.454	0.112
Number of clusters	20	20	20	20
Observations	817	817	743	743

Pooled logit regressions. Dependent variable: Realization of the respective voting outcome. Standard errors are clustered at the session level and given in parentheses: * p<0.10, ** p<0.05, *** p<0.01. NoChat serves as baseline treatment in all regressions.

Our Hypothesis 2a is fully confirmed: The logit regression results presented in Table 7 reveal that the whites' propensity to vote for the efficient policy A after majority signal X does not differ significantly between *Deliberation*, *TopDown*, and *TopDownClosed*. It is significantly higher in these treatments than in *NoChat* (see Model (1) and the respective Wald test results, where the difference between *TopDown* and *NoChat* is only significant at the 10% level). A similar picture emerges when we consider the whites' voting behavior given majority signal Y. As predicted by Hypothesis 2a, the whites' propensity to vote for the efficient policy B is not significantly different across the communication treatments. It is, however, significantly higher in these treatments compared to *NoChat* (see Model (5) and the respective Wald test results).

5.4 Blues' individual votes

In Table 8 we study the blue players' voting decisions in logit regressions that are analogous to the ones that we considered in Table 7.

Result 2b (Blues' voting decisions) *Given majority signal X (Y), the frequency ordering of blues' votes for A (B) is:* $D \overset{w}{>} TD > NoC$ *and* $D > TDC > NoC$ *(*$D > TD > TDC > NoC$*). Given majority signal X (Y), the frequency ordering of blues' votes for C is:* $D \overset{w}{<} TD < NoC$ *and* $D < TDC < NoC$ *(*$D \overset{w}{<} NoC$ *and* $D < TD$ *and* $TD < TDC$*).*

Table 7: Individual voting decisions of the whites

	Majority signal: X		Majority message: X		Majority signal: Y		Majority message: Y	
	(1) A vote	(2) B vote	(3) A vote	(4) B vote	(5) A vote	(6) B vote	(7) A vote	(8) B vote
Deliberation	2.050***	−2.030***	0.813	−0.655	−1.853***	1.840***	−0.371	0.315
	(0.541)	(0.546)	(0.860)	(0.905)	(0.256)	(0.259)	(0.339)	(0.339)
TopDown	1.391*	−1.544*			−2.090***	2.102***		
	(0.737)	(0.809)			(0.297)	(0.303)		
TopDownClosed	2.440***	−2.424***	−1.720**	1.850**	−1.986***	1.974***	−0.491	0.465
	(0.554)	(0.557)	(0.832)	(0.893)	(0.328)	(0.328)	(0.611)	(0.596)
Period	0.062**	−0.066***	−0.039**	0.037**	0.050***	−0.051***	0.021	−0.024*
	(0.025)	(0.023)	(0.015)	(0.016)	(0.012)	(0.013)	(0.013)	(0.014)
Constant	1.711***	−1.695***	4.198***	−4.323***	0.620***	−0.616***	−1.626***	1.663***
	(0.258)	(0.252)	(0.672)	(0.761)	(0.190)	(0.200)	(0.315)	(0.322)
Wald test results for comparison of treatment coefficients (p values):								
D vs. TDC	0.595	0.593	0.001	0.001	0.669	0.660	0.845	0.801
D vs. TD	0.459	0.609			0.397	0.344		
TD vs. TDC	0.237	0.353			0.764	0.707		
Pseudo R^2	0.115	0.119	0.121	0.123	0.124	0.124	0.009	0.009
Number of clusters	20	20	15	15	20	20	15	15
Observations	2451	2451	2112	2112	2229	2229	1278	1278

Pooled logit regressions. Dependent variable: Decision to vote for the respective policy. Standard errors are clustered at the session level and 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 in the other regressions.

As predicted in Hypothesis 2b, given majority signal X, the blues vote significantly more often for the efficient policy A in *Deliberation* than in *TopDownClosed* and *NoChat*, and more often in *TopDown* than in *NoChat* (see Model (1) in Table 8). We find a small deviation from the hypothesis in that in *Deliberation* blues vote weakly significantly more for A than in *TopDown*. If we focus on those periods in which the majority signal is Y (Model (5)), we observe that the blues' propensity to vote for the efficient policy B is significantly higher in *Deliberation* than in *TopDown*, *TopDownClosed* and *NoChat*, again as predicted in Hypothesis 2b. However, Table 8 also reveals deviations: Blues vote more often for A after majority signal X in *TopDownClosed*, compared to *NoChat*, and in *Deliberation*, compared to *TopDown* (Model (1)); and they vote more often for B after majority signal Y in *TopDown* than in *TopDownClosed*.

5.5 Whites' lying behavior

For the purpose of our data analysis, we define truthtelling (lying) as the subgroup of white players' reporting majority message Y (X) if, in fact, their majority signal was Y. This means that we consider only pivotal lies.[10] In Table 9, we report pooled logit regressions that are restricted to periods where the majority signal is Y. We regress the

[10]Considering the descriptive statistics (not reported in the tables), we find that 22.16% (11.44 %, 35.08) of white subgroups lie in *Deliberation* (*TopDown*, *TopDownClosed*). In addition, there are 12.04% of silent white subgroups in *TopDownClosed* and none in the other two communication treatments.

Table 8: Individual voting decisions of the blues

	Majority signal: X		Majority message: X		Majority signal: Y		Majority message: Y	
	(1) A vote	(2) C vote	(3) A vote	(4) C vote	(5) B vote	(6) C vote	(7) B vote	(8) C vote
Deliberation	2.688***	−2.016***	0.418	−0.429	2.227***	−0.780*	0.717***	−0.646**
	(0.260)	(0.398)	(0.280)	(0.298)	(0.406)	(0.468)	(0.247)	(0.272)
TopDown	2.237***	−1.533***			1.726***	−0.096		
	(0.279)	(0.400)			(0.327)	(0.422)		
TopDownClosed	1.892***	−1.273***	−0.297	0.238	1.174***	−0.559	0.002	−0.081
	(0.223)	(0.355)	(0.219)	(0.221)	(0.399)	(0.404)	(0.179)	(0.188)
Period	−0.085***	0.086***	−0.097***	0.104***	−0.109***	0.067***	−0.105***	0.094***
	(0.013)	(0.013)	(0.013)	(0.011)	(0.023)	(0.014)	(0.025)	(0.023)
Constant	−1.144***	0.366	1.189***	−1.330***	−2.430***	0.664*	−0.640***	0.503***
	(0.218)	(0.335)	(0.246)	(0.245)	(0.328)	(0.356)	(0.196)	(0.190)
Wald test results for comparison of treatment coefficients (p values):								
D vs. TDC	0.000	0.000	0.002	0.005	0.002	0.414	0.014	0.053
D vs. TD	0.081	0.075			0.048	0.024		
TD vs. TDC	0.106	0.208			0.023	0.010		
Pseudo R^2	0.169	0.120	0.065	0.069	0.115	0.041	0.072	0.059
Number of clusters	20	20	15	15	20	20	15	15
Observations	2451	2451	2112	2112	2229	2229	1278	1278

Pooled logit regressions. Dependent variable: Decision to vote for the respective policy. Standard errors are clustered at the session level and 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 in the other regressions.

dummy variable for a white sub-group reporting majority message Y (truthtelling) on dummies for treatments *Deliberation* and *TopDown* and interaction terms.

Result 3a (Whites' truthtelling) *The frequency ordering of instances in which the majority message equals the majority signal is:* $TD \overset{w}{>} D \overset{w}{>} TDC$ *and* $TD > TDC$.

In line with Hypothesis 3a, we find that truthtelling is significantly more pronounced in *Deliberation* and *TopDown* than in *TopDownClosed* with weak significance for *Deliberation*. However, we also find a small deviation: Truthtelling is weakly significantly more pronounced in *TopDown* than in *Deliberation* with a Wald test result of $p = 0.062$.

5.6 Blues' trustfulness

To test our Hypothesis 3b, we reconsider the regression results presented in Table 8.

Result 3b (Blues' trustfulness) *The frequency ordering of blues' votes for A (C) after majority message X is:* $D > TDC$ ($D < TDC$).

As predicted, the blues vote significantly more (less) often for A (C) after majority message X in *Deliberation* than in *TopDownClosed* (see Models (3) and (4)). However, we again find a deviation: We do not find the predicted treatment differences in blues'

Table 9: Truthtelling if the majority signal is Y

	Majority message: Y
Deliberation (D)	0.846*
	(0.441)
TopDown (TD)	1.636***
	(0.512)
Constant	0.410
	(0.372)
Wald test result for comparison of treatment coefficients (p value):	
D vs. TD	0.062
Pseudo R^2	0.068
Number of clusters	15
Observations	554

Pooled logit regressions. Data set restricted to periods where the majority signal is Y. Dependent variable: Reported majority message: Y (dummy). Standard errors are clustered at the session level and given in parentheses: * p<0.10, ** p<0.05, *** p<0.01. Note that the TopDownClosed treatment serves as baseline treatment.

votes for A and C between *TopDown* and *TopDownClosed*. After majority message X, blues also vote significantly more (less) often for A (C) in *TopDownClosed* than predicted, compared to *TopDown* (see, again, Models (3) and (4)). We can hence only partly validate Hypothesis 3b: Blue players are more trusting in *Deliberation* than in *TopDownClosed*, but they do not trust more in *TopDown* than in *TopDownClosed*. We call this phenomenon *flat (dis-)trust*.

5.7 Earnings and efficiency

Table 10 summarizes the predicted and actual (expected) joint earnings that are realized by the voting group and the color groups, respectively. Regressions of these expected joint earnings on treatment dummies are presented in Table 11.

Result 4a (Efficiency ranking) *The frequency ordering of total expected joint earnings is: D, TD, TDC > NoC.*

Result 4b (Earnings' rankings) *The frequency ordering of whites' (blues') expected earnings is: D, TD, TDC > NoC (TD, TDC < NoC and D* $\overset{w}{>}$ *TDC).*

For total and whites' expected joint earnings, our efficiency results are more (though also not fully) in line with our Hypothesis 4b, compared to blues' expected joint earnings. Comparing total expected earnings – our measure of efficiency – between the communication treatments in Table 10, we find that they are highest in *Deliberation* (80.98) and lowest in *TopDownClosed* (78.28), with *TopDown* (79.21) in between, as predicted. However, the Wald test results from Model (1) presented in the bottom part of Table 11

reveal that none of the communication-treatment comparisons in (voting-group or color-group) expected earnings is statistically significant. The only statistical differences are between the communication treatments and *NoChat*. Hence, *Deliberation* is less and *TopDownClosed* more efficient (and favorable to the whites) than predicted, compared to *TopDown*.

Table 10: Rankings over earnings

Efficiency		White players' joint earnings		Blue players' joint earnings	
Predicted expected total payoffs	Empirical outcome*	Predicted expected total payoffs	Empirical outcome*	Predicted expected payoffs	Empirical outcome*
D (85.56)	D (80.98, 28.53)	D (50.28)	D (42.16, 15.65)	TD (42.78)	NoC (39.33, 12.52)
TD (81.30)	TD (79.21, 27.06)	TD (38.52)	TD (41.42, 15.36)	TDC (40.14)	D (38.83, 12.53)]
TDC (65.28)	TDC (78.28, 27.41)	TDC (25.14)	TDC (41.04, 14.67)	NoC (37.5)	TD (37.79, 11.13)
NoC (60)	NoC (70.77, 33.26)	NoC (22.5)	NoC (31.44, 19.11)	D (35.28)	TDC (37.25, 12.39)

* The numbers in columns labeled "empirical outcomes" are average expected period earnings and their respective standard deviations (in points). They are calculated based on the players' types (white or blue), the signals that the computer reported to the whites, the conditional probabilities of the states of the world (each signal is true with 70% probability) and the actual votes in a given period. In case of a voting tie, the expected earnings are based on the probabilities with which the policies are implemented ($\frac{1}{3}$ in case there is a tie between three policies, $\frac{1}{2}$ in case there is a tie between two policies).

The blues earn in expectation 4.99 points less than predicted in *TopDown*, 2.89 points less than predicted in *TopDownClosed* and 3.55 points more than predicted in *Deliberation*. They also earn 3.72 points more than predicted in *NoChat*. This preserves the predicted ranking between *TopDown* and *TopDownClosed*, but reverses all other predicted rankings. The differences in blues' expected earnings across communication treatments are mostly not significant, with the exception of the difference between *Deliberation* and *TopDownClosed*. Hence, *Deliberation* generates higher expected earnings of the blues than predicted, compared to the other treatments. To conclude, although *Deliberation* is less efficient than predicted, compared to *TopDown*, it is still the socially most desirable treatment since in it the whites gain significantly and the blues are not hurt.

6 What we learn from the deviations

6.1 Systematic summary of all deviations

As already mentioned in section 5.7, *TopDownClosed* is more efficient than predicted and *Deliberation* is less efficient than predicted, compared to *TopDown*. In line with that, we can order all deviating findings (numbered from 1 to 10) according to two categories: deviations that we find when comparing *TopDownClosed* with other treatments, and deviations that we find when comparing *Deliberation* to other treatments.

Table 11: Expected period earnings – across treatments

	(1) All Players	(2) Whites	(3) Blues	(4) All Players
Deliberation	1.702***	3.572***	−0.168	−0.168
	(0.448)	(1.024)	(0.283)	(0.283)
TopDown	1.407***	3.327***	−0.512**	−0.512**
	(0.408)	(0.953)	(0.226)	(0.226)
TopDownClosed	1.252**	3.199***	−0.694**	−0.694**
	(0.442)	(0.938)	(0.285)	(0.285)
White player				−2.629**
				(1.034)
Deliberation × White player				3.740***
				(1.207)
TopDown × White player				3.839***
				(1.118)
TopDownClosed × White player				3.893***
				(1.068)
Constant	11.795***	10.481***	13.110***	13.110***
	(0.370)	(0.877)	(0.197)	(0.197)
Wald-test results for comparison of treatment coefficients (p values):				
Deliberation vs. TopDown	0.347	0.709	0.154	
TopDown vs. TopDownClosed	0.607	0.800	0.446	
Deliberation vs. TopDownClosed	0.213	0.557	0.085	
R^2	0.018	0.069	0.005	0.047
Number of clusters	20	20	20	20
Observations	9360	4680	4680	9360

Pooled OLS regressions. Dependent variable: Expected earnings (in points), conditional on received signals. Standard errors are clustered at the session level and given in parentheses: * p<0.10, ** p<0.05, *** p<0.01. NoChat serves as baseline treatment in all regressions.

TopDownClosed (1) The conflict outcome *A/C* under majority signal X is less frequent in *TopDownClosed* than even in *NoChat*, although it was predicted to be the other way around. (2) Comparing *TopDownClosed* with *TopDown*, *A/A*-outcomes (*A/C*-outcomes) are not significantly less (more) frequent in the former than in the latter treatment, other than predicted. (3) Relatedly, the blues trust the whites equally often in both treatments: They do not vote less often for A after majority message X in *TopDownClosed*, although they were predicted to do so. (4) In line with this, *TopDownClosed* is not significantly less efficient than *TopDown*, other than predicted. Actually, it is much more efficient than predicted (by 13 expected points; see Table 10, not tested for significance).

Deliberation t(5) After majority signal X, blues vote more often for A in *Deliberation* than in *TopDown* (Table 8, weakly significant; no longer significant if conditioning on majority message X). (6) In line with this, we find more *A/A* outcomes in *Deliberation* than in *TopDown* (Table 6), although no difference was predicted. (7) Outcome B/C after majority signal Y is *not* significantly less pronounced in *Deliberation* than in *TopDown* (Table 6), although it was predicted to be so. Note that this occurs despite the fact that under majority signal Y the blues vote significantly more often for B and significantly

less often for C in *Deliberation*, compared to *TopDown* – both as predicted (Table 8). (8) Moreover, whites do not lie equally little in *Deliberation* as in *TopDown*, other than predicted. On the contrary, they lie more (Table 9, weakly significant). Together, the two deviations (7) and (8) already indicate the reasons for the most important deviation that we find with respect to *Deliberation*: (9) It is not significantly more efficient than *TopDown*. In fact, it is less efficient than predicted (by 4.58 expected points; see Table 10, not tested for significance). The two subsections below investigate the reasons for the lower-than-expected efficiency of *Deliberation* in more detail. (10) Finally, the ranking of blues' earnings deviates from our prediction: The blues earn more in *Deliberation* and less in *TopDown* than predicted, so that their earnings do not differ significantly across these two treatments. The reasons will become evident from the subsection below.

Blues' (dis-)trust With the exception of deviations (7), (8), (9), and (10), all deviations mentioned above reduce to *flat (dis-)trust* of the blues: The blues condition their trust in whites' truthtelling (i.e., voting for A after majority message – or signal – X) insufficiently (and wrongly) on the deliberative structure: They seem to expect whites to be equally or even a little more honest in *Deliberation*, compared to *TopDown*, although there is a significant difference in the opposite direction to what they expect. Moreover, they do not account for how much *TopDownClosed* induces the whites to lie, compared to *TopDown*. The former deviation yields the higher-than-predicted frequency of blues' A-votes and total A/A-otcomes in *Deliberation*, compared to *TopDown*, after majority signal X. The latter deviation yields the surprisingly high relative efficiency of *TopDown-Closed*, and the blues' behavior leading to it. In what follows, we explain deviations (7) and (8) in more detail and conduct additional analyses to further investigate the reasons behind deviations (9) and (10).

6.2 Having a voice enables blues to better coordinate among themselves

At first sight, the finding that *Deliberation* turns out not to be more efficient than *TopDown* seems to fit uneasily with how voting behavior differs between the two treatments: Blues vote less often for C and more often for A in *Deliberation* than in *TopDown*. Voters coordinate more often on A/A and less often on A/C in *Deliberation*, when the majority signal is X, and not more often on B/C when the majority signal is Y. So why is *Deliberation* not more efficient than *TopDown*? And why do blues earn more in the former than in the latter treatment (though not significantly so)?

To anticipate the answer, consider coordination on voting for a particular policy among the blues. First, note that the possibility of blues talking to each other in *Deliberation*

facilitates coordination between them. Second, note that, if already all whites perfectly coordinate on the efficient policy (say, A), the blues' perfect (rather than only imperfect) coordination on it does no longer affect policy implementation. By contrast, perfect rather than imperfect coordination of the blues on policy C reduces the probability of implementing the efficient policy *by half* if all whites are coordinated on a different policy. These asymmetric effects of coordination among the blues already constitute a disadvantage of *Deliberation* in terms of efficiency, compared to the other communication treatments. Now remember that B/C does not occur significantly less frequently in *Deliberation* than in *TopDown*, although blues vote less often for C in the former treatment. Together, this already indicates that C-votes are more often coordinated in *Deliberation* than in *TopDown*, which has the detrimental effect mentioned above.

To investigate this further, consider the Almost A/C and the Almost B/C outcomes in Tables 5 and 12. Remember that Almost-outcomes are voting outcomes from which at most one player per color group deviates. Hence, the category of Almost A/C (Almost B/C) comprises different voting outcomes that induce different policy implementations, namely A, or C (B, or C), or a tie between the two. In Table 12, we disaggregate the Almost A/C and the Almost B/C outcomes, depending on which policy implementation they induce. Almost A/C 1 (Almost B/C 1) induces the efficient policy A (B); and Almost A/C 2 (Almost B/C 2) induces the inefficient policy C. The residual category induces a tie.

Consider first the Almost A/C outcomes. As can be seen from Table 12, if the majority signal is X, Almost A/C occurs (weakly) significantly more often in *TopDown* than in *Deliberation*. However, this is entirely due to Almost A/C 1 which leads to the efficient policy A and not to Almost A/C 2 which leads to C. In Almost A/C 1, two blues vote for C; while three blues vote for C in Almost A/C 2. Hence, the relatively more frequent C-votes of the blues in *TopDown* under majority message X are less coordinated, compared to *Deliberation*, and in such extent that efficiency is affected positively, if at all (see Table A.1).

Next, consider the Almost B/C outcomes under majority signal Y. As Table 12 reveals, they are not significantly different between *Deliberation* and *TopDown* on the aggregate level. However, Almost B/C 1 which yields the efficient policy B occurs significantly more often in *TopDown* than in *Deliberation*, whereas there is no difference with respect to Almost B/C 2 which yields C. Again, this reveals the lower frequency of perfectly coordinated C-votes of blues in *TopDown*, compared to *Deliberation*. Moreover, the two subcategories of Almost B/C 1 affect efficiency positively (Table A.1, model 3).

In sum, the aggregate information that *Deliberation* has less "Almost" conflict outcomes than *TopDown* (significant in the case of A/C) is misleading: The only "Almost" conflict outcomes that occur *significantly* less often in *Deliberation* than in *TopDown* are

those of category 1, i.e., those that lead to the efficient policies A or B. This reveals the better coordination of blues' C-votes in *Deliberation*, compared to *TopDown*, and hence contributes to the explanation of why the former treatment is not more efficient than the latter and why the blues earn no more in *TopDown* than in *Deliberation*, although they place more selfish votes in *TopDown*.

Table 12: Frequencies of voting outcomes at the group level conditional on the received majority signal – further details (For the full table, see Table A.2)

	Majority signal: X		Majority signal: Y	
	Deliberation	TopDown	Deliberation	TopDown
Almost A/C outcome	0.181	0.307	0.086	0.095
Almost A/C outcome 1	0.181	0.285	0.043	0.065
Almost A/C outcome 2	0.000	0.000	0.043	0.025
Almost B/C outcome	0	0.006	0.238	0.328
Almost B/C outcome 1	0.000	0.000	0.162	0.299
Almost B/C outcome 2	0.000	0.006	0.049	0.025
Mann-Whitney ranksum test results:				
Deliberation vs. TopDown				
Almost A/C outcome	$p = 0.076$		$p = 0.916$	
Almost A/C outcome 1	$p = 0.076$		$p = 0.402$	
Almost A/C outcome 2	$p = 0.317$		$p = 0.245$	
Deliberation vs. TopDown				
Almost B/C outcome			$p = 0.175$	
Almost B/C outcome 1			$p = 0.047$	
Almost B/C outcome 2			$p = 0.401$	

Note: In "Almost" outcomes at most one player per color group deviates from the respective outcome.
Outcome definitions:
Almost A/C outcome 1: either 3 whites vote for A, 2 blues vote for C or 2 whites vote for A, 1 white votes for B, 2 blues vote for C, 1 blue votes for A
Almost B/C outcome 1: either 3 whites vote for B, 2 blues vote for C or 2 whites vote for B, 1 white votes for A, 2 blues vote for C, 1 blue votes for B
Almost A/C (B/C) outcome 2: 2 whites vote for A (B), 3 blues vote for C

6.3 The dynamics reveal deterioration of deliberation

In this section we study the dynamics of behavior. The significant *Period* coefficients in the regressions from Table 6 as well as summary statistics presented in Table SOM.1 and Table SOM.2 show that the incidences of conflict outcomes (A/C and B/C) increases in all four treatments over time. In the communication treatments, after an initial phase of high cooperation and low conflict, the opportunity to chat does not lead to *sustainable* coordination on the efficient A/A (B/B) outcome in case the received majority signal is X (Y). Instead, if the majority signal indicates state X, more and more often A/C is realized in later periods. If the majority signal indicates Y, we increasingly often observe the B/C outcome. The effectiveness of deliberative democracy hence seems to deteriorate over time. This very likely leads to a floor effect, i.e., the measurable differences between the communication treatments decline with time, which contributes to explaining why we find fewer significant efficiency differences between them.

We now move on to analyzing how the interaction between blues' voting decisions, whites' lying behavior and the content of chat conversations leads to changes in behavior over time.[11] Table 13 shows blues' voting decisions in the communication treatments as a function of a number of chat categories. In our analysis we focus on chat classifications that we observed in more than 15% of all chat messages. Consider, first, the left part of the table 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. In logit regressions (1) and (2) we regress the blues' propensity to vote for A and C, respectively, on treatment dummies for *Deliberation* and *TopDownClosed* (*TopDown* serves as baseline treatment in these regressions), 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 and disrespectful language). Moreover, we include four additional dummy variables that capture whether the whites mention the 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 add a control variable for the period of play.

Interestingly, under majority message X we find significant negative but no positive effects of whites' chat-message content and tone on blues' trust (voting for A after majority message X). Conversely, there are only positive but no negative effects of content and tone on blues' distrust (voting for C after majority message X). To be precise, voting for A (voting for C) after majority message X is significantly less likely (more 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. By contrast, whites' justifying themselves by referring to the experimental environment (e.g., stating that wrong messages are 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 A (C) after majority signal X.

Next we turn to those periods in which the whites report majority message Y. It is well understood that this majority message is not a lie, since, given the monetary payoffs, the whites have no incentive to make the blues believe that Y prevails if in fact they believe that it is X. Hence, what we study here are not the effects on trust of the blues but on their cooperativeness, i.e., their B-votes, and on their uncooperativeness, i.e., their

[11] A full list of the dimensions in which the chat messages were coded can be found in Appendix B. Two research assistants coded the chat messages independently from each other (we refer to them as Coder #1 and Coder #2). In the regressions presented in the main part of this paper, we rely on the work done by Coder #1. All significant results presented in Table 13 and Table 14 are similarly found when relying on the codings of Coder #2 instead, see Tables A.3 and A.4 in Appendix A.

Table 13: Communication treatments: Individual voting decisions of the blues

	Majority message: X		Majority message: Y	
	(1) A vote	(2) C vote	(3) B vote	(4) C vote
Deliberation	0.528*	−0.555*	0.971***	−0.823***
	(0.277)	(0.290)	(0.313)	(0.305)
TopDownClosed	−0.298	0.213	−0.037	−0.044
	(0.205)	(0.213)	(0.209)	(0.203)
Potential lie in previous period	−0.224**	0.289***	−0.089	0.146
	(0.102)	(0.100)	(0.255)	(0.189)
Respectful whites	0.079	−0.085	0.554***	−0.587***
	(0.134)	(0.136)	(0.201)	(0.179)
Disrespectful whites	−0.583***	0.596***	−1.149***	0.914***
	(0.143)	(0.143)	(0.326)	(0.338)
Whites mention the experimental environment as information	0.037	−0.033	−0.156	0.269
	(0.079)	(0.080)	(0.261)	(0.196)
Whites mention the experimental environment to justify their behavior	0.102	−0.117	−0.241	0.002
	(0.194)	(0.191)	(0.275)	(0.268)
Whites mention the public spirit	0.081	−0.080	0.059	−0.042
	(0.157)	(0.146)	(0.181)	(0.177)
Whites mention whites' and blues' joint payoffs	0.071	−0.083	0.229	−0.022
	(0.157)	(0.143)	(0.216)	(0.272)
Period	−0.088***	0.092***	−0.089***	0.079***
	(0.014)	(0.013)	(0.026)	(0.022)
Constant	1.110***	−1.227***	−0.754***	0.564***
	(0.261)	(0.267)	(0.235)	(0.208)
Wald test results for comparison of treatment coefficients (p values):				
D vs. TDC	0.001	0.001	0.008	0.018
Pseudo R^2	0.061	0.065	0.081	0.068
Number of clusters	15	15	15	15
Observations	2001	2001	1230	1230

Pooled logit regressions. Dependent variable: Decision to vote for the respective policy. Standard errors are clustered at the session level and given in parentheses: * p<0.10, ** p<0.05, *** p<0.01. TopDown serves as baseline treatment in all 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.

C-votes, given majority message Y. For this analysis we regress the blues' propensity to vote for B (model (3)) and C (model (4)) on the same explanatory variables that we used in models (1) and (2). As evident, if the reported majority message is Y, voting for B (voting for C) does not depend on the perceived correctness of the previous state of the world (see the insignificant coefficient of "Potential lie"). However, disrespectfulness is again effective: Voting for B (voting for C) is on average less likely (more likely) if the whites treat the blues disrespectfully. Moreover, treating the blues respectfully now has the opposite effect, potentially reinforcing the general positive effect of telling the truth to the blues. 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. Lastly, also for majority message Y, voting for B (voting for C) is on average more likely in earlier periods (in later periods).

To summarize: If the majority message is X, blues vote for C more often when they suspect having been lied to in the previous period and when being treated disrespectfully. This behavior could be considered both an indication for blue players' general distrust in the reported message or a desire for revenge or spitefulness. Suspecting having been lied to in the previous period has less of an effect on the blues' voting decisions if the reported majority message is Y. The impact of disrespectful language is, however, still sizable.

We have seen that when blues discover that the state of the world was Y, although the majority message they received from the whites was X (a potential lie), they significantly move their votes away from A and towards C, when in the next period they receive majority singal X. The question arises why the whites lie to the blues and – considering that they do so also in the public chat in *Deliberation* and *TopDown* – why they do it even at the expense of lying to their fellow whites. The regression specifications in Table 14 attempt to shed light on this question. In the reported logit specification we regress the individual white players' decisions to lie on all chat content categories that were recorded in at least 15% of the blues' chat messages in *Deliberation*. We also include a dummy that takes the value 1 if all blue players that a white player was matched to in the previous period voted for C in that period. Furthermore, we 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 and a variable capturing the periods.

Model (1) considers only the *Deliberation* treatment. As evident, the whites' propensity to report a wrong majority message (report X instead of Y) increases if they encountered at least one blue player who recommended voting for C and if all blue players voted for C in the previous period. If, however, a blue recommended voting for B in the previous period, the whites' propensity to lie decreases on average. Also, the more successful a lie was in the previous period (measured as number of convinced blues), the higher is a white's propensity to lie again.

When considering all communication treatments (see model (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*. Nevertheless, we observe similar behavioral patterns: The whites' propensity to lie increases in the number of blue players who voted for A following a lie in the previous period and it increases if all blue players voted for C in the previous period.

To conclude, the dynamics of communication reveal a general deterioration of blues' cooperativeness and trust and whites' truthfulness in all communication treatments. Due to the floor effect, this general deterioration potentially reduces the measurable differences between the communication treatments. Moreover, in *Deliberation* there is an additional factor that enforces deterioration, namely disrespectful language of the whites and uncooperative talk of the blues. This also contributes to explaining why *Deliberation* does not

Table 14: Communication treatments: Lying decisions of the whites, conditional on receiving signal Y

	Only Deliberation treatment	All communication treatments
	(1)	(2)
Suspicious blue in previous period	−0.322	
	(0.288)	
Blue recommended voting for A in previous period	0.376	
	(0.266)	
Blue recommended voting for B in previous period	−0.489*	
	(0.264)	
Blue recommended voting for C in previous period	0.633**	
	(0.291)	
Disrespectful blue in previous period	−0.097	
	(0.313)	
All blues voted for C in previous period	0.394***	0.275*
	(0.054)	(0.166)
# convinced blues in previous lie	0.359***	0.458***
	(0.115)	(0.094)
Period	0.035**	0.077***
	(0.016)	(0.014)
Constant	−2.232***	−2.679***
	(0.438)	(0.287)
Pseudo R^2	0.052	0.039
Number of clusters	5	15
Observations	545	1555

Pooled logit regressions. Dependent variable: Decision to lie. Standard errors are clustered at the session level and given in parentheses: * p<0.10, ** p<0.05, *** p<0.01. In Model (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.

turn out more efficient than *TopDown*.

7 Discussion

We use a laboratory experiment to shed light on an important socio-economic issue: The difficulty of reaching an efficient outcome in a democratic environment in which two social groups with different material interests have also different information and different access to communication channels. We propose a theoretical model and derive formal hypotheses about comparative statics, which we test in the laboratory. Both the findings consistent with our hypotheses and the deviations teach us important lessons.

Compared with the setting without any communication, we find that communication leads to efficiency gains. However, for all three deliberative structures most efficiency gains go to the whites, not only, as predicted, for *Deliberation*. Hence, our findings suggest that

in material terms all types of societal communication ultimately serve the well informed.

This pattern of deviations of expected earnings from the prediction can be traced back to *Deliberation* leading to substantially lower efficiency than predicted and *TopDownClosed* to substantially higher efficiency than predicted, with *TopDown* remaining in the middle. This is largely the consequence of blues' notable voting behavior in *Deliberation*. They either coordinate on cooperating with the whites or they coordinate on conflict against the whites. The fact that in this deliberative structure blues can express themselves and, in particular, communicate with fellow group members facilitates coordination, either way. Coordinating on conflict strongly lowers whites' earnings and hurts efficiency, but it protects the blues to some extent.

Studying the dynamics of the chat and voting behavior, we find that the interaction between deliberative structures and color is not the only relevant factor: Potential lies interact with the dynamics of our experimental setting in a way that affects outcomes. Given the information structure of our environment, it is both possible that the whites lie to the blues and also that whites seem to lie, although they do not. Blues detect a potential lie when at the end of a period they find out that there is a discrepancy between what the whites told them and the true state of the world. A potential lie naturally increases political conflict between whites and blues. The dynamics of the chat and voting behavior in *Deliberation* reveals the existence of a vicious circle: A blue recommends an egoistic vote to the other blues. In reaction, more whites tend to lie to the blues in the next period. This tends to increase the discrepancy between announced and ex-post observed state of the world. In reaction, more blues recommend the egoistic vote to the other blues. The good news is that unrestricted communication also allows for a virtuous circle that, although less prevalent than the vicious circle, also occurs in *Deliberation*: A blue recommends the efficient, not the egoistic, vote, to the other blues; the whites tend to lie less in the next period, and the blues are less likely to observe a potential lie. Hence, they tend to be more trustful and recommend the efficient vote again. However, 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 *Deliberation* treatment may lead to an escalation in animosity.

We believe that the phenomena we observe are relevant beyond our experiment. First, in unequal societies communication between groups increases efficiency but mostly favors the informed. Second, 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 (flat) 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. Third, with free communication the immediacy and anonymity of 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.

References

Akerlof, G. A., and R. E. Kranton (2000): "Economics and identity," *The Quarterly Journal of Economics*, 115(3), 715–753.

——— (2010): *Identity Economics, How Our Identities Shape Our Work, Wages, and Well-Being*. Princeton University Press.

Ambrus, A., E. M. Azevedo, and Y. Kamada (2013): "Hierarchical cheap talk," *Theoretical Economics*, 8(1), 233–261.

Apffelstaedt, A., and L. Mechtenberg (2018): "Competition over context-sensitive consumers," Discussion paper, Working Paper, University of Hamburg.

Benoît, J.-P., and J. Dubra (2014): "A theory of rational attitude polarization," *Working Paper, Social Science Research Network*.

Bock, O., I. Baetge, and A. Nicklisch (2014): "hroot: Hamburg registration and organization online tool," *European Economic Review*, 71, 117–120.

Borgonovi, F., B. d'Hombres, and B. Hoskins (2010): "Voter turnout, information acquisition and education: Evidence from 15 European countries," *The BE Journal of Economic Analysis & Policy*, 10(1).

Bowles, S., and S. Polanía-Reyes (2012): "Economic incentives and social preferences: substitutes or complements?," *Journal of Economic Literature*, 50(2), 368–425.

Brandts, J., and D. J. Cooper (2007): "It's 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.

Buechel, B., and L. Mechtenberg (2017): "The swing voter's curse in social networks," *Working Paper, University of Hamburg*.

Chen, R., and Y. Chen (2011): "The potential of social identity for equilibrium selection," *American Economic Review*, 101(6), 2562–2589.

Chen, Y., and S. X. Li (2009): "Group identity and social preferences," *The American Economic Review*, 99(1), 431–457.

Cohen, J. (1989): "The good polity," pp. 67–92.

Cooper, D. J., and J. H. Kagel (2005): "Are two heads better than one? Team versus individual play in signaling games," *American Economic Review*, 95(3), 477–509.

Dawes, R. M., A. J. Van de Kragt, and J. M. Orbell (1990): "Cooperation for the benefit of us: Not me, or my conscience," in *Beyond Self-Interest*, ed. by J. Mansbridge, pp. 97–110. The University of Chicago Press.

Dryzek, J. S., and C. List (2003): "Social choice theory and deliberative democracy: a reconciliation," *British Journal of Political Science*, 33(1), 1–28.

Feddersen, T. J., and W. Pesendorfer (1996): "The swing voter's curse," *American Economic Review*, pp. 408–424.

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

Goeree, J. K., and L. Yariv (2011): "An experimental study of collective deliberation," *Econometrica*, 79(3), 893–921.

Guarnaschelli, S., R. D. McKelvey, and T. R. Palfrey (2000): "An experimental study of jury decision rules," *American Political Science Review*, 94(2), 407–423.

Gutmann, A., and D. Thompson (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, J. (2015): *Between facts and norms: Contributions to a discourse theory of law and democracy*. Cambridge, MA: MIT Press.

Karpowitz, C. F., and T. Mendelberg (2011): "An experimental approach to citizen deliberation," in *Cambridge Handbook of Experimental Political Science*, ed. by J. N. Druckman, D. P. Green, J. H. Kuklinski, and A. Lupia, pp. 258–272. Cambridge University Press Cambridge.

Landwehr, C. (2010): "Discourse and Coordination: Modes of Interaction and their Roles in Political Decision-Making," *Journal of Political Philosophy*, 18(1), 101–122.

Morton, R. B., and J.-R. Tyran (2011): "Let the experts decide? Asymmetric information, abstention, and coordination in standing committees," *Games and Economic Behavior*, 72(2), 485–509.

Myers, C. D., and T. Mendelberg (2013): "Political deliberation," in *Oxford Handbook of Political Psychology*, ed. by D. S. Leonie Huddy, and J. Levy, pp. 699–734. Oxford University Press.

Orbell, J. M., A. J. Van de Kragt, and R. M. Dawes (1988): "Explaining discussion-induced cooperation.," *Journal of Personality and Social Psychology*, 54(5), 811–819.

Palfrey, T. R. (2016): "Experiments in political economy," in *Handbook of Experimental Economics*, ed. by A.Roth, and J. H.Kagel, vol. 2, pp. 347–434. Princeton University Press.

Palfrey, T. R., and K. Pogorelskiy (2017): "Communication Among Voters Benefits the Majority Party," *Economic Journal*, 129(618), 961–990.

Pande, R. (2011): "Can informed voters enforce better governance? Experiments in low-income democracies," *Annual Review of Economics*, 3, 215–237.

Pronin, K., and J. Woon (2017): "Public Deliberation, Private Communication, and Collective Choice," *Working Paper, New York University*.

Robalo, P., A. Schram, and J. Sonnemans (2017): "Other-Regarding Preferences, In-Group Bias and Political Participation: An Experiment," *Journal of Economic Psychology*, 62, 130–154.

Schotter, A. (2015): "On the relationship between economic theory and experiments," in *The Handbook of Experimental Economic Methodology*, ed. by G. .Frechette, and A. Schotter, pp. 58–85. Oxford University Press.

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

Appendix

A Additional Tables

Table A.1: Expected period earnings – across treatments

	(1) All Players	(2) All Players	(3) All Players	(4) Whites	(5) Whites	(6) Whites	(7) Blues	(8) Blues	(9) Blues
Deliberation	1.702***	1.611***	1.415***	3.572***	2.702***	2.643***	−0.168	0.521**	0.187
	(0.448)	(0.402)	(0.387)	(1.024)	(0.844)	(0.818)	(0.283)	(0.191)	(0.186)
TopDown	1.407***	1.200***	1.217***	3.327***	1.822**	1.964**	−0.512**	0.579**	0.470**
	(0.408)	(0.350)	(0.370)	(0.953)	(0.747)	(0.752)	(0.226)	(0.204)	(0.191)
TopDownClosed	1.252**	0.940**	0.960***	3.199***	1.761**	1.930***	−0.694**	0.118	−0.009
	(0.442)	(0.344)	(0.328)	(0.938)	(0.676)	(0.674)	(0.285)	(0.207)	(0.152)
Almost A/C 1.1		2.675***	2.096***		4.437***	3.236***		0.913**	0.956***
		(0.386)	(0.475)		(0.580)	(0.691)		(0.329)	(0.295)
Almost A/C 1.2		−0.210	1.562***		2.940***	3.397***		−3.361***	−0.273
		(0.622)	(0.415)		(0.771)	(0.719)		(0.657)	(0.389)
Almost A/C 2		−4.219***	−7.762***		−11.718***	−13.367***		3.279***	−2.158***
		(0.357)	(0.453)		(0.743)	(0.829)		(0.453)	(0.240)
Almost B/C 1.1		−0.667**	1.923***		3.383***	6.027***		−4.716***	−2.181***
		(0.314)	(0.258)		(0.502)	(0.508)		(0.194)	(0.282)
Almost B/C 1.2		−1.551***	1.051**		2.138***	4.789***		−5.240***	−2.686***
		(0.413)	(0.380)		(0.633)	(0.658)		(0.233)	(0.292)
Almost B/C 2		−3.962***	−1.951***		−12.544***	−10.454***		4.621***	6.552***
		(0.318)	(0.458)		(0.639)	(0.560)		(0.443)	(0.788)
Majority signal: Y			−4.783***			−4.927***			−4.639***
			(0.242)			(0.384)			(0.343)
Almost A/C 1.1 × Majority signal: Y			−3.129***			−0.168			−6.089***
			(0.407)			(0.492)			(0.450)
Almost A/C1.2 × Majority signal: Y			−2.354***			0.149			−4.857***
			(0.436)			(0.456)			(0.648)
Almost A/C 2 × Majority signal: Y			6.802***			4.093***			9.511***
			(0.296)			(0.379)			(0.449)
Constant	11.795***	11.818***	14.041***	10.481***	11.112***	13.291***	13.110***	12.524***	14.790***
	(0.370)	(0.360)	(0.458)	(0.877)	(0.777)	(0.830)	(0.197)	(0.124)	(0.173)
Wald test results for comparison of treatment coefficients (p values):									
D vs. TDC	0.2130	0.0447	0.0229	0.5571	0.0895	0.1083	0.0848	0.1197	0.2512
D vs. TD	0.3472	0.2164	0.4298	0.7088	0.1643	0.2134	0.1538	0.8140	0.1803
TD vs. TDC	0.6072	0.3139	0.1597	0.8002	0.8805	0.9177	0.4463	0.0666	0.0051
R^2	0.018	0.122	0.365	0.069	0.516	0.668	0.005	0.194	0.655
Number of clusters	20	20	20	20	20	20	20	20	20
Observations	9360	9360	9360	4680	4680	4680	4680	4680	4680

Pooled OLS regressions. Dependent variable: Expected earnings, conditional on received signals. Standard errors are clustered at the session level and given in parentheses: * p<0.10, ** p<0.05, *** p<0.01. NoChat serves as baseline treatment in all regressions.

Table A.2: Frequencies of voting outcomes at the group level conditional on the received majority signal – Further details

	Majority signal: X				Majority signal: Y			
	NoChat	Deliberation	TopDown	TopDownClosed	NoChat	Deliberation	TopDown	TopDownClosed
Almost A/C outcome	0.513	0.181	0.307	0.339	0.476	0.086	0.095	0.126
Almost A/C outcome 1	0.368	0.181	0.285	0.339	0.241	0.043	0.065	0.084
Almost A/C outcome 1.1	0.321	0.172	0.274	0.339	0.163	0.043	0.060	0.058
Almost A/C outcome 1.2	0.047	0.009	0.011	0.000	0.078	0.000	0.005	0.026
Almost A/C outcome 2	0.107	0.000	0.000	0.000	0.181	0.043	0.025	0.042
Almost B/C outcome	0.013	0	0.006	0	0.090	0.238	0.328	0.257
Almost B/C outcome 1	0.000	0.000	0.000	0.000	0.012	0.162	0.299	0.209
Almost B/C outcome 1.1	0.000	0.000	0.000	0.000	0.012	0.141	0.279	0.178
Almost B/C outcome 1.2	0.000	0.000	0.000	0.000	0.000	0.022	0.020	0.031
Almost B/C outcome 2	0.013	0.000	0.006	0.000	0.048	0.049	0.025	0.031
Mann-Whitney ranksum test results:								
Deliberation vs. TopDown								
Almost A/C outcome			$p = 0.076$				$p = 0.916$	
Almost A/C outcome 1			$p = 0.076$				$p = 0.402$	
Almost A/C outcome 1.1			$p = 0.047$				$p = 0.402$	
Almost A/C outcome 1.2			$p = 0.699$				$p = 0.317$	
Almost A/C outcome 2			$p = 0.317$				$p = 0.245$	
Deliberation vs. TopDownClosed								
Almost A/C outcome			$p = 0.076$				$p = 0.347$	
Almost A/C outcome 1			$p = 0.076$				$p = 0.117$	
Almost A/C outcome 1.1			$p = 0.028$				$p = 0.295$	
Almost A/C outcome 1.2			$p = 0.134$				$p = 0.054$	
Almost A/C outcome 2			-				$p = 1$	
TopDown vs. TopDownClosed								
Almost A/C outcome			$p = 0.530$				$p = 0.754$	
Almost A/C outcome 1			$p = 0.295$				$p = 0.754$	
Almost A/C outcome 1.1			$p = 0.245$				$p = 0.754$	
Almost A/C outcome 1.2			$p = 0.317$				$p = 0.196$	
Almost A/C outcome 2			$p = 0.317$				$p = 0.245$	
Deliberation vs. TopDown								
Almost B/C outcome							$p = 0.175$	
Almost B/C outcome 1							$p = 0.047$	
Almost B/C outcome 1.1							$p = 0.047$	
Almost B/C outcome 1.2							$p = 0.914$	
Almost B/C outcome 2							$p = 0.401$	
Deliberation vs. TopDownClosed								
Almost B/C outcome							$p = 0.754$	
Almost B/C outcome 1							$p = 0.347$	
Almost B/C outcome 1.1							$p = 0.602$	
Almost B/C outcome 1.2							$p = 0.523$	
Almost B/C outcome 2							$p = 0.600$	
TopDown vs. TopDownClosed								
Almost B/C outcome							$p = 0.251$	
Almost B/C outcome 1							$p = 0.076$	
Almost B/C outcome 1.1							$p = 0.076$	
Almost B/C outcome 1.2							$p = 0.459$	
Almost B/C outcome 2							$p = 0.834$	

Note: In "Almost" outcomes at most one player per color group deviates from the respective outcome.
Outcome definitions:
Almost A/C (B/C) outcome 1.1: 3 whites vote for A (B), 2 blues vote for C
Almost A/C (B/C) outcome 1.2: 2 whites vote for A (B), 1 white votes for B (A), 2 blues vote for C, 1 blue votes for A (B)
Almost A/C (B/C) outcome 1: either Almost A/C (B/C) outcome 1.1 or Almost A/C (B/C) outcome 1.2
Almost A/C (B/C) outcome 2: 2 whites vote for A (B), 3 blues vote for C

Table A.3: Communication treatments: Individual voting decisions of the blues - Regressions based on rating done by Coder #2

	Majority message: X		Majority message: Y	
	(1) A vote	(2) C vote	(3) B vote	(4) C vote
Deliberation	0.672**	−0.712**	1.014***	−0.885***
	(0.301)	(0.318)	(0.274)	(0.279)
TopDownClosed	−0.279	0.198	−0.154	0.085
	(0.219)	(0.228)	(0.213)	(0.201)
Potential lie in previous period	−0.213**	0.277***	−0.074	0.142
	(0.097)	(0.095)	(0.233)	(0.175)
Respectful whites	0.239	−0.241	1.453***	−1.679***
	(0.246)	(0.244)	(0.358)	(0.404)
Disrespectful whites	−0.738***	0.768***	−1.413***	1.346***
	(0.268)	(0.291)	(0.429)	(0.361)
Whites mention the experimental environment as information	0.150	−0.142	0.033	0.187
	(0.131)	(0.133)	(0.208)	(0.226)
Whites mention the experimental environment to justify their behavior	−0.018	0.028	−0.552**	0.573***
	(0.122)	(0.125)	(0.257)	(0.206)
Whites mention the public spirit	−0.317*	0.362**	−0.454***	0.494***
	(0.179)	(0.168)	(0.170)	(0.186)
Whites mention whites' and blues' joint payoffs	0.273*	−0.351**	0.848***	−0.695***
	(0.158)	(0.155)	(0.213)	(0.223)
Period	−0.088***	0.093***	−0.092***	0.083***
	(0.013)	(0.013)	(0.026)	(0.021)
Constant	1.113***	−1.234***	−0.817***	0.517***
	(0.278)	(0.283)	(0.202)	(0.178)
Wald test results for comparison of treatment coefficients (p values):				
D vs. TDC	0.000	0.000	0.000	0.001
Pseudo R^2	0.065	0.069	0.095	0.087
Number of clusters	15	15	15	15
Observations	2001	2001	1230	1230

Pooled logit regressions. Dependent variable: Decision to vote for the respective policy. Standard errors are clustered at the session level and given in parentheses: * p<0.10, ** p<0.05, *** p<0.01. TopDown serves as baseline treatment in all 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 A.4: Communication treatments: Lying decisions of the whites, conditional on receiving signal Y - Regressions based on rating done by Coder #2

	Only Deliberation treatment	All communication treatments
	(1)	(2)
Suspicious blue in previous period	−0.006	
	(0.387)	
Blue recommended voting for A in previous period	0.251	
	(0.172)	
Blue recommended voting for B in previous period	−0.533**	
	(0.255)	
Blue recommended voting for C in previous period	0.517*	
	(0.284)	
Disrespectful blue in previous period	0.458**	
	(0.212)	
All blues voted for C in previous period	0.280***	0.275*
	(0.102)	(0.166)
# convinced blues in previous lie	0.368***	0.458***
	(0.108)	(0.094)
Period	0.031**	0.077***
	(0.014)	(0.014)
Constant	−2.269***	−2.679***
	(0.468)	(0.287)
Pseudo R^2	0.056	0.039
Number of clusters	5	15
Observations	545	1555

Pooled logit regressions. Dependent variable: Decision to lie. Standard errors are clustered at the session level and given in parentheses: * p<0.10, ** p<0.05, *** p<0.01. In Model (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.

B Chat dimensions

Dimension 1 – General classification of chats

Table B.1: Rater 1: Fraction of groups in which the whites...

	Deliberation	TopDown	TopDownClosed
... stress the public spirit	0.22	0.22	0.23
... stress the interests of the own color group	0.11	0.12	0.01
... suspect lying	0.03	0.01	0.01
... stress trust	0.01	0.01	0.02
... mention circumstances as justification for behavior	0.22	0.14	0.10
... mention circumstances as information	0.27	0.28	0.21
... stress hope or optimism	0.07	0.13	0.10

Table B.2: Rater 1: Fraction of groups in which the blues...

	Deliberation
... stress the public spirit	0.09
... stress the interests of the own color group	0.13
... suspect lying	0.16
... stress trust	0.03
... mention circumstances as justification for behavior	0.13
... mention circumstances as information	0.23
... stress hope or optimism	0.04

Table B.3: Rater 2: Fraction of groups in which the whites...

	Deliberation	TopDown	TopDownClosed
... stress the public spirit	0.25	0.24	0.27
... stress the interests of the own color group	0.08	0.03	0.00
... suspect lying	0.02	0.00	0.00
... stress trust	0.02	0.00	0.02
... mention circumstances as justification for behavior	0.13	0.16	0.03
... mention circumstances as information	0.16	0.18	0.07
... stress hope or optimism	0.07	0.16	0.03

Table B.4: Rater 2: Fraction of groups in which the blues...

	Deliberation
... stress the public spirit	0.16
... stress the interests of the own color group	0.11
... suspect lying	0.18
... stress trust	0.05
... mention circumstances as justification for behavior	0.12
... mention circumstances as information	0.13
... stress hope or optimism	0.07

Dimension 2 – Recommendations

Table B.5: Rater 1: Fraction of groups in which the whites...

	Deliberation	TopDown	TopDownClosed
... recommend voting for A	0.73	0.70	0.71
... recommend voting for B	0.41	0.49	0.29
... recommend voting for C	0.06	0.01	0.01
... recommend something different	0.04	0.01	0.06

Table B.6: Rater 1: Fraction of groups in which the blues...

	Deliberation
... recommend voting for A	0.53
... recommend voting for B	0.30
... recommend voting for C	0.54
... recommend something different	0.04

Table B.7: Rater 2: Fraction of groups in which the whites...

	Deliberation	TopDown	TopDownClosed
... recommend voting for A	0.73	0.70	0.71
... recommend voting for B	0.41	0.49	0.29
... recommend voting for C	0.08	0.02	0.01
... recommend something different	0.04	0.01	0.06

Table B.8: Rater 2: Fraction of groups in which the blues...

	Deliberation
... recommend voting for A	0.55
... recommend voting for B	0.30
... recommend voting for C	0.54
... recommend something different	0.03

Dimension 3 – Addressees

Table B.9: Rater 1: Fraction of groups in which the whites...

	Deliberation	TopDown	TopDownClosed
... address the speech to all	1.00	1.00	1.00
... address their speech to their own color group only	0.28	0.22	0.01
... address their speech to the other color group only	0.34	0.19	0.14
... directly address a speech to s.o. from own color group	0.13	0.14	0.06
... directly address a speech to s.o. from other color group	0.18	0.00	0.00

Table B.10: Rater 1: Fraction of groups in which the blues...

	Deliberation
... address the speech to all	0.99
... address their speech to their own color group only	0.37
... address their speech to the other color group only	0.38
... directly address a speech to s.o. from own color group	0.15
... directly address a speech to s.o. from other color group	0.21

Table B.11: Rater 2: Fraction of groups in which the whites...

	Deliberation	TopDown	TopDownClosed
... address the speech to all	1.00	1.00	1.00
... address their speech to their own color group only	0.20	0.01	0.00
... address their speech to the other color group only	0.22	0.06	0.03
... directly address a speech to s.o. from own color group	0.09	0.04	0.02
... directly address a speech to s.o. from other color group	0.11	0.01	0.00

Table B.12: Rater 2: Fraction of groups in which the blues...

	Deliberation
... address the speech to all	0.99
... address the speech to their own color group only	0.29
... address the speech to the other color group only	0.18
... directly address a speech to s.o. from own color group	0.09
... directly address a speech to s.o. from other color group	0.11

Dimension 4 – Showing respect

Table B.13: Rater 1: Fraction of groups in which the whites...

	Deliberation	TopDown	TopDownClosed
... show respect	0.17	0.17	0.15
... behave disrespectfully	0.31	0.14	0.04
... behave neutrally	1.00	1.00	1.00

Table B.14: Rater 1: Fraction of groups in which the blues...

	Deliberation
... show respect	0.13
... behave disrespectfully	0.28
... behave neutrally	1.00

Table B.15: Rater 2: Fraction of groups in which the whites...

	Deliberation	TopDown	TopDownClosed
... show respect	0.06	0.01	0.03
... behave disrespectfully	0.31	0.02	0.01
... behave neutrally	1.00	1.00	1.00

Table B.16: Rater 2: Fraction of groups in which the blues...

	Deliberation
... show respect	0.06
... behave disrespectfully	0.27
... behave neutrally	1.00

Dimension 5 – Specific recommendations I

Table B.17: Rater 1: Fraction of groups in which the whites...

	Deliberation	TopDown	TopDownClosed
... recommend LTED	0.00	0.00	0.00
... recommend A/C	0.02	0.00	0.00
... recommend A/A	0.72	0.69	0.54
... recommend B/B	0.40	0.46	0.26
... recommend C/C	0.05	0.01	0.01
... recommend voting acc. to majority signal	0.14	0.14	0.43
... do not give any such recommendation	0.04	0.01	0.06

Table B.18: Rater 1: Fraction of groups in which the blues...

	Deliberation
... recommend LTED	0.00
... recommend A/C	0.07
... recommend A/A	0.52
... recommend B/B	0.27
... recommend C/C	0.53
... recommend voting acc. to maj. signal	0.06
... do not give any such recommendation	0.03

Table B.19: Rater 2: Fraction of groups in which the whites...

	Deliberation	TopDown	TopDownClosed
... recommend LTED	0.00	0.00	0.00
... recommend A/C	0.02	0.00	0.00
... recommend A/A	0.70	0.70	0.71
... recommend B/B	0.40	0.49	0.29
... recommend C/C	0.06	0.02	0.01
... recommend voting acc. to majority signal	0.03	0.00	0.00
... do not give any such recommendation	0.05	0.01	0.06

Table B.20: Rater 2: Fraction of groups in which the blues...

	Deliberation
... recommend LTED	0.00
... recommend A/C	0.06
... recommend A/A	0.52
... recommend B/B	0.29
... recommend C/C	0.47
... recommend voting acc. to majority signal	0.01
... do not give any such recommendation	0.05

Dimension 6 – Specific recommendations II

Table B.21: Rater 1: Fraction of groups in which the whites give recommendations to both color groups...

	Deliberation	TopDown	TopDownClosed
..., not mentioning signals or abstentions	0.95	0.98	0.75
..., mentioning signals, but not abstentions	0.08	0.14	0.43
..., not mentioning signals, but abstentions	0.01	0.00	0.01
... do not give any such recommendation	0.04	0.01	0.06

Table B.22: Rater 1: Fraction of groups in which the blues give recommendations to both color groups...

	Deliberation
..., not mentioning signals or abstentions	0.93
..., mentioning signals, but not abstentions	0.03
..., not mentioning signals, but abstentions	0.01
... do not give any such recommendation	0.06

Table B.23: Rater 2: Fraction of groups in which the whites give recommendations to both color groups...

	Deliberation	TopDown	TopDownClosed
..., not mentioning signals or abstentions	0.94	0.86	0.85
..., mentioning signals, but not abstentions	0.16	0.42	0.17
..., not mentioning signals, but abstentions	0.00	0.00	0.00
... do not give any such recommendation	0.05	0.01	0.06

Table B.24: Rater 2: Fraction of groups in which the blues give recommendations to both color groups...

	Deliberation
..., not mentioning signals or abstentions	0.94
..., mentioning signals, but not abstentions	0.08
..., not mentioning signals, but abstentions	0.00
... do not give any such recommendation	0.05

Dimension 7 – Inter-group fairness and efficiency

Table B.25: Rater 1: Fraction of groups in which the whites...

	Deliberation	TopDown	TopDownClosed
... mention relative payoffs (W vs. B)	0.10	0.09	0.09
... mention joint payoffs (W + B)	0.19	0.20	0.13

Table B.26: Rater 1: Fraction of groups in which the blues...

	Deliberation
... mention relative payoffs (W vs. B)	0.11
... mention joint payoffs (W + B)	0.04

Table B.27: Rater 2: Fraction of groups in which the whites...

	Deliberation	TopDown	TopDownClosed
... mention relative payoffs (W vs. B)	0.08	0.10	0.03
... mention joint payoffs (W + B)	0.13	0.23	0.26

Table B.28: Rater 2: Fraction of groups in which the blues...

	Deliberation
... mention relative payoffs (W vs. B)	0.09
... mention joint payoffs (W + B)	0.04

C Theoretical appendix

C.1 The game

There are two states of the world, X and Y, and six players that form a group G, with three white and three blue players forming two respective subgroups, G_w and G_b. A player's color is publicly observable. The players have to choose a policy P from three alternative policies , A, B, and C, by a vote. Policies generate state-dependent payoffs that may differ across colors. These payoffs are depicted in Table 1. Nature draws the state of the world ω, which is either X or Y with equal probability, at the beginning of the game. Afterwards, nature randomly draws an informative private signal $s_i \in \{x, y\}$ on the state of the world for each white player i and sends an empty signal $s_i = \emptyset$ to the blue players. Informative signals are conditionally independent and true with probability $p := \{s_i = x \mid \omega = X\} = 0.7$. The subsequent collective policy choice has two stages, the communication stage and the voting stage, in treatments *Deliberation*, *TopDown* and *TopDownClosed*. Treatment *NoChat* has no communication stage.

The voting stage is identical across treatments and is structured as follows: All six players simultaneously and individually place a vote for A, B, or C, or abstain. The winning alternative is determined by the plurality rule, i.e., the alternative with the most votes is implemented. If there is a tie, the winning alternative is chosen randomly, with equal probability of both alternatives. In the end, payoffs from the winning alternative are realized, given the true state of the word.

In all treatments with communication stage, this stage is structured as follows: There is a set of senders $S \supseteq G_w$ and a set of receivers $R(G_\tau)$ to which players in the subgroup G_τ, $\tau \in \{w, b\}$, can send messages. Let M denote the set of all messages that can be constructed in the common language spoken by the six players, including the empty set. Then, any player $i \in G_\tau \subseteq S$ sends a message $m_i \in M$ to $R(G_\tau)$. In *Deliberation*, $S = G = R(G_w) = R(G_b)$, i.e., communication is public and involves everyone as both sender and receiver. In particular, the whites may reveal their signal to the entire group of six (or lie or be silent about it), and both the whites and the blues may recommend a specific voting profile for the group. *TopDown* differs from *Deliberation* in that $S = G_w$, $R(G_w) = G$, and $R(G_b) = \emptyset$. Thus, blues are no longer senders, but messages are still received by everyone. In *TopDownClosed*, by contrast, $S = G_w = R$ on the first communication stage, i.e., the blues are entirely excluded from the communication on that stage, and the whites send messages to the subgroup of white players only. On the second communication stage in *TopDownClosed*, the whites can talk to the entire group; hence, $R = G$.

C.2 Preferences

According to our main hypothesis *MH*, a player's expected utility is the expected sum of his own payoff and the payoffs of his receivers on the - first - communication stage. Hence, players have treatment-dependent preferences that can be described as follows. Let $R^1(G_\tau)$ denote the set of receivers of individuals in G_τ on the first communication stage in the game (which is also the final communication stage in all treatments except *TopDownClosed*). Let $\pi_i(\omega, P)$ denote the final payoff of player i, given the state of the world ω and the chosen policy P. Moreover, let $q_i(\omega \mid m, s_i)$ denote player i's posterior belief about how likely state of the world ω is, given the sent messages m and his own signal s_i; let $\sigma_P(v) \in \left\{0, \frac{1}{2}, 1\right\}$ denote the probability of P being the winning policy, given the voting profile v; and let $z(P) \in \{1,2,3\}$ be an index of the policy and $z(\omega) \in \{1,2\}$ an index of the state of the world. Hence, we can define the utility function as follows:

$$u_i(v) = \sum_{z(\omega)=1}^{2} q_i(\omega \mid m, s_i) \sum_{z(P)=1}^{3} \sigma_{P_{z(P)}}(v \mid \omega) \left(\pi_i\left(\omega, P_{z(P)}\right) + \sum_{\substack{j \in R^1(G_\tau) \\ i \in G_\tau}} \pi_j\left(\omega, P_{z(P)}\right) \right).$$

C.3 Equilibrium concept

We are solving the game for all Perfect Bayesian Nash equilibria in pure strategies that fulfill the following selection criteria.

Definition 1 *(**WU**) Any equilibrium is in **weakly undominated** strategies.*

Definition 2 ***(DT)** Players exhibit **dominant truthtelling**: If there exists a truthtelling equilibrium, no babbling equilibrium is played; i.e., if there exists an equilibrium in which all whites reveal their signal on the / a communication stage, no equilibrium is played in which not all whites reveal their signal on that stage.*

Definition 3 *(**SCT**) Players exhibit **same-color trust**: If the message of a player i to the entire group contradicts a message he has sent to players of his own color only, players of the same color as i believe the message that i has sent to them and disbelieve the message he has sent to the entire group.*

Definition 4 *(**MC**) Players exhibit **minimal coordination** in the following sense: For any $\tau \in \{w, b\}$, players $i \in G_\tau$ who move at the same information set I and hence know that they have identical beliefs $q_i(\omega \mid I) = q$ coordinate on sending the same message and / or voting for the same policy such that they maximize their (joint and individual) expected utility, given the strategies of the other voters.*

Definition 5 *(**LS**) Whites exhibit **literal speaking**: They send the message "x" if they want to indicate that their signal was "x", and they send the message "x" if they want to indicate that their signal was "y".*

Definition 6 *(**CORS**) All players exhibit **conditioning on revealed signals**: They may condition their strategies on the signals that are revealed by the messages but do not use messages as coordination devices otherwise.*

Criterion **WU** excludes equilibria in which *selfish* whites vote for policy C. Criterion **DT** is typical for the cheap-talk literature and selects the equilibrium with the highest degree of information transmission. Criterion **SCT** exludes equilibria in *TopDownClosed* in which the whites cannot lie to the blues without changing the beliefs of the other whites, too. (Note that such equilibria would be extremely implausible since in *TopDownClosed*, the whites can even tell each other that they intend to lie to the blues.) Criterion **MC** restricts attention to equilibria in which the blues coordinate on the same voting strategy (since the blues always have the same information). Moreover, **MC** guarantees that in any truthtelling equilibrium (in which the whites, too, have the same information) the whites also coordinate on the same voting strategy. Criterion **LS** reduces the syntax of the language in which signals are communicated to a binary set and hence simplifies (the proofs in) equilibrium description. **CORS** restricts the function of communication as a coordination device to what is implied by **MC** and allows us to focus on information aggregation rather than pure (uninformed) coordination. The resulting effect of **CORS** is to restrict the number of outcome equivalent equilibria that differ in strategy profiles.

C.4 Equilibria in *Deliberation*

Since in *Deliberation* both the whites and the blues send messages to the entire group, they fully internalize group utility, i.e., they have efficiency preferences. Hence, their (joint) utility is

$$\begin{aligned} u_i^D(v) &= \sum_{z(\omega)=1}^{2} q_i(\omega \mid m, s_i) \sum_{z(P)=1}^{3} \sigma_{P_{z(P)}}(v \mid \omega) \left(\pi_i\left(\omega, P_{z(P)}\right) + \sum_{j \in G} \pi_j\left(\omega, P_{z(P)}\right) \right) \\ &= q_i(X \mid m, s_i)\left(\sigma_A(v \mid X)\, 120 + \sigma_C(v \mid X)\, 30\right) + \\ &+ q_i(Y \mid m, s_i)\left(\sigma_A(v \mid Y)\, 30 + \sigma_B(v \mid Y)\, 90 + \sigma_C(v \mid Y)\, 60\right). \end{aligned}$$

Consider a candidate equilibrium in which the whites truthfully reveal their signals to the public. In such an equilibrium, all players have the same information and hence then same belief about ω: $q_i(\omega \mid m, s_i) = q(\omega \mid m)\, \forall i, \omega$. Hence, *MC* applies to both whites

and blues: Whites coordinate on the same vote and blues coordinate on the same vote, maximizing the expected group payoff, given the strategy of the other color. Strategies of the whites may condition on the private signal or on the messages. Note that under truthtelling, the state of the world that is more likely than the other is indicated both by the signal that is received most often (i.e., twice or even three times) and by the message that is sent most often. Hereafter, we will call this signal and message the *majority signal* and the *majority message*.

Definition 7 *An equilibrium is* ***efficient*** *if and only if A is the winning policy whenever the majority signal is X and B is the winning policy otherwise.*

Proposition 1 *(i) There is a set of truthtelling equilibria in Deliberation that fulfill the selection criteria. They have the following properties: (a) The whites reveal their signals. The whites vote for A if the majority message indicates X and for B otherwise, and the blues abstain (LTED: "let the experts decide"). If (off equilibrium) there is no majority message, then arbitrary off-equilibrium beliefs about the unrevealed signal and voting profiles consistent with these beliefs can be assumed. (b) The whites reveal their signals. Everyone votes for A if the majority message indicates X and votes for B otherwise (A/B). If (off equilibrium) there is no majority message, then arbitrary off-equilibrium beliefs about the unrevealed signal and voting profiles consistent with these beliefs can be assumed. (ii) These equilibria are outcome equivalent in the sense that A is the winning policy if the majority signal indicates X, and B is the winning policy otherwise. (iii) Hence, both types of truthtelling equilibria are efficient. (iv) These equilibria are the unique pure-strategy equilibria in Deliberation that fulfill all selection criteria.*

Proof. We first show that the voting profiles described in (a) and (b) are efficient equilibria of the continuation game on the voting stage, given truthtelling. We then show that truthtelling is an equilibrium strategy of the whites, given the voting profiles in (a) and (b). Finally, we prove (i), i.e., that the two equilibria are the only truthtelling equilibria that fulfill our selection criteria. Consider (a) and (b). Due to truthtelling, the majority message always equals the majority signal. Hence, LTED and AB are efficient. Efficiency preferences imply that no-one has a deviation incentive. Thus, the voting profiles in (a) and (b) are equilibria of the continuation game on the voting stage. Consider now the communication stage preceding the voting stage with LTED. If a white deviates from telling the truth in that he lies about his signal, then he either does not change the majority message, hence leaving the voting outcome unchanged, too, or he changes the majority message and hence moves the voting outcome away from efficiency. Thus, no white wants to deviate to lying. Now consider a deviation to silence. Again, this

either changes nothing or moves the voting outcome away from efficiency, depending on the off-equilibrium voting strategies. Hence, again, the whites do not want to deviate. The same kind of argument holds true for the communication stage that precedes a voting profile described in (b). Thus, in *Deliberation* there exist truthtelling equilibria with voting profiles as described in (a) and (b). Parts (ii) and (iii) follow directly.

It now remains to show that these two sets of equilibria defined above contain the only truthtelling equilibria in *Deliberation* that fulfill our selection criteria. Note first that *CORS* excludes equilibria in which strategies condition on messages without conditioning on beliefs. Furthermore, note that under truthtelling, *MC* applies both to the whites and the blues. If the whites have revealed their signals and the blues abstain, then *MC* and efficiency preferences imply that the whites coordinate on voting for A or B, depending on the majority message. If the whites do this, and if they have revealed their signals, then *MC* and efficiency preferences imply that the blues coordinate on a strategy that never distorts the voting outcome away from efficiency. Hence, in this case the only two voting profiles of the blues that fulfill *MC* are abstention and voting along with the whites. Finally, note that *DT* excludes equilibria with partial truthtelling. Hence, *CORS*, *MC*, *DT*, and efficiency preferences pin down all truthtelling equilibria in *Deliberation* to the ones that are characterized in (a) and (b). Part (iv) follows directly from this and *DT*. ■

From Proposition 1, the following outcome-related result can be derived:

Result 1: *In Deliberation, (a) the whites truthfully reveal their signal; and (b) if the majority signal indicates X, all votes that are placed are for A (A/A); whereas (c) if the majority signal indicates Y, all votes that are placed are for B (B/B).*

C.5 Equilibria in *TopDown*

In *TopDown*, the whites can still address the entire group on the communication stage, but the blues are no longer senders. Hence, the whites still have efficiency preferences, but the blues become self-interested. Still, the blues have the same information, so MC still applies to them: $q_i(\omega \mid m) = q(\omega \mid m)\ \forall i \in G_b$. Moreover, note that payoffs are perfectly aligned across players of the same color; thus we can define $\pi_i\left(\omega, P_{z(P)}\right) := \pi_b\left(\omega, P_{z(P)}\right)$ $\forall\ i \in G_b$. Hence, in *TopDown* a player i has utility u_i^{TD} as follows:

$$u_i^{TD} = u_i^D(v) \text{ if } i \in G_w,$$

$$u_i^{TD} = \sum_{z(\omega)=1}^{2} q(\omega \mid m) \sum_{z(P)=1}^{3} \sigma_{P_{z(P)}}(v \mid \omega)\, \pi_b\left(\omega, P_{z(P)}\right) \text{ if } i \in G_b.$$

Consider truthtelling equilibria.

Proposition 2 (i) *There is a set of truthtelling equilibria in TopDown that fulfill the selection criteria. They have the following properties: The whites reveal their signals. If the majority message indicates X, then (a) all vote for A, (b) all whites vote for A and all blues abstain, or (c) all whites abstain and all blues vote for A. If the majority message indicates Y, then the whites vote for B and the blues for C (B/C). If (off equilibrium) there is no majority message, we restrict off-equilibrium beliefs as follows: If there is a one-shot deviation of one white player to being silent and the remaining revealed signals contradict each other(i.e., there is no majority message), then the blues have a belief $q(X \mid m) < \frac{2}{3}$. Then in all voting profiles consistent with off-equilibrium beliefs after such a deviation, the blues vote for C. (ii) These equilibria are outcome equivalent: They generate winning policy A if the majority signal is X and a tie between B and C if the majority signal is Y. (iii) These equilibria are inefficient. (iv) These equilibria are the unique pure-strategy equilibria that fulfill all selection criteria.*

Proof. We first show that with truthtelling of the whites on the communication stage, voting profiles with the properties described in (i) are equilibria of the continuation game. Second, we show that then, truthtelling must be part of the equilibrium. Part (ii) directly follows from part (i); and (iii) directly follows from (ii) and the definition of efficient equilibrium.

Assume now truthtelling of the whites, and consider the blues first. For $q(\omega \mid m) < \frac{2}{3}$, policy C is strictly better for a blue player than the other policies, otherwise, policy A is better than the other policies. If x is the majority message, we have $q(\omega \mid m) \geq 0.7 > \frac{2}{3}$, and if y is the majority message, we have $q(\omega \mid m) \leq 0.3 < \frac{1}{3}$. Thus, if x is the majority message, then A is better for any blue player than (a tie with) any other policy; and if y is the majority message, then (a tie with) C is better for any blue player than (a tie with) any other policy. Then, MC implies that all blues vote for A or abstain if x is the majority message and vote for C otherwise. If there is no majority message (i.e., there are only two messages that contradict each other), then the off-equilibrium belief of the blues, $q(\omega \mid m) < \frac{2}{3}$, and MC imply that all blues vote for C.

Consider now the whites on the voting stage. Remember that they have efficiency preferences. If the majority message is x, then any white prefers A over all other policies. MC then implies that all whites coordinate on an action that makes A the winning policy; i.e., voting for A, or, (only) if the blues vote for A, abstention. If the majority message is y, then any white anticipates the three blue votes for C but prefers B over all other policies himself. Hence, he also prefers a tie between B and C over C or any other tie with C. Thus, MC implies that all whites vote for B. If there is no majority message, i.e., if there are only two messages that contradict each other, then any off-

equilibrium belief about the unrevealed signal and any consistent voting strategy of the whites can be assumed. Note that regardless of the voting strategy of the whites after such a deviation, the resulting efficiency level (group payoff) cannot exceed the level implied by the equilibrium strategies (because strategies cannot improve upon conditioning on the full information about all signals). Thus far, we have shown that under truthtelling, voting profiles with the properties described in (a), (b), and (c) are equilibria of the continuation game on the voting stage.

Consider now the communication stage. We check the incentive of an arbitrary white player i to deviate to a lie or to being silent about his signal. Consider now a white who has received a signal s_i. If he lies or is silent about s_i, then he is either not pivotal, the other two messages being $m_{-i} = (y, y)$ or $m_{-i} = (x, x)$, in which case the deviation does not change anything. Or i is pivotal, in which case the other two whites have contradicting signals and s_i is the majority signal. Then, i's efficiency preferences imply that he cannot do better than revealing his signal. Thus, there is no deviation incentive on the communication stage.

We now proceed to proving (i) by showing that *all* truthtelling equilibria in *TopDown* have the properties that are described in (a), (b), and (c). Note first that *CORS* excludes equilibria in which strategies condition on messages without conditioning on beliefs. Second, under truthtelling, *MC* applies to both colors. Hence, under truthtelling each color will coordinate on an action that maximizes the probability of the policy preferred by this color, given the strategy of the other color and the common beliefs about the state of the world. But then, (a), (b), and (c) describe all voting profiles under truthtelling. Moreover, *DT* excludes partial truthtelling and babbling equilibria. Thus, *MC*, *CORS,* and *DT* restrict all pure-strategy equilibria in *TopDown* to the set described in Proposition 2, which proves part (iv). ■

From Proposition 2, the following outcome-related result can be derived:

Result 2: *In TopDown, (a) the whites truthfully reveal their signal; and (b) if the majority signal indicates X, all votes that are placed are for A (A/A); whereas (c) if the majority signal indicates Y, the whites vote for B and the blues for C (B/C).*

C.6 Equilibria in *TopDownClosed*

In *TopDownClosed*, our main hypothesis *MH* implies that the whites do not have efficiency preferences any longer but maximize the joint payoffs of their own color group instead (color-group identity). Note that this is equivalent to being selfish since payoffs are perfectly aligned between individuals of the same color. Importantly, *WU* implies that selfish whites never vote for C, since they prefer any possible outcome of the vote over C,

regardless of their beliefs about the state of the world, so that voting for C is a weakly dominated strategy for selfish whites. The blues, too, are selfish, as in *TopDown*.

Consider now potential equilibria in which the whites truthfully reveal their signals to each other on the first communication stage. Note that in such equilibria, the whites have identical beliefs on the voting stage, so that *MC* applies to them. Note that MC always applies to the blues, regardless of whether they are told the true signals or not.

Proposition 3 *(i) There is a set of equilibria in TopDownClosed that fulfill the selection criteria. They have the following properties: The whites reveal their signals to each other, but babble to the blues. The whites vote for A if the majority message indicates X and for B otherwise, and the blues vote for C (AC/BC). If (off equilibrium) there is no majority message on the first communication stage, arbitrary off-equilibrium beliefs of the whites and white votes consistent with these beliefs can be assumed; but the blues (unobservant of the deviation) are restricted to keep their prior beliefs and hence to vote for C. (ii) These equilibria are inefficient. (iii) These equilibria are the unique pure-strategy equilibria that fulfill all selection criteria.*

Proof. We first show that given truthtelling on the first communication stage, there can be no truthtelling on the second communication stage. We then show existence of the AC-BC equilibria as characterized in (i). Part (ii) - inefficiency - directly follows from (i) and the definition of efficiency. Finally, we will prove (iii).

Assume now that the whites truthfully reveal their signals to each other on the communication stage. Assume for the sake of argument that there is also truthtelling on the second communication stage. Consider now a situation in which the majority signal indicates Y, but there has been one signal indicating X. On the voting stage, both the whites and the blues hence believe that the state of the world is Y with probability 0.7. But then, their preferences and *MC* imply that the whites vote for B and the blues for C. Under truthtelling, the whites' expected utility is $0.3 \times 0 + 0.7\left(\frac{1}{2} \times 20 + \frac{1}{2} \times 0\right) = 7$. If, by contrast, one of the whites who have received the signal indicating Y deviates to a lie, saying that his signal indicates X, the beliefs of the whites will not change since this is precluded by *SCT*, but the blues will believe that the state of the world is X with probability 0.7. Then, *MC* and the players' preferences imply that the whites will still vote for B and the blues will vote for A. For the whites, this yields an expected utility of

$$0.3\left(\frac{1}{2} \times 20 + \frac{1}{2} \times 0\right) + 0.7\left(\frac{1}{2} \times 10 + \frac{1}{2} \times 20\right) = 13.5.$$

Thus, the lie strictly increases the expected utility of the whites. Consider now a white i whose signal was $s_i = Y$. This white is pivotal on the second communication stage in

the sense that his message determines the majority message sent to the blues (since the other two whites are assumed to tell their true - contradictory - signals). Thus, this white has a strict incentive to lie on the second communication stage. This proves that under truthtelling on the first communication stage, there can be no truthtelling on the second communication stage if the signal distribution is $2:1$.

Consider now a situation in which all three signals indicate Y. Then, under truthtelling on both commication stages, no white is the pivotal sender on the second communication stage any longer, and the individual lying incentive does no longer exist on the equilibrium path. Instead, a given white in this situation is indifferent between lying and revealing his signal, given that the other two whites reveal that their signals indicated Y. (Note that the whites know the signal distribution on the second communication stage since we assume truthtelling on the first communication stage.) However, the whites still *prefer* that the blues vote for A rather than C. To see this, note that their expected utility if the blues vote for A (and they themselves for B) would be

$$\frac{0.3^3}{0.3^3+0.7^3}\left(\frac{1}{2}\times 3\times 20+\frac{1}{2}\times 3\times 0\right)+\frac{0.7^3}{0.3^3+0.7^3}\left(\frac{1}{2}\times 3\times 10+\frac{1}{2}\times 3\times 20\right)$$
$$=43.905.$$

By contrast, if the blues vote for C, the whites' expected utility amounts to

$$\frac{0.3^3}{0.3^3+0.7^3}\times 0+\frac{0.7^3}{0.3^3+0.7^3}\left(\frac{1}{2}\times 3\times 20+\frac{1}{2}\times 3\times 0\right)=27.811.$$

Thus, the whites have a higher expected utility if the blues vote for A rather than C. Now note that the whites have identical beliefs on the second communication stage due to truthtelling on the first communication stage. Thus, MC applies to them on the second communication stage. But sending a majority message that indicates Y and thus making the blues vote for C violates *MC*. Thus, our selection criteria exclude equilibria in which any signal distribution leads to truthtelling on the second communication stage.

Consider now potential equilibria with truthtelling on the first communication stage and babbling on the second communication stage. Consider the voting stage first. The blues have their prior belief that both states of the world are equally likely. Thus, their selfish preferences and *MC* imply that they coordinate on voting for C. The whites, by contrast, know the actual signal distribution s. They prefer A whenever $q(X \mid s)\geq 0.7$ and B otherwise. Hence, they also prefer a tie between A and C whenever $q(X \mid s)\geq 0.7$ and a tie between B and C otherwise. But then, MC implies that they coordinate on voting for A whenever the majority signal indicates X and on voting for B otherwise. This proves the voting profile AC/BC on the equilibrium path.

Consider now the second communication stage. Given that the blues do not condition their beliefs on the messages sent, no white has an incentive to deviate from babbling to conditioning his message on his signal. Note that this also holds true off equilibrium, i.e., after a deviation of a white / some whites on the first communication stage.

Now consider the first communication stage. Given that the whites believe each other, no white has an incentive to deviate to being silent or to lying. To see this, note that such a deviation would either change nothing or would distort the beliefs of the other two whites away from the true signal distribution. This distortion, in its turn, would either change nothing or distort the votes of the other two whites away from the voting profile that maximizes the whites' expected utility, given that the blues vote for C. Note that the blues cannot observe any deviation on the first communication stage. Hence, they cannot respond to such a deviation and will vote for C after it, too.

This proves parts (i) and (iii) of Proposition 3. Part (ii) is trivial. ■

From Proposition 3, the following outcome-related result can be derived:

Result 3: *In TopDownClosed, (a) the whites truthfully reveal their signals to each other but babble to the blues, and (b) if the majority signal indicates X, the whites vote for A but the blues for C (A/C), whereas (c) if the majority signal indicates Y, the whites vote for B and the blues still for C (B/C).*

C.7 Equilibria in *NoChat*

In the *NoChat* treatment, there is no possibility to communicate. Therefore, both colors become self-interested and maximize the utility of their own color. Moreover, only the blues (know that they) have the same information set, namely their prior belief that the two states of the world are equally likely. The whites, however, have private independent information on the true state. Hence, *MC* applies to the blues but not to the whites.

Proposition 4 *(i) There is a set of equilibria in NoChat that fulfill the selection criteria. They have the following properties: The blues vote for C, and (a) the whites vote for A (A/C), or (b) the whites vote for B (B/C), or the whites vote for A if their signal indicates X and for B otherwise (split-whites). (ii) These equilibria are inefficient.*

Proof. Since *MC* applies to the blues, we only have equilibria in which the blues coordinate on the same vote. Since the blues are self-interested in *NoChat*, votes other than C are weakly dominated for them. Hence, MC and WU restrict the analysis to equilibria in which the blues vote for C. The whites are self-interested, too. Given that the blues vote for C, each white will minimize the probability of the implementation of C (since C

provides strictly lower expected payoffs than any other policy for a self-interested white, regardless of his signal). Voting for C is hence weakly dominated for the whites. Thus, WU exludes equilibria in which some whites, too, vote for C. Consider now an arbitrary white i. If the two other whites vote for the same policy (that is not C), then i's best response is to vote for this policy, too, in order to decrease the probability of C from 1 to 0.5. Hence, AC and BC are equilibria. If, now, the two other whites vote for the policy indicated by their signal (A if the signal indicates X and B otherwise), the best response of i is to vote in line with his signal, too. To see this, note that this strategy maximizes the probability of hitting the vote of the other two whites if they voted for the same policy, and hence minimizes the probability of C. Thus, split-whites is an equilibrium, too. Note that self-interest, MC (for the blues) and WU (for both colors) exlude other possible equilibria. This proves (i). Part (ii) follows from the definition of efficiency. ■

Proposition 4 implies the following outcome-related result:

Result 4: *In NoChat, the blues vote for C, regardless of the majority signal; and the whites vote for A or B or according to their signal (resulting in outcome A/C or B/C).*

Excluding all equilibrium outcomes with abstention, results 1-4 imply our testable hypotheses 1-4 in Section 4 of the paper. In these hypotheses, we focus on the treatment comparisons, i.e., on the comparative statics, rather than on point predictions.

C.8 Predictions with standard preferences

Standard preferences would imply that all players are selfish maximizers of their own expected payoff. Due to our design, this is equivalent to assuming a color-group identity for both colors in all treatments. Hence, the predictions for treatments *NoChat* and *TopDownClosed* would not change if we assumed standard preferences.

By contrast, our predictions for *Deliberation* and *TopDown* would change: As is easy to show, there would not be any truthtelling equilibria but only babbling equilibria in these two treatments since each white would have an incentive to lie to the blues and report "x" even if her signal indicated "y". We omit the proof, but a crucial point in the proof is that selfish whites prefer A/A over B/C even under majority signal Y. Accordingly, if players were selfish, the predictions for *Deliberation* and *TopDown* would coincide with those for *NoChat*.

D Supplementary online material

Additional tables

Table SOM.1: Voting outcomes at the group level over time – Conditional on the received majority signal – First 10 periods

	Majority signal: X				Majority signal: Y			
	NoC	D	TD	TDC	NoC	D	TD	TDC
A/C outcome	**0.303**	0.058	0.047	0.076	**0.128**	0.052	0.106	0.020
of this: split-whites	0.197	0.019	0.012	0.022	-	-	-	-
Almost A/C outcome	0.516	0.087	0.233	0.228	0.513	0.042	0.058	0.122
(Almost) A/C outcome	0.820	0.144	0.279	0.304	0.641	0.094	0.163	0.143
B/C outcome	0	0	0	0	0.026	**0.240**	**0.279**	**0.194**
of this: split-whites	-	-	-	-	0.013	0.115	0.154	0.082
(Almost) B/C outcome	0.016	0	0	0	0.179	0.479	0.654	0.469
Split-whites	0.279	0.019	0.012	0.022	0.051	0.135	0.173	0.092
LTED	0	0	0	0	0	0	0	0
Almost LTED	0.025	0	0	0	0.013	0	0	0.010
(Almost) LTED	0.025	0	0	0	0.013	0	0	0.010
A/A	0	**0.548**	**0.337**	**0.228**	0	0.063	0	0.020
Almost A/A	0.041	0.250	0.349	0.424	0.038	0.042	0.038	0.133
(Almost) A/A	0.041	0.798	0.686	0.652	0.038	0.104	0.038	0.153
B/B	0	0	0	0	0	0.167	0.010	0.010
Almost B/B	0	0.010	0	0.011	0	0.125	0.115	0.102
(Almost) B/B	0	0.010	0	0.011	0	0.292	0.125	0.112
Other	0.098	0.048	0.035	0.033	0.128	0.031	0.019	0.112

Treatment names are abbreviated with NoC (NoChat), D (Deliberation), TD (TopDown) and TDC (TopDownClosed). In "Almost" outcomes at most one player per color group deviates from the respective outcome. LTED refers to the "Let the experts decide equilibrium". Figures printed in bold highlight the observed modal voting outcomes for the respective treatment and signal combination.

Table SOM.2: Voting outcomes at the group level over time – Conditional on the received majority signal – Last 10 periods

	Majority signal: X				Majority signal: Y			
	NoC	D	TD	TDC	NoC	D	TD	TDC
A/C outcome	**0.375**	0.180	**0.161**	**0.237**	**0.443**	0.157	0.124	0.086
of this: split-whites	0.170	0.027	0.011	0.052	-	-	-	-
Almost A/C outcome	0.509	0.270	0.376	0.443	0.443	0.135	0.134	0.129
(Almost) A/C outcome	0.884	0.450	0.538	0.680	0.886	0.292	0.258	0.215
B/C outcome	0	0	0	0	0	**0.191**	**0.351**	**0.376**
of this: split-whites	-	-	-	-	0	0.067	0.175	0.226
Almost B/C outcome	0.009	0	0.011	0	0.034	0.236	0.278	0.237
(Almost) B/C outcome	0.009	0	0.011	0	0.034	0.427	0.629	0.613
Split-whites	0.259	0.027	0.022	0.052	0.011	0.079	0.175	0.247
LTED	0	0	0	0	0	0	0	0
Almost LTED	0.018	0	0	0	0.023	0	0	0
(Almost) LTED	0.018	0	0	0	0.023	0	0	0
A/A	0	**0.234**	0.075	0.062	0	0.056	0	0.022
Almost A/A	0.036	0.315	0.366	0.237	0	0.079	0.072	0.075
(Almost) A/A	0.036	0.550	0.441	0.299	0	0.135	0.072	0.097
B/B	0	0	0	0	0	0	0	0
Almost B/B	0	0	0	0	0	0.124	0.010	0.011
(Almost) B/B	0	0	0	0	0	0.124	0.010	0.011
Other	0.054	0	0.011	0.021	0.057	0.022	0.031	0.065

Treatment names are abbreviated with NoC (NoChat), D (Deliberation), TD (TopDown) and TDC (TopDownClosed). In "Almost" outcomes at most one player per color group deviates from the respective outcome. LTED refers to the "Let the experts decide equilibrium". Figures printed in bold highlight the observed modal voting outcomes for the respective treatment and signal combination.

Table SOM.3: Summary of the Hypotheses Testing

Hypothesis	Behavior consistent with the hypothesis	Behavior inconsistent with the hypothesis
1a. Voting outcomes given majority signal X	A/A more frequent in D and TD than in TDC and NoC. A/C more frequent in NoC than in D and TDC	A/C more frequent in TDC than in D or TDC
1b. Voting outcomes given majority signal Y	B/B more frequent in D than in all other treatments	B/C not more frequent in TD or TDC than in D
2a. Whites' voting decisions	Given majority X(Y), whites' votes for $A(B)$ not different between D, TD and TDC and higher than in NoC.	–
2b. Blues' voting decisions	Given majority X, blues' votes for A higher in D than in TDC and NoC and higher in TD than in NoC. Given majority Y, blues' votes for B higher in D than TD, TDC and NoC.	Given majority X, blues' votes for A is not different between TD and TDC.
3a. Whites' lying	Lying less frequent in D and TD than in TDC	Lying is weakly more frequent in D than in TD
3b. Blues' trustfulness	Blues more trusting in D than in TDC	Blues not more trusting in TD than in TDC
4a. Efficiency ranking	Efficiency in all communication treatment higher than in NoC.	Efficiency ranking of the communication treatments as predicted, but differences are not significant.
4b. Earnings' ranking	Whites' earnings in all communication treatments higher than in NoC.	Whites' earnings ranking of the communication treatments as predicted, but differences are not significant. Blues' earnings ranking is D = NoC > TD = TDC, unlike the predicted TD > TDC > NoC > D

Treatment names are abbreviated with NoC (NoChat), D (Deliberation), TD (TopDown) and TDC (TopDownClosed). Moreover, to enhance readability, 'significantly higher' is abbreviated with 'higher'.

Translated instructions

Welcome to today's experiment!

You are taking part in a decision situation and it is possible for you to earn some money. The amount of money that you are able to win depends on your decisions and on the decisions of the other participants that are assigned to you. Moreover, it is influenced by the role that is randomly allocated to you. After having finished the experiment, we would like to ask you to fill in a short questionnaire.

Please note that from now on and throughout the experiment it is **not allowed to communicate** unless the computer explicitly asks you to do so. If you have any questions, please raise your hand out of your cubicle. One of the experimenters will come to you then. Throughout the experiment, it is forbidden to use mobile phones, smartphones, tablets or the like. Any violation of the rules leads to exclusion from the experiment and payment. All decisions are made anonymously, i.e. none of the participants learns about the identity of the others. Also the payment will be made anonymously at the end of the experiment.

Instructions

1. What's it about – An overview

[NoChat:] This experiment is about making a decision within a group between three different options A, B and C by way of vote.

[Deliberation / TopDown / TopDownClosed:] This experiment is about making a decision within a group between three different options A, B and C through communication and by way of vote.

A group consists of three „white" and three „blue" members. Your payment depends on the decision that the group makes regarding the possible options. It depends, first, on the fact which of the options will be implemented. Second, it is determined by the role you are assigned to – the "white" one or the "blue" one. And third, it also depends on the situation that occurs – this can be either X or Y. The graph below, comprising two tables, shows how many points a white and blue group member can earn given the three options and depending the situation that occurs – X (left table) or Y (right table).

Situation X

Options	White members	Blue members
A	20	20
B	0	0
C	0	10

Situation Y

Options	White members	Blue members
A	10	0
B	20	10
C	0	20

The following applies for situation X: If option A is implemented, the white members and the blue members earn 20 points; if option B is implemented none of the members earns anything. If option C is implemented, the white members do not earn anything and the blue members earn 10 points.

The analogue applies for situation Y: If option A is implemented, the white members earn 10 points and the blue members do not earn anything; if option B is implemented, the white members earn 20 points and the blue members earn 10 points. If option C is implemented, the white members do not earn anything and the blue members earn 20 points.

The situation is not directly observable, but is selected randomly by the computer; both situations X and Y are equally likely i.e. they will be realized with a probability of 50%. The situation that is chosen by the computer is valid for the entire group; i.e. the payments for the white members as well as for the blue members are determined by either the left table or the right table. Thus, one could also say that the computer selects randomly one out of the two tables for the entire group, whereby both tables are equally likely.

Besides the partly different payments, there is also another difference between the white members and the blue members within a group: Each white member receives independent information by the computer on whether situation X or situation Y occurs. This information is true with a 70% probability (i.e. it is true in 70 out of 100 cases and wrong in 30 out of 100 cases). Thus, as this information does not always have to be true, it is possible that not all three white members receive the same information by the computer. The blue members do not receive any information by the computer.

[NoChat:] In order to make a decision between the three options, the group goes through a two-stage process. On the first stage, all group members can take notes in order to sort out their thoughts. On the second stage the voting will be carried out. The option with the most votes will be implemented.

[Deliberation:] In order to make a decision between the three options, the group goes through a two-stage process. On the first stage, all group members can chat together. On the second stage the voting will be carried out. The option with the most votes will be implemented.

[TopDown:] In order to make a decision between the three options, the group goes through a two-stage process. On the first stage, all white group members can chat together and send messages to the entire group. The blue members can read these messages, but they cannot actively take part in chatting. On the second stage the voting will be carried out. The option with the most votes will be implemented.

[TopDownClosed:] In order to make a decision between the three options, the group goes through a three-stage process. On the first stage, all white group members can chat together. The blue members cannot read these messages. On the second stage, all white group members can chat together and send messages to the entire group. The blue members can read these messages, but

they cannot actively take part in chatting. On the third stage the voting will be carried out. The option with the most votes will be implemented.

The experiment comprises 20 rounds.

In the following, the experiment will be explained in detail:

1. The allocation of the roles

At the beginning of the experiment, the computer randomly assigns every participant either the role of a white member or that of a blue member. The **roles** remain **constant throughout the whole experiment**, i.e. one's own role will not change between rounds. Instead, in each round the group constellation will be re-determined: In each round the computer randomly allocates the participants to groups of six, consisting of three white members and three blue members.

In the following the course of an (arbitrary) round will be described. The experiment consists of **20 rounds**. The payments in any given round only depend on what happens in that round – they are independent of former rounds. The situation that occurs in a given round is likewise independent of the situations that have occurred in former rounds.

2. Course of a round

At the beginning of each round, the computer randomly assigns the whites and the blues to groups of six, consisting of three white members and three blue members. Then each **white member** receives **information** by the computer on whether situation X or Y prevails, i.e. if the left or right table is correct. This information is true with a 70% probability. The blue members do not get any information.

[NoChat:] Then a **"note"**-window opens where you can write down notes. Please use the window only for taking notes regarding things that are relevant for the experiment. The window will disappear after **two minutes**. You will see in the top right corner how much time you have left.

[Deliberation:] Then a **"chat"**- window opens where **all group members**, the white and the blue members, can chat together. The computer randomly assigns everyone who enters a message a number that will be shown at the beginning of the message sent together with the role (white or blue). A possible pseudonym is for example „Blue 1". Please note: The pseudonyms are only valid **for this round**. With the help of these pseudonyms you can address each other and keep track of which messages are sent from the same person during the chat. Throughout the chat you can try to influence the voting decisions of the others. Please only use this chat for exchanging views on things that are relevant for the experiment. It is not allowed to uncover one's own identity or the identity of other group members. The chat window will disappear after **two minutes**. You will see in the top right corner how much time you have left.

[TopDown:] Then a "**chat"**- window opens where the **white group members** can chat together and send messages to the entire group. The blue members can read these messages, but they cannot

actively take part in chatting. The computer randomly assigns all white members who enter a message a number that will be shown at the beginning of the messages sent. A possible pseudonym is for example „White 1“.Please note: The pseudonyms are only valid **for this round**. With the help of these pseudonyms you can address each other and keep track of which messages are sent from the same person during the chat. Throughout the chat you can try to influence the voting decisions by the others. Please only use this chat for exchanging views on things that are relevant for the experiment. It is not allowed to uncover one's own identity or the identity of other group members. The chat window will disappear after **two minutes**. You will see in the top right corner how much time you have left.

[TopDownClosed:] Then a **"chat"**- window opens for the white group members where they can chat together. The blue members cannot read these messages. They have to wait for the experiment to proceed. Subsequently, another chat window opens where the white group members can chat together and send messages to the entire group. The blue members can read these messages, but they cannot actively take part in chatting. In both chats, the computer assigns all white members who enter a message randomly a number that will be shown at the beginning of the messages sent. A possible pseudonym is for example „White 1“. Please notice: The pseudonyms are only valid **for this round**. With the help of these pseudonyms you can address each other and keep track of which messages are sent from the same person during the chat. Throughout the chat you can try to influence the voting decisions by the others. Please only use this chat for exchanging views on things that are relevant for the experiment. It is not allowed to uncover one's own identity or the identity of other group members. Each of these chat windows will disappear after **one minute**. You will see in the top right corner how much time you have left.

[NoChat:] In the next step there is a secret **vote over the three options**. That means each group member can vote anonymously either for A or B or C or abstain from voting. Ultimately, the computer implements the **option with the most votes**. (In case of parity of votes the computer randomly chooses between the options with the most votes. Also in case that all group members abstain from voting, the computer randomly chooses one of the three options.)

[Deliberation / TopDown:] After the chat there is a secret **vote over the three options**. That means, each group member can vote anonymously either for A or B or C or abstain from voting. Ultimately, the computer implements the **option with the most votes**. (In case of parity of votes the computer randomly chooses between the options with the most votes. Also in case that all group members abstain from voting, the computer randomly chooses one of the three options.)

[TopDownClosed:] After the second chat there is a secret **vote over the three options**. That means, each group member can vote either for A or B or C or abstain from voting. Ultimately, the computer implements the **option with the most votes**. (In case of parity of votes the computer randomly chooses between the options with the most votes. Also in case that all group members abstain from voting, the computer chooses one of the three options.)

Then all group members are informed about the option that has been elected and they learn about the distribution of votes, i.e. how many votes option A has received, how many votes option B has received, how many votes option C has received and how many abstentions there have been. Moreover, the computer screen informs each group member about the situation that has occurred and how many points he or she has earned in the given round.

3. Total payment for the experiment

At the end of the experiment the computer will randomly, and independently from each other, selected three rounds. All rounds are equally likely. The payments that you have earned in these selected rounds will be summed up and converted into EURO with the **exchange rate 1 EURO = 3 POINTS**. Your total earnings from the experiment consist of the resulting amount plus the show-up fee of 10 EURO.