Exiting the Echo Chamber: Can Discussions in Randomly Formed Groups Change Opinions and Votes?

Hendrik Hüning[*] *Lydia Mechtenberg*[*] *Stephanie W. Wang*[†]

February 27, 2022

Abstract

Political polarization has increased recently with people withdrawing into their respective echo chambers. We test a potential remedy: discussions in randomly formed groups with varying political composition. In a two-wave experiment, we collected voting intentions and textual data from interactions before the ballot and self-reported votes afterwards. Using NLP and ML methods, we find a striking "double standard." Participants do change opinions and votes when confronted with opposing views supported by strong enough arguments. However, they are more likely to maintain their prior opinions when more chat partners share them, even if these like-minded peers do not provide supporting arguments.

Keywords: Chat, Survey Experiment, Voting
JEL: D01; D04; D72; D83

[*]Department of Economics, Hamburg University, Von-Melle-Park 5, 20146 Hamburg, Germany, Email addresses: hendrik.huening@uni-hamburg.de and lydia.mechtenberg@uni-hamburg.de

[†]Department of Economics, University of Pittsburgh, 230 South Bouquet Street, Pittsburgh, PA 15260, USA, Email: swwang@pitt.edu.

We are grateful to Sophia Schulze-Schleithoff, Miriam Hinternesch and Sofia Schnitzler for excellent research assistance and the WISO lab of Hamburg University for the outstanding technical assistance. This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 822590. Any dissemination of results here presented reflects only the authors' view. The Agency is not responsible for any use that may be made of the information it contains.

1 Introduction

There has been a marked rise in political polarization in recent years (Bail et al. 2018; Pew Research Center 2014). People consume media sources that share their political orientation, talk with like-minded people, and are generally ensconced in their respective echo chambers. Political opinions are entrenched, and misinformation that confirms prior views proliferates, unchallenged within each group.

We explore an understudied potential interruption to this information trap: communication in a *randomly* composed group of people about the issue at hand. Much of the existing experimental interventions aimed at changing political beliefs focus on information provision (Haaland et al. 2022; Ahler and Sood 2018; Kendall et al. 2015).[1] As yet, these interventions usually avoid direct peer-to-peer interaction.[2] We ask whether people would learn from each other if they deliberated a contentious issue online in a group with those who held different opinions, i.e. people they were unlikely to encounter in their usual ecosystem. This is an easily implementable and scalable counterfactual to the usual exchanges that take place in homogeneous social groups. In contrast to much of the existing literature, we also elicit (self-reported) votes in an actual ballot that our participants placed some days after their chat interaction. Hence, we test whether direct chat interaction in a randomly matched group triggers both changes in opinions and changes in behavior.

Will deliberation in a heterogeneous group lead to opinion change? Most economic models assume that people weigh new information from others according to quality, and update their beliefs. Under this assumption, models of information aggregation usually predict that communication amongst peers improves accuracy and leads to belief convergence. Specifically, informative communication can increase the accuracy of beliefs according to the theoretical literature on persuasive cheap talk (Crawford and Sobel 1982; Crawford 1998; Green and Stokey 2007; Chakraborty and Harbaugh 2010), Bayesian persuasion (Kamenica and Gentzkow 2011; Kamenica 2019), Bayesian learning in social networks (Gale and Kariv 2003; Mueller-Frank 2013; Mossel et al. 2015; Mossel et al. 2016), and deliberation (Gerardi and Yariv 2007, Goeree and Yariv 2011, Iaryczower et al. 2018). Opinion change towards the better informed can also be found in lab experiments testing cheap-talk models (Blume et al. 2020), models of Bayesian persuasion (Fréchette

[1]Balietti et al. 2021 aim at the negative effects of political polarization and exogenously expose the participants of their survey experiment to opinions they would not encounter otherwise. However, Balietti et al. 2021 do not let participants interact directly and do not systematically investigate the effect of argument use on persuasion.

[2]One potential reason for this reluctance is the assumption, formed upon observing polarization in social media, that direct interaction could easily get out of hand, escalating into unfriendly exchanges that increase rather than decrease polarization (Balietti et al. 2021).

et al. 2018), communication networks (Buechel and Mechtenberg 2019; Grimm and Mengel 2020) and deliberation (Goeree and Yariv 2011).[3]

However, optimistic predictions on opinion change have also been challenged by observational evidence: opinions on important economic issues differ widely and persistently.[4] Meanwhile, experiments, usually in the lab, have documented under-weighing and biased interpretation of countervailing information (see Golman et al. 2017, for a survey), providing evidence for self-serving biases that could prevent opinion change, such as confirmatory bias (Rabin and Schrag 1999, Charness and Dave 2017) or overconfidence (Ortoleva and Snowberg 2015).

We hypothesize that in major debates on issues of consequence, both the quality of the messages exchanged and self-serving biases determine the extent to which people learn from each other. To investigate this, we conduct a survey-chat experiment in the field. We report findings from a two-wave experiment in which groups of randomly matched participants discussed how to vote on the Local Rent Control Initiative on the 2018 California ballot. This initiative would allow local governments to implement rent control. After the election, we asked participants in the follow-up survey how they actually voted. Altogether, we collected the participants' prior opinions on how they would vote, all chat text, and self-reported actual votes. We analyzed the chat data using argument-mining techniques that combine NLP approaches such as the language model BERT with machine learning methods. In particular, we measured argument use on the chat-message level. We did so by training a machine learning algorithm on a subset of chat data as well as on textual data generated by participants who did not chat but wrote down one argument in favor of rent control and one argument against.[5] We then applied the trained algorithm to the overall data set. It assigned each chat message a probability of being an argument, instead of a claim only, with an accuracy of 91 percent.[6] We also used human coders as a robustness check.

26 percent of our sample changed their opinion. Opinion composition of the chat group matters. The more chat partners in the group who share her views, the

[3] See DellaVigna and Gentzkow (2010) for an overview on the empirical literature on persuasion that contains a section on persuading voters.

[4] For a sociological approach on why economists tend to disagree on economic issues, see Marietta and Perlman (2000).

[5] Thereby, we make methodological contributions to the burgeoning economic literature using textual analysis (Gentzkow et al. 2019).

[6] To be more precise, the algorithm was trained to detect arguments versus messages that contained no argument, which includes both unjustified claims and unrelated messages such as greetings. Chat discussions were very focused on the topic, though; hence off-topic messages, except greetings, were rare. Importantly, this method of tagging chat messages does not presume an external standard of argument use (e.g., expert opinions on rent control). Instead, it relies on purely internal standards implicitly applied by participants themselves.

less likely a participant is to change her opinion. This is in line with echo chambers forestalling the convergence of opinions in societies. Interestingly, opponents to the rent-control initiative, which are the minority in our sample, are significantly more likely to use arguments than the initiative's supporters. Accordingly, arguments affect opinion change only for those who were initially in favor of rent control. That is, participants are more likely to change to voting against rent control when their chat partners expressed more anti-rent-control arguments. Together, these two countervailing effects - the majority of supporters of rent control confirming each other and the minority of opponents using persuasive arguments - could explain why average voting behavior does not differ between the chat treatment and a control treatment without chat. In sum, opinion change, in the direction of anti-rent-control, seems to be driven by persuasive arguments; however, the more chat partners with congruent prior opinions in the group, the less likely a participant is to change opinions. Hence, we find evidence for a double standard consistent with self-serving biases: While peers who share our opinions make us more entrenched regardless of whether they use arguments, peers who oppose our opinions need to use arguments to convince us to change our views. A majority, however, insulate their existing opinions against countervailing arguments. This may be analogous to the asymmetric response to good news vs. bad news found in previous experiments on individual updating (Eil and Rao 2011; Möbius et al. 2011) whereby good news (peers in agreement in our case) is easily incorporated in the belief updating whereas bad news (peers in opposition) is more likely to be ignored.

We ask what effect anonymous online exchanges have on opinions and votes on an important economic issue and through which channels. Our setting is a typical one for public discourse in this digital age. We provide field evidence that reconciles two strands of literature, one showing that individuals weigh information communicated by others according to signal quality and another which suggests that people place too much weight on information that confirms their priors. Moreover, we bridge the gap between commonplace discussions, for which precise measures of information quality are hard to obtain, and experimental explorations of models with precise signal quality. Our results highlight an important cost of political polarization: when confronted with strong enough arguments from an opposing side, people do change their opinions, but such exchanges are becoming increasingly rare since people are left alone in their echo chambers.

2 Experimental Design

The online survey experiment was conducted in two waves around the Local Rent Control Initiative ballot on the November 6, 2018 California election. With this ballot, Californians could vote in favor or against Proposition 10 that expands local governments' authority to enact rent control in their communities. Wave 1 started eight days before the ballot, i.e. on October 29th, and was terminated on November 5th. The second wave started ten days after the ballot on November 16th and was terminated on December 4th. Recruitment and payment of subjects was delegated to Respondi.[7] The experiment itself was programmed in o-tree and conducted by the WISO laboratory at Hamburg University.

The surveys from wave 1 and 2 were both filled out by participants online. Wave 1 comprises 80 questions and elicits participants' voting intentions (likeliness to vote in favor from *very likely* to *not at all likely*), prior beliefs about the effects of rent control, participants' media consumption, their written arguments in favor of or against rent control, and socio-demographic information (e.g. age, gender, and whether they were renting, renting out, or owning a house). Wave 2 comprises 20 questions and elicits subjects' final votes, the importance of economic-, liberty-, and fairness-based arguments in their voting decision, and questions about participation in other ballot questions.

Half of all subjects were randomly invited to participate in an online chat to discuss the pros and cons of rent control at the end of wave 1, i.e., prior to the actual ballot. Subjects were assigned randomly to chat-groups of five individuals. This random assignment allows us to analyze the effects of opinions and arguments in the chat groups on opinion change and voting behavior in a causal way. The chat environment was similar in design to WhatsApp, a chat platform likely familiar to most of our subjects.

3 Data

In total, 2934 subjects participated in wave 1 of our online survey experiment (Compare Figure A1). At the end of wave 1, 2404 of those participants were randomly invited to chat, leaving 530 uninvited. Participants had to wait in a digital waiting room until five subjects could be grouped for a chat. Chat invitations were oversampled compared to non-invitations because we required that chat-groups always start with five subjects resulting in some chat-groups not being formed due to time delays. In our case, 1278 subjects were allocated to *NoChat* because the chat-group could

[7] www.respondi.com

not be formed. Thus, 1126 subjects ended up in 264 chats. In some cases, chat-groups only contained four or three subjects because subjects left the experiment. Chats lasted on average 10.7 minutes and created 6415 messages. Out of the 1808 subjects assigned to *NoChat*, 817 subjects participated in wave 2, while 743 out of the 1126 chatters participated in wave 2. Attrition between wave 1 and 2 amounts to 1374 participants. In Appendix C we show that our results are robust to this selection effect using a Heckman selection procedure.

From the 1560 subjects that participated in both waves, we excluded 54 subjects because they stated that they already voted before the survey (early voters), leaving us with 1506 subjects in both waves. In wave 2, 704 (47%) subjects stated that they voted in favor of rent control and 586 (39%) stated that they voted against rent control. Finally, 216 (14%) subjects declined to answer this question. The sample is thus significantly more in favor of rent control than the actual outcome of the ballot (41% in favor and 59% against rent control).[8]

Table 1 and Table A1 summarize some descriptive statistics about our participants. Importantly, 44% of subjects are renters and only 11% are renting out. With regard to personal experience with rent control 16% state that they live or lived in a rent-controlled area. Regarding party affiliation, 45% of subjects describe themselves as Democrat while 22% see themselves as Republican and 29% as Independent. With regard to voter turnout, Table 2 provides some details for our sample. It

Table 1: Summary statistics

	Mean			Overall		
Variable	Overall	NoChat	Chat	St. Dev.	Min	Max
female	0.63	0.63	0.63	0.48	0	1
number of children	1.07	0.98	1.17	1.22	0	4
renting	0.44	0.45	0.44	0.50	0	1
renting out	0.11	0.10	0.13	0.32	0	1
republican	0.22	0.20	0.23	0.41	0	1
democrat	0.45	0.46	0.45	0.50	0	1
independent	0.29	0.28	0.29	0.45	0	1
lived rent controlled before	0.16	0.16	0.15	0.36	0	1
chat participation	0.48	0.00	1.00	0.50	0	1

Notes: The table summarizes key statistics from the survey questions from Wave 1.

indicates that the vast majority of subjects followed through with their plan of casting a ballot (81%). A few subjects did not plan to vote and actually did not vote (4.8%) and even fewer who planned to vote did not show up (3.9%). Moreover, we see only minor differences in voter turnout for subjects who participated in the chat and those who did not.

[8]See `https://ballotpedia.org/California_Proposition_10,_Local_Rent_Control_Initiative_(2018)`

Table 2: Voter turnout

Type	Overall	NoChat	Chat
PlannedNoShow	71 (4.8%)	34 (2.3%)	37 (2.5%)
PlannedShow	1210 (81.0%)	613 (41.0%)	597 (40.0%)
UnplannedNoShow	58 (3.9%)	40 (2.7%)	18 (1.2%)
UnplannedShow	14 (0.9%)	7 (0.5%)	7 (0.5%)
UnsureNoShow	78 (5.2%)	48 (3.2%)	30 (2.0%)
UnsureShow	63 (4.2%)	36 (2.4%)	27 (1.8%)

Notes: The table displays frequencies and percentages of planned and actual voting behavior, i.e. planning to vote or not and showing up or not. Twelve subjects did not answer.

4 Methods

4.1 Opinion Change

Since we are interested in whether and how the chat discussions change the participants' opinions, we construct the following opinion change variables using subjects' answers to the question on how likely they will vote in favor of rent control, gathered in wave 1, and their reported actual votes in wave 2. If a subject states that she is *very likely* or *pretty likely* to vote in favor of rent control but finally voted against it, she is typed a *YesNo* opinion changer. Similarly, a subject who claimed to be *not that likely or not at all likely* to vote in favor of rent control but finally did so is defined as type *NoYes*. Those that first claim to be *neither not likely nor likely* to vote in favor and finally voted against or in favor of rent control are defined as types *UnsureNo* and *UnsureYes*, depending on their final vote. Importantly, we do not classify these latter two as opinion changers since they had no preconceived opinion to begin with.[9] All other subjects who do not change their opinion are defined as *NoNo* and *YesYes* types. We then aggregate all types of opinion changers to a binary opinion change variable (*opinion_change_bin*) and a categorical opinion change variable with three categories (*opinion_change_cat*). Table 3 summarizes the construction of both variables and contains the frequencies of all types. A majority of 74% did not change their opinion, while 10% changed their opinion to a No-vote. Both measures are subsequently used in binary and multinomial regressions to investigate the effect of chat content and composition of prior opinions in the chat on opinion change.

Our third variable of interest is the distance of a subject's actual voting behavior, i.e., her reported voting decision in wave 2, to her prior voting intention. This distance variable is constructed as follows. First, we normalize the prior voting intention, i.e. the likeliness of voting in favor of rent control, to a range of -1 (*not at*

[9] All main results are robust to including the unsure types as opinion changers.

Table 3: Construction of opinion change

Type	Frequency	Opinion_change_bin	Opinion_change_cat
NoNo	360 (30%)	No_change (=0)	No_change (Cat. 1)
YesYes	535 (44%)		
YesNo	117 (10%)	Change (=1)	Ch_to_No (Cat. 2)
NoYes	47 (4%)		Ch_to_Yes (Cat. 3)
UnsureNo	73 (6%)		
UnsureYes	69 (6%)		

Notes: Due to missing values, *opinion_change_bin* and *opinion_change_cat* cannot be calculated for 305 (20%) subjects. In the following, No_change (Category 1) serves as the benchmark in our multinomial regression analysis.

all likely) to 1 (*very likely*). Second, we re-label actual voting behavior to -1 (against rent control) and 1 (in favor of rent control). Third, we subtract the normalized prior voting intention from the re-labeled actual voting behavior. Finally, we divide the resulting measure by two to construct a normalized measure that ranges from -1 to 1. We denote this variable as *opinion_change_dist* and Figure 1 illustrates its distribution. The variable *opinion_change_dist* measures the magnitude of a subject's

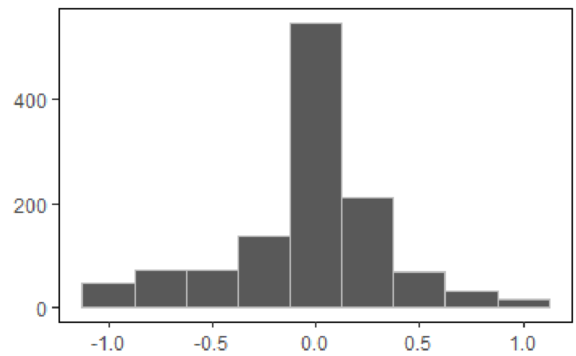


Figure 1: Distributions of *opinion_change_dist*

opinion change. For instance, consider a subject who is a priori *pretty likely* to vote in favor of rent control but ends up voting against it. This change of opinion is stronger in magnitude (value -0.75) than a subject who is a priori *not that likely* to vote in favor and finally votes against rent control (value -0.25). Overall, individuals who followed through with their clear ex-ante voting intention are located in the middle of the scale at zero. Individuals who changed to a Yes-vote are located in the positive domain, and individuals changing to a No-vote are located in the negative domain. It is important to note that, unlike our first two opinion change variables, this distance variable explicitly considers participants who are ex-ante unsure how to vote. An unsure individual can either vote in favor of rent control, receiving the distance value 0.5, or vote against, receiving the distance value -0.5.

4.2 Pre-chat Positions on Rent Control

The heterogeneity of opinions about rent control among chat partners is a potentially important influential factor in the chat discussions that affects an individual's voting decision. We therefore construct a pre-chat position measure from the answers to the question how likely a subject will vote in favor of rent control. We label a subject stating that she will *very likely* or *pretty likely* vote in favor of rent control as having the *Position* equal to Yes $(= 1)$. Similarly, we label a subject who is *not that likely* or *not at all likely* to vote in favor of rent control as having the *Position* equal to No $(= -1)$. For each such subject we calculate the number of opposing positions minus the number of aligned positions from all peers matched to her in one chat group. More formally, for an individual i that is a priori against rent control, we calculate $\sum_{j=1}^{n} Position_j$, with $j \neq i$, while n is the number of subjects in $i's$ chat group without subject i. For an individual who is a priori in favor of rent control, we calculate $(-1) * \sum_{j=1}^{n} Position_j$.

In other words, for each individual, we calculate the number of subjects opposing her in the chat minus the number of subjects who share her position and call it *diff_exante_pos*. This variable takes values between -4 and 4. For instance, consider an individual who is a priori in favor of rent control, and assume that one of her four chat partners is also in favor of it while three are against it. Then, we formally get: *diff_exante_pos*$= 3 - 1 = 2$. Thus, we account for the heterogeneity of positions in the chat conditional on a subject's prior position. This measure can be calculated for 598 subjects. 271 subjects face an "overweight" of aligned positions, 187 subjects face more opposing views than aligned views in the chat, and 140 experience a balanced chat group with regard to the other chat members' positions on rent control.

In using this distance measure, we impose a model in which any subject assigns the same weight to contrary opinions as to opinions aligned to her own. Hence, the model corresponding to this measure excludes confirmatory bias as well as observable heterogeneity in the quality of information. We additionally construct the variables *exante_pos_op* and *exante_pos_al* equal to the number of chat partners with opposing and aligned positions, respectively. These variables are used in separate regressions in order to test for potential biases in information processing.

Note that there is a shortcoming of these measures. A subject stating a positive or negative position towards rent control during the survey does not necessarily communicate this during the chat discussion.[10] If the decision to remain silent in the chat depends on the prior position, our measures are biased. Hence, we test for such a bias and re-do all analyses with re-constructed measures that do not take into

[10]see also Biermann, Hüning, and Mechtenberg (2021) for a similar finding.

account the silent subjects.

4.3 In-chat Argumentative Positions

Besides mere opinions on rent control, the argumentative discourse among chat partners also potentially affects an individual's voting decision. We measure the heterogeneity in argumentative positions towards rent control among chat partners by applying argument mining techniques (see Lippi and Torroni (2016) and Cabrio and Villata (2018) for an overview of the literature). Argument mining uses NLP and Machine Learning techniques to detect arguments, or components thereof, in natural language text. In the following, we summarize our procedure.[11]

First, a random forest classification model is trained to classify out-of-sample chat messages as containing an argument or not. An argument is defined as a message containing both a claim and a premise or a premise where the claim is implicit (Toulmin 1958, Walton 2009). Second, all chat messages detected in the first step as containing arguments are fed into a second algorithm predicting the position of that argument, i.e. pro or contra rent control. This results in raw probabilities for each argumentative message being in favor of or against rent control. Raw probabilities against rent control are multiplied by -1. Third, the sum of these modified raw probabilities measures an individual's average argumentative position on rent control. For instance, an individual formulates three arguments, two in favor of rent control and one against (modified raw probabilities are 0.6, 0.8 and -0.7). Then, her average argumentative position is 0.7 and positive, i.e. the individual argues more in favor of than against rent control.

Finally, the heterogeneity of argumentative positions among chat partners is summarized in the same way as the pre-chat positions on rent control. We thus calculate for each individual the strength of arguments that are opposing her position minus the strength of arguments that align with her position. More specifically, for an individual i who is a priori against rent control, we calculate $\sum_{j=1}^{n} ArgPosition_j$, with $j \neq i$, where n is the number of other subjects in $i's$ chat group. In contrast, for an individual that is a priori in favor of rent control, we calculate $(-1) * \sum_{j=1}^{n} ArgPosition_j$. We refer to this variable as *diff_arg_scores*. For instance, consider a subject who is a priori in favor of rent control, and suppose that two chat-partners argue more in favor than against rent control (Aligned Scores: 2.7, 3.1), while two others argue more against than in favor of rent control (Opposing Scores: 4.1, 0.5). Then, we formally get:

[11] For details of this procedure and methods see Appendix B.

$$diff\_arg\_score = Opposing\_Scores - Aligned\_Scores$$
$$= (4.1 + 0.5) - (2.7 + 3.1) = -1.2$$

Note that the variable *diff_arg_scores* is constructed to impose a rational-voter model: heterogeneity of individual message quality, measured by average argument score, is observable; and subjects weigh information that their matched peers pass on to them according to that quality only. Hence, biased information processing as in models of confirmatory bias is again excluded by assumption. In order to allow for biased information processing we construct the variables *arg_score_op* and *arg_score_al* that measure the argument strength of chat partners with opposing and with aligned arguments, respectively. These variables are used in separate regressions.

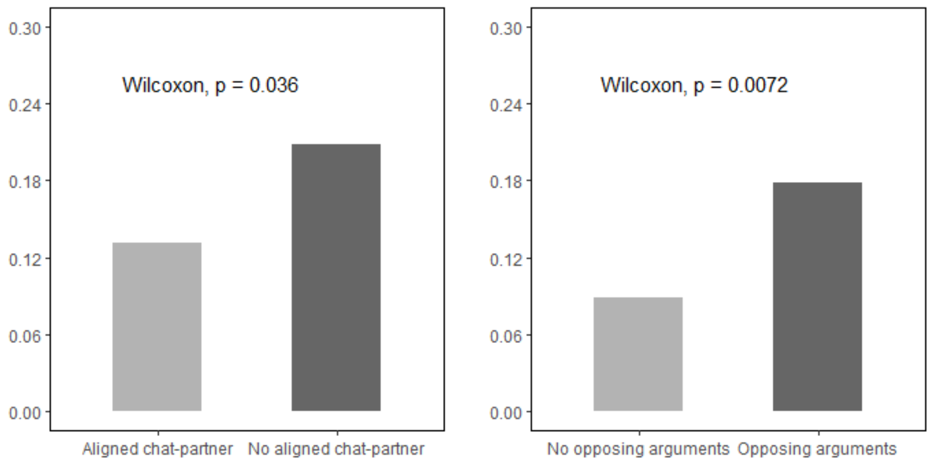
The two variables *diff_exante_pos* and *diff_arg_scores* are both constructed conditional on a subject's own prior position on rent control. This is necessary for our investigation of opinion change since we consider changing opinions mutually in both directions. For our investigation of *opinion_change_dist*, however, we construct the same variables not conditioning on a subject's own position. The variable *exante_pos_avg* (*arg_score_avg*) measures the number of (argumentative) positions in favor of rent control minus the number of (argumentative) positions against rent control by the subject's chat partners.

Finally, before presenting our regression results, we illustrate Spearman's correlations of all independent variables used in the regression analysis in Table A3. For the number of opposed and aligned positions within one chat group, i.e. *exante_pos_op* and *exante_pos_al*, one would expect a strong negative correlation since the total number of chat participants per group is fixed at five. The correlation coefficient, however, is rather moderate at -0.25 (column 12). This has two reasons. First, some chat participants state that they are unsure how to vote and thus are not counted in the number of opposed and aligned positions within one chat group. Second, if one of individual i's chat partner did not answer the question with regard to her ex-ante voting intention, the variables *exante_pos_op* and *exante_pos_al* are still calculated for all non-missing chat partners.

5 Results

We start by highlighting a key result (Figure 2). Participants who face at least one chat-partner with an aligned prior position on rent control are more reluctant to change opinion compared to the rest (left panel: 13% vs. 20%; MWU, p-value:

0.036). By contrast, participants who face at least one chat partner providing arguments against her prior position are twice as likely to change their opinion compared to the rest (right panel: 8.9% vs. 17.8%; MWU, p-value: 0.007). Participants get more entrenched in their positions when meeting like-minded people, but opposing arguments can convince them to change their opinion.



(a) Chat-partners with and without aligned views

(b) Chat-partner with and without opposing arguments

Figure 2: Opinion change (in %) by aligned chat partners and opposing arguments

We first investigate the determinants of any opinion change (*opinion_change_bin*). Next, we distinguish between the two directions of change, i.e. from being a priori in favor of rent control to a No-vote and vice versa (*opinion_change_cat*). We also investigate opinion change as the distance between an individual's prior position and final vote (*opinion_change_dist*). In all cases, we present results for the overall dataset (All) and a subsample with only chat participants (Chat).

5.1 Opinion Change - Binary

Table 4 summarizes the results for our simplest measure of opinion change, i.e. comparing those who changed opinion versus those who did not (*opinion_change_bin*).

First, we do not find evidence that chat participation per se has an effect on opinion change (column 1).[12] Second, the more a subject is confronted with positions in her chat group that are against her own, measured by *diff_exante_pos*, the higher are

[12] As Figure A1 highlights, 991 subjects participated in wave 1 (in NoChat) but did not participate in wave 2. As a robustness check for our null finding , we add those subjects to the regressions and assume that they did not change their opinion. Results, depicted in Table A10, indicate that correcting for this selection effect does not change our overall results.

the odds that she will change her opinion. Message quality matters, too: the more opposing arguments relative to aligned arguments a subject faces from chat partners, the higher the odds of changing opinion (columns 3). After adding *diff_arg_score* to the regression with *diff_exante_pos*, the effect of the latter gets reduced and insignificant (column 4). A Sobel's test indicates that this reduction is significant (p-value: 0.017). Hence, chat partners using arguments may mediate the effect of the chat-group composition on opinion change. This mediation effect is validated in more detail in Appendix B.

In columns 5 to 7, we decompose *diff_exante_pos* and *diff_arg_score* into their respective components. While the variable *exante_pos_op* (*exante_pos_al*) denotes the number of chat partners with opposed (aligned) views, the variables *arg_score_op* and *arg_score_al* denote opposing and aligned argument strength of chat partners, respectively. Here, it becomes evident that the number of aligned positions and the quality of opposing arguments determine the tendency to change opinion: while the number of aligned views decreases the odds of changing opinion regardless of arguments, the strength of opposing arguments increases the odds of changing opinion. In other words, only the strength of opposing arguments can counteract potential confirmatory bias in our setting.

Up to now we did not focus on the direction of opinion change. However, Table 5 reveals systematic and relevant differences between subjects who are a priori in favor and subjects who are a priori against rent control. Most importantly, compared to subjects who are a priori against rent control, subjects a priori in favor are more reluctant to express their opinion in the chat and much less likely to use arguments.[13] Hence, our findings on the effects of chat composition and argument use may mask some asymmetries in opinion change between the two sides. Therefore, we now move on to investigate opinion change in each direction separately.

5.2 Opinion Change - Directional

Table 7 summarizes the results that inform us on opinion change splitting those that do change opinion into those who change to *No* (*Ch_to_No*) and those who change to *Yes* (*Ch_to_Yes*).

First, consistent with our previous finding, we find no evidence that chat participation per se affects opinion change in either direction (columns 1 and 2). Second, we investigate how the chat partners' initial positions on rent control and their argumentative strength in the chat affect individual opinion change when we im-

[13] Moreover, compared to those a priori in favor, those a priori against hold the other subjects' understanding of the issue of rent control in lower esteem and have less respect for those who change their opinion. They also have higher income and are less likely to be renters.

Table 4: Opinion change (Binary)

	All	Chat	Chat	Chat	Chat	Chat	Chat
chat	−0.080						
	(0.174)						
diff_exante_pos		0.186		0.124			
		(0.078)		(0.084)			
diff_arg_score			0.157	0.124			
			(0.056)	(0.060)			
exante_pos_op					−0.002		−0.095
					(0.147)		(0.161)
exante_pos_al					−0.380		−0.333
					(0.150)		(0.156)
arg_score_op						0.165	0.159
						(0.080)	(0.088)
arg_score_al						−0.146	−0.081
						(0.107)	(0.108)
Constant	−1.459	−1.621	−1.664	−1.634	−1.268	−1.667	−1.243
	(0.330)	(0.633)	(0.633)	(0.637)	(0.669)	(0.633)	(0.676)
Obs.	1,039	518	518	518	518	518	518
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Akaike Inf. Crit.	905.551	440.111	437.816	437.664	439.760	439.800	439.045

Notes: The table reports results of binomial regressions with *opinion_change_bin* as the dependent variable. The variable *chat* is a dummy equal to one for chat participants and zero otherwise. The variable textitdiff_exante_pos denotes the chat composition an individual faces (opposed minus aligned views of chat-partners) and *diff_arg_score* denotes the argumentative positions an individual faces in the chat (opposed arguments minus aligned arguments of chat-partners). Furthermore, *exante_pos_op* (*exante_pos_al*) denotes the number of chat partners with opposed (aligned) views. The variable *arg_score_op* (*arg_score_al*) denotes the sum of opposing (aligned) arguments of chat partners. The regressions include the following *Controls*: The variable *value_ opinionchange* denotes an individual's attitude towards opinion change and *diff_understand_rentcon* denotes an individual's perceived understanding of rent control compared to the average understanding. The variable *female* (*renting out*) is a dummy that is equal to one for female subjects (subjects that rent out property) and zero otherwise. The variable *age* reflects eight age categories ranging from "18-24" to "85 or older". The variable *info_yes_camp* (*info_no_camp*) is a dummy indicating if a subject received information from the Yes (No) campaign and zero otherwise. *chat_length* controls for the length of the chat in minutes. Log odds are reported as coefficients and standard errors are in parentheses.

Table 5: Characteristics by Prior

	Against Mean	In Favor Mean	MWU p-value
diff_understand_rentcon	1.16	0.68	0.00
understand_rentcon	3.65	3.61	0.20
understand_rentcon (others)	2.49	2.93	0.00
opinion_expressed_bin	0.81	0.75	0.07
opinion_expressed_count	1.93	1.45	0.00
abs_argument_strength	1.02	0.73	0.00
school_educ	3.58	3.47	0.11
female	0.61	0.63	0.81
age	3.89	3.19	0.00
value_opinionchange	7.00	7.54	0.00
democrat	0.32	0.58	0.00
republican	0.36	0.16	0.00
independent	0.31	0.24	0.10
household_inc	8.49	7.06	0.00
renting	0.26	0.54	0.00
renting_out	0.16	0.09	0.00

Notes: The table displays means by prior voting intention. Additionally the p-value of a MWU-test is presented. Variables are as described in Table 4. Moreover, *opinion_expressed_bin* is equal to one if the participant expressed her opinion in the chat and zero otherwise. *opinion_expressed_count* is the corresponding count variable, i.e. how often participants expressed their opinion.

Table 6: Characteristics by Opinion Change Type

	No Change Mean	Opinion Change Mean	MWU p-value
diff_understand_rentcon	0.89	0.76	0.03
understand_rentcon	3.65	3.47	0.01
understand_rentcon (others)	2.77	2.72	0.90
opinion_expressed_bin	0.81	0.68	0.00
opinion_expressed_count	1.73	1.46	0.04
abs_argument_strength	0.84	0.78	0.11
school_educ	3.53	3.49	0.69
female	0.62	0.64	0.39
age	3.48	3.36	0.62
value_opinionchange	7.35	7.29	0.95
democrat	0.48	0.49	0.71
republican	0.23	0.27	0.23
independent	0.28	0.22	0.11
household_inc	7.56	7.72	0.55
renting	0.43	0.45	0.69
renting_out	0.11	0.12	0.87

Notes: The table displays means by opinion change type, i.e. no change versus change. Additionally the p-value of a MWU-test is presented. Variables are as described in Table 5.

pose equal weights on aligned and opposing positions and arguments. That is, we regress opinion change on *diff_exante_pos* and *diff_arg_score*. We re-constructed *diff_exante_pos* to account for the asymmetry in opinion expression between subjects a priori in favor and subjects a priori against rent control, and excluded subjects who remained fully silent in the chat from the construction of this variable. We do not find significant effects of the opinion composition of chat groups under the equal-weights assumption (columns 3 and 4).

By contrast, when considering arguments, we find that they do matter, but only for opinion changes toward voting *No* (column 3): the difference in peers' argumentative position measured by *diff_arg_score* increases the odds of being an opinion changer of type *Ch_to_No* (column 3), but not of type *Ch_to_Yes* (column 4).

Intuitively, this asymmetry in the effects of *diff_arg_score* is expected, given that subjects a priori in favor of rent control are making much fewer arguments than those a priori against. If the opposing side refrains from using arguments, opinion change in that direction is unlikely to be triggered by said arguments. In addition, subjects who are initially against rent control also exhibit higher confidence in understanding the issue, compared to both subjects who are initially in favor of rent control and those who are unsure (Wilcoxon-rank-sum tests, p-values: 0.001). There is weak evidence that confidence tends to reduce the odds of changing opinion (Table A4, column 1). Hence, subjects initially opposed to rent control may be less open to arguments against their position than subjects initially in favor.

Next, we split the two variables *diff_exante_pos* and *diff_arg_score* into their directional components and test these variables in separate regressions (columns 5 and 6), as we did before for the binary regressions, but with *exante_pos_al* and *exante_pos_op* re-constructed dropping subjects who remained fully silent in the chat. Hence, we now drop the assumption that subjects assign equal weights to positions or arguments opposed to their own position and those aligned to it, while again taking into account the asymmetric chat behavior of the two sides of the debate. The results reveal that the more opinions aligned with her own a subject encounters in her chat group (*exante_pos_al*), the lower the subject's odds of changing her opinion. This effect holds true for both directions of opinion change. In contrast, we do not find any effect of opposing positions on opinion change. Instead, we again find that stronger opposing *arguments* (*arg_score_op*) increase the odds of changing to a *No*-vote, though still not the odds of changing to a Yes-vote. Together, these findings indicate a confirmatory bias: opinions of peers are effective only if aligned with the subject's own opinion; then, they insulate the latter against opinion change. Opinions opposing the subject's own, however, need to be backed up by arguments.

Table 7: Opinion change (Multinomial)

	All	All	Chat	Chat	Chat	Chat
	Ch_to_No	*Ch_to_Yes*	*Ch_to_No*	*Ch_to_Yes*	*Ch_to_No*	*Ch_to_Yes*
chat	0.001	−0.288				
	(0.202)	(0.309)				
diff_exante_pos			0.167	−0.001		
			(0.109)	(0.182)		
diff_arg_score			0.170	0.018		
			(0.069)	(0.107)		
exante_pos_op					−0.126	−0.308
					(0.186)	(0.328)
exante_pos_al					−0.411	−0.707
					(0.194)	(0.342)
arg_score_op					0.220	0.069
					(0.100)	(0.181)
arg_score_al					−0.141	0.159
					(0.138)	(0.161)
Constant	−1.413	−3.798	−1.434	−4.474	−1.286	−4.365
	(0.378)	(0.622)	(0.715)	(1.269)	(0.727)	(1.287)
Obs.	1039	1039	518	518	518	518
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Akaike Inf. Crit.	1,096.652	1,096.652	535.991	535.991	535.674	535.674

Notes: The table reports results of multinomial regressions with *opinion_change_cat* as the dependent variable. All independent variables are as described in Table 4. Log odds are reported as coefficients and standard errors are in parentheses.

5.3 Opinion change - Distance to Prior

Finally, we investigate how chat composition and argument strength affect not only the direction, but also the magnitude of opinion change, using our opinion-change distance variable. Results are summarized in Table 8. Remember that *exante_pos_avg* and *arg_score_avg* are unconditional on an individual's prior position, i.e. high values of these variables indicate more chat partners being in favor of rent control and higher argument strength in favor of rent control, respectively.

Consistent with our previous findings, we find that the more the average position of the partners is in favor of rent control, measured by *exante_pos_avg*, the stronger is the move to a Yes-vote. Regarding argumentative strength, we also find a positive effect, i.e. the higher the argumentative strength of chat partners in favor of rent control, the more likely a subject changes to a Yes-vote. When considering both effects simultaneously, however, only *exante_pos_avg* remains weakly significant.

Table 8: Opinion change (Distance to prior)

	All	Chat	Chat	Chat
chat	−0.010			
	(0.022)			
exante_pos_avg		0.025		0.020
		(0.010)		(0.011)
arg_score_avg			0.013	0.008
			(0.006)	(0.007)
Constant	−0.043	−0.081	−0.072	−0.086
	(0.043)	(0.078)	(0.076)	(0.077)
Obs.	1,170	569	569	569
Controls	Yes	Yes	Yes	Yes
R^2	0.005	0.016	0.013	0.018
F Statistic	0.789	1.023	0.799	1.045

Notes: The table reports results of OLS regressions with *opinion_change_dist* as the dependent variable. All independent variables are as described in Table 4. Furthermore, the variable *exante_pos_avg* denotes the chat composition of chat-partners only (positions in favor of minus positions against rent control) and *arg_score_avg* denotes the argumentative positions of chat-partners only (arguments in favor of minus arguments against rent control). Heteroskedasticity-consistent standard errors are reported in parentheses.

5.4 Robustness Checks

We perform the following robustness analyses to check the validity of our results. First, as Figure A1 indicates, 383 subjects participated in wave 1 and the chat discussions but decided to not participate in wave 2. In order to check if this attrition affects our results, we add those subjects and assume that they did not change opinion, i.e. we set their reported vote in wave 2 equal to their initial voting intention in wave 1, and rerun our regressions. Results are depicted in Table A7 to Table A9 and indicate that our findings are robust to the inclusion of those subjects. As a more sophisticated approach we train a random forest on participants that took part in both waves (separately for Chat and NoChat participants) and perform an out-of-sample prediction for voting behavior on those who only participated in wave 1.[14] The two algorithms predict 20% (NoChat) and 33% (Chat) opinion changes which is substantially different from our heuristic that assumes no opinion change of wave 1 only participants. Finally, we add those predictions and rerun our regressions. Results depicted in Table A11 to Table A13 indicate that our results are robust to the inclusion of participants who only participated in wave 1. Moreover, we perform a Heckman selection procedure to account for these selection effects (Appendix C). Results, depicted in Table C2 to Table C4, indicate that our results are robust to these selection effects.

Second, for the construction of *diff_arg_score* we used Machine Learning and NLP technqiues to detect arguments and their positions in chat-messages. As a robustness check, we use the agreement of three manual annotations for all chat-messages instead of the Machine Learning predictions. Results depicted in Table A14 to Table A16 indicate that our findings are not a result of the ML exercise. Our main findings remain robust using simple manual annotations.

Third, although chat discussions always started with five subjects in each chat group, some chats suffered from dropouts. Overall, 50% of chat groups remained at the size of five, 40% went to four, 9% went to three, and 1% of the chat groups ended up containing only two participants. As a robustness check, we rerun our regressions only with those chat groups with four or five participants. Result in Table A17 to Table A19 indicate that our results are robust to the exclusion of chats with fewer than four participants.

Fourth, some subjects might rush through the survey and pay little attention. Others might not complete the survey in one turn but are busy or distracted doing other things online in the meantime. Hence, as a robustness check we remove those subjects who belong to the 10% fastest or the 10% slowest subjects (wave 1). As Table A20 to Table A25 show, our main results that subjects are less likely to change opinion the more chat partners confirm them in their initial position and the more likely to change opinion (in the direction to a No-vote) the stronger opposing arguments they encounter are robust to the exclusion of the fastest 10% of subjects.

[14] Accuracy of the two models are 87% (Chat) and 90% (NoChat) indicating that voting behavior can be predicted reasonably well.

Removing the slowest 10%, results with regard to argument strength are less robust in the case of binary regressions of opinion change.

Fifth, for the opinion-change regressions in Table 4 we chose a binary regression model with a logistic link function that converts the linear combination of the independent variables to a scale of probabilities as well as a multinomial regression model in Table 7. Our results, however, are also robust using a simple linear probability model (OLS). In the case of the multinomial model, one category is modeled against the remaining two categories. Results are available upon request.

Finally, with regard to our distance-to-prior opinion change variable, we chose simple OLS regressions (Table 8). The dependent variable, however, might not be sufficiently normal to justify OLS. As a robustness check, we perform ordered logit regressions and find virtually the same results (Table A26), i.e. an individual is the more likely to switch to a Yes-vote, the more the average position of her chat partners is in favor of rent control (measured by *exante_pos_avg*).

6 Discussion and Conclusion

In this paper, we studied whether and why randomized chat groups trigger opinion change in voters, moving them toward voting the opposite of what they intended. We focused on the opinion composition of chat groups and the mediating effect of argument use as potential explanations. Our study was conducted in the context of the Local Rent Control Initiative on the 2018 California ballot. Half of the subjects of our online survey experiment had the chance to discuss the topic of rent control in randomized chat-groups of up to five individuals. We measured the aligned and opposing views in the group prior to the chat as well as the argumentative strength in favor and against rent control that is expressed in the chat. For the latter, we used Machine Learning techniques together with a state-of-the-art language model to automatically detect argumentative reasoning in chat messages.

Our main finding is that arguments against rent control communicated during the chat discussions convince subjects to vote accordingly, i.e. against the Initiative that was on the ballot. In contrast, we do not find that initial opponents of rent control are convinced by arguments of proponents to vote in favor of the Initiative. We argue that this asymmetry is likely due to both the higher tendency of those initially opposing rent control to use arguments, compared to those who initially support it, and to the opponents' higher confidence, potentially resulting in less openness to arguments contrary to their own position. Moreover, the chat composition, i.e. the number of aligned and opposing positions in the chat, affects an individual's decision to change opinion on this matter. More specifically, the more chat partners who are in line with an individual's prior voting intention, the less likely she will change opinion. We attribute this effect to confirmatory bias.

We assumed no systematic misreporting of prior voting intentions and actual votes in our analysis. It is implausible that our results could have been generated by such misreporting. Let us consider several possible ways participants could misrepresent their voting behavior. Suppose people who misreported voting No because they wanted to be on the winning side after the election results actually stuck to their intended Yes vote or did not vote at all. Such behavior would not explain why we find that argument strength drives the opinion change from Yes to No. Similarly, imagine participants misrepresenting their initial intention to vote No by saying they would vote Yes because they thought that was the socially acceptable choice, but truthfully reporting that they did vote No. Then what we mistakenly observed as changes from Yes to No should again be independent of opinion composition or argument strength in the group because it was the actual election result that revealed the social acceptability of voting No, not what happened in the chat. It is also unlikely that participants who intended to vote Yes thought that voting No was the socially acceptable choice and misreported that as their intention since the Yes voters are the overwhelming majority in our sample and the No voters were the defensive ones with many arguments in the chats.

Overall, we find that the fundamental assumption underlying the literature on information aggregation and deliberation captures an important part of reality: People do let peers persuade them; and they do account for expertise, which they assess from their peers' arguments. However, the picture does become more complicated with an interesting double standard: people get confirmed in their initial belief by like-minded peers who do not use arguments, while arguments are needed to make countervailing positions more persuasive.

We were able to offer a discussion platform with random, neutral group matching to our participants. Many of the prominent online platforms, however, use biased matching algorithms that tend to group together like-minded peers. Our findings suggest a two-fold effect of such matching bias: it not only induces people to get more entrenched in their own beliefs, but also likely lowers the standards of discussion since influencing like-minded peers does not seem to require the use of arguments. Hence, there could be long-term consequences, beyond belief polarization, for the ability of citizens to form and weigh arguments from others.

Our intervention with random chat assignments can be easily scaled up to large numbers of citizens. Take for instance the popularity of voting advice applications such as the *Wahl-O-Mat* in Germany that was requested 21 million times before the

German federal elections in 2021.[15]. The *Wahl-O-Mat* elicits parties' policy platforms and users' policy preferences and then informs voters about the party closest to their preferences. It would be easy to implement chat invitations at the very end of this application and allow voters to discuss the most controversial topics in randomly formed discussion groups. Such an implementation could scale up citizens' interactions with views and arguments outside their echo chambers.

[15]The *Wahl-O-Mat* is usually available approx. four weeks before a federal or state election. Archived versions of each *Wahl-O-Mat* can be accessed online. Here the archived version for the last German federal election in 2021: `https://www.bpb.de/themen/wahl-o-mat/45484/archiv/`, accessed on the 25/02/2022.

References

Ahler, D. J. and G. Sood (2018). The parties in our heads: Misperceptions about party composition and their consequences. *The Journal of Politics 80*(3).

Bail, C. A., L. P. Argyle, T. W. Brown, J. P. Bumpus, H. Chen, M. B. F. Hunzaker, J. Lee, M. Mann, F. Merhout, and A. Volfovsky (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences 115*(37), 9216–9221.

Balietti, S., L. Getoor, D. G. Goldstein, and D. J. Watts (2021). Reducing opinion polarization: Effects of exposure to similar people with differing political views. *Proceedings of the National Academy of Sciences 118*(52).

Blume, A., E. K. Lai, and W. Lim (2020). Strategic information transmission: A survey of experiments and theoretical foundations. In C. M. Capra, R. Croson, M. Rigdon, and T. Rosenblat (Eds.), *Handbook of Experimental Game Theory*, pp. 311–347. Northampton, Massachusetts: Edward Elgar Publishing.

Breiman, L. (2001). Random forests. *Machine Learning 45*(1), 5–21.

Buechel, B. and L. Mechtenberg (2019). The swing voter's curse in social networks. *Games and Economic Behavior 118*, 241–268.

Cabrio, E. and S. Villata (2018). Five years of argument mining: a data-driven analysis. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, Stockholm, Sweden, pp. 5427–5433.

Chakraborty, A. and R. Harbaugh (2010). Persuasion by cheap talk. *American Economic Review 100*(5), 2361–2382.

Charness, G. and C. Dave (2017). Confirmation bias with motivated beliefs. *Games and Economic Behavior 104*, 1–23.

Crawford, V. (1998). A survey of experiments on communication via cheap talk. *Journal of Economic Theory 78*(2), 286–298.

Crawford, V. P. and J. Sobel (1982). Strategic information transmission. *Econometrica: Journal of the Econometric Society 50*(6), 1431–1451.

DellaVigna, S. and M. Gentzkow (2010). Persuasion: Empirical Evidence. *Annual Review of Economics 2*(1), 643–669.

Devlin, J., M. Chang, K. Lee, and K. Toutanova (2018). BERT: pre-training of deep bidirectional transformers for language understanding. *Computer Science abs/1810.04805*, 4171–4186.

Eil, D. and J. M. Rao (2011). The good news-bad news effect: Asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics 3*(2), 114–138.

Fréchette, G. R., A. Lizzeri, and J. Perego (2018). Rules and commitment in communication. Working Paper 26404, National Bureau of Economic Research.

Gale, D. and S. Kariv (2003). Bayesian learning in social networks. *Games and Economic Behavior 45*, 329–346.

Gentzkow, M., B. Kelly, and M. Taddy (2019). Text as data. *Journal of Economic Literature 57*(3), 699–746.

Gerardi, D. and L. Yariv (2007). Deliberative voting. *Journal of Economic Theory 134*(1), 317–338.

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

Golman, R., D. Hagmann, and G. Loewenstein (2017). Information avoidance. *Journal of Economic Literature 55*(1), 96–135.

Green, J. R. and N. L. Stokey (2007). A two-person game of information transmission. *Journal of Economic Theory 135*, 90–104.

Grimm, V. and F. Mengel (2020). Experiments on belief formation in networks. *Journal of the European Economic Association 18*(1), 49–82.

Haaland, I., C. Roth, and J. Wohlfart (2022). Designing information provision experiments. *Journal of Economic Literature (forthcoming).*

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica 47*(1), 153–161.

Huber, M. (2020). Mediation analysis. In *Handbook of Labor, Human Resources and Population Economics*, pp. 1–38. Springer.

Hüning, H., L. Mechtenberg, and S. W. Wang (2021). Detecting argumentative discourse in online chat experiments. Working paper, Hamburg University.

Iaryczower, M., X. Shi, and M. Shum (2018). Can words get in the way? the effect of deliberation in collective decision-making. *Journal of Political Economy 126*(2), 688–734.

Imai, K., L. Keele, and D. Tingley (2010a). A general approach to causal mediation analysis. *Psychological Methods 15*(4), 309–334.

Imai, K., L. Keele, and D. Tingley (2010b). Identification, inference, and sensitivity analysis for causal mediation effects. *Statistical Science 25*(1), 51–71.

Kamenica, E. (2019). Bayesian persuasion and information design. *Annual Review of Economics 11*(1), 249–272.

Kamenica, E. and M. Gentzkow (2011). Bayesian persuasion. *American Economic Review 101*(6), 2590–2615.

Kendall, C., T. Nannicini, and F. Trebbi (2015, January). How do voters respond to information? evidence from a randomized campaign. *American Economic Review 105*(1), 322–53.

Lippi, M. and P. Torroni (2016). Argumentation mining: State of the art and emerging trends. *ACM Transactions on Internet Technology 16*(2), 10:1–10:25.

Marietta, M. and M. Perlman (2000). The uses of authority in economics. *American Journal of Economics and Sociology 59*(2), 151–189.

Möbius, M. M., M. Niederle, P. Niehaus, and T. S. Rosenblat (2011). Managing self-confidence: Theory and experimental evidence. Working paper, NBER.

Mossel, E., N. Olsman, and O. Tamuz (2016). Efficient bayesian learning in social networks with gaussian estimators. *2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton).*

Mossel, E., A. Sly, and O. Tamuz (2015). Strategic learning and the topology of social networks. *Econometrica 83*, 1755–1794.

Mueller-Frank, M. (2013). A general framework for rational learning in social networks. *Theoretical Economics 8*, 1–40.

Ortoleva, P. and E. Snowberg (2015). Overconfidence in political behavior. *American Economic Review 105*(2), 504–35.

Penczynski, S. (2019). Using machine learning for communication classifcation. *Experimental Economics 22*, 1002–1029.

Pew Research Center (2014). Political polarization in the american public. https://www.pewresearch.org/politics/2014/06/12/political-polarization-in-the-american-public/. Accessed: 2022-02-21.

Rabin, M. and J. L. Schrag (1999). First impressions matter: A model of confirmatory bias. *The Quarterly Journal of Economics 114*(1), 37–82.

Rinott, R., L. Dankin, C. Alzate Perez, M. M. Khapra, E. Aharoni, and N. Slonim (2015). Show me your evidence - an automatic method for context dependent evidence detection. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, Lisbon, Portugal, pp. 440–450. Association for Computational Linguistics.

Tingley, D., T. Yamamoto, K. Hirose, K. Imai, and L. Keele (2014). mediation: R package for causal mediation analysis. *Journal of Statistical Software 59*(5), 1–38.

Toomet, O. and A. Henningsen (2008). Sample selection models in R: Package sampleSelection. *Journal of Statistical Software 27*(7).

Toulmin, S. E. (1958). *The Use of Argument*. Cambridge University Press.

VanderWeele, T. (2016). Mediation analysis: A practitioner's guide. *Annual Review of Public Health 37*, 17–32.

Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives 28*(2), 3–28.

Walton, D. (2009). Argumentation theory: A very short introduction. In G. Simari and I. Rahwan (Eds.), *Argumentation in Artificial Intelligence*, pp. 1–22. Boston, Massachusetts: Springer.

Appendix A

Table A1: Age distribution

Category	Age class	Frequency	Percentage
0	Under 18	0	0%
1	18-24	145	9.6%
2	25-34	380	25.2%
3	35-44	374	24.8%
4	45-54	233	15.5%
5	55-64	229	15.2%
6	65-74	121	8.0%
7	75-84	23	1.5%
8	85 or older	1	0.1%

Notes: The table displays frequencies and percentages of responses across eight age categories.

Table A2: Attrition

	W1 only		Both Waves			
Variable	Mean	SD	Mean	SD	Mean-Diff	p-value
female	0.67	0.47	0.63	0.48	0.04	0.036
age	2.90	1.49	3.32	1.51	-0.42	0.000
number of children	1.05	1.21	1.07	1.22	-0.02	0.730
renting	0.52	0.50	0.44	0.50	0.08	0.000
renting out	0.11	0.31	0.11	0.32	0.00	0.640
republican	0.22	0.42	0.22	0.41	0.00	0.754
democrat	0.46	0.50	0.45	0.50	0.01	0.766
independent	0.25	0.43	0.29	0.45	-0.04	0.042
prior in favor	0.55	0.50	0.54	0.50	0.01	0.749
prior against	0.28	0.45	0.33	0.47	-0.05	0.004
prior unsure	0.17	0.38	0.12	0.33	0.05	0.001
lived rent controlled before	0.18	0.38	0.16	0.36	0.02	0.121

Notes: The table compares key statistics for survey respondents that participated in Wave 1 only (n=1374) to those that participated in both waves (n=1506). Age is a categorical variable ranging from 1 to 8, compare Table A1. P-values are from t-tests of the mean-difference.

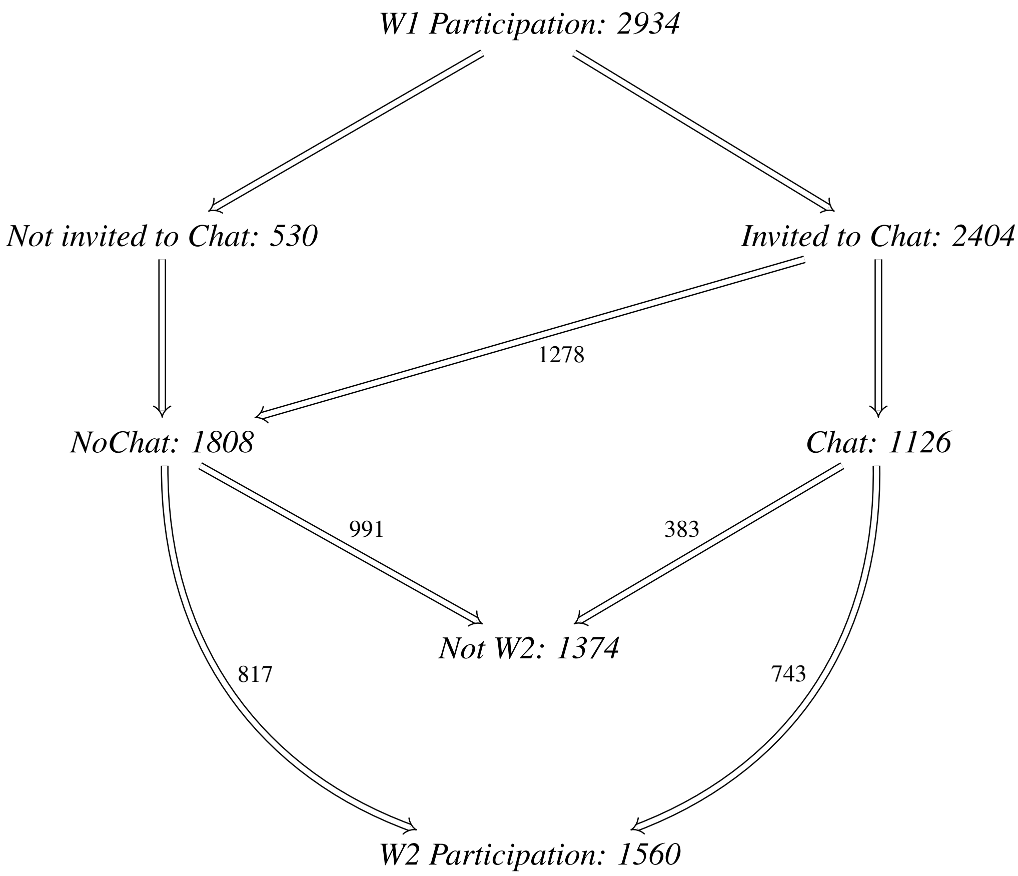


Figure A1: Number of subjects in each stage of the experiment

Table A3: Correlation matrix

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.	21.
1. chat	1																				
2. age	0.04	1																			
3. value_opinionchange	0.03	0.01	1																		
4. renting_out	0.05	-0.03	0	1																	
5. diff_understand_rentcon	0.05	0.12	0	0.08	1																
6. female	-0.01	-0.04	0.01	-0.06	-0.14	1															
7. arg_score_avg		-0.04	-0.02	-0.09	-0.08	0.05	1														
8. diff_arg_score		-0.04	-0.02	-0.08	-0.08	-0.05	-0.19	1													
9. arg_score_op		0.01	-0.01	-0.04	-0.06	-0.02	-0.1	0.66	1												
10. arg_score_al		0.07	0.02	0.09	0.09	0.05	0.06	-0.71	-0.12	1											
11. exante_pos_avg		0.04	-0.02	-0.02	-0.06	-0.04	0.46	-0.1	-0.03	0.08	1										
12. diff_exante_pos		0.04	0.02	0.06	0.05	0.01	-0.22	0.38	0.31	-0.22	-0.34	1									
13. exante_pos_op		0.05	0.06	0.06	0.04	0.02	-0.18	0.25	0.46	0.06	-0.17	0.76	1								
14. exante_pos_al		-0.01	0.04	-0.03	-0.01	-0.01	0.11	-0.33	-0.04	0.45	0.25	-0.75	-0.25	1							
15. arg_coders_avg		-0.06	-0.02	-0.13	-0.09	0.01	0.75	-0.13	-0.04	0.08	0.55	-0.28	-0.2	0.19	1						
16. diff_arg_coders		-0.05	0.01	0	-0.09	-0.02	-0.14	0.74	0.51	-0.54	-0.13	0.51	0.38	-0.41	-0.2	1					
17. arg_coders_op		0	0.02	-0.02	-0.04	-0.01	-0.13	0.45	0.73	-0.02	-0.11	0.4	0.56	-0.1	-0.09	0.61	1				
18. arg_coders_al		0.07	0.03	-0.01	0.1	0.03	0	-0.53	-0.1	0.74	0.05	-0.33	-0.08	0.49	0.11	-0.72	-0.04	1			
19. info_no_camp	0.06	0.17	0.02	0.06	0.13	-0.09	-0.07	-0.02	0.06	0.08	-0.03	0.01	0.07	0.06	-0.06	-0.04	0.04	0.09	1		
20. info_yes_camp	0.07	0.12	0.04	0.04	0.16	-0.08	-0.01	0.05	0.08	-0.03	-0.03	0.05	0.06	0	-0.03	0.02	0.04	0.02	0.57	1	
21. chat_length		0.1	-0.1	0.04	0.04	0.02	-0.08	0	0.16	0.13	-0.01	-0.05	0.01	0.07	-0.04	0	0.15	0.13	0.04	0.03	1

Notes: The table presents Spearman's correlations among all independent variables used in the regressions.

Table A4: Opinion change (binary, controls shown)

	All	Chat	Chat	Chat	Chat	Chat	Chat
value_opinionchange	−0.079	−0.468	−0.448	−0.459	−0.417	−0.450	−0.408
	(0.220)	(0.320)	(0.320)	(0.322)	(0.322)	(0.321)	(0.324)
diff_understand_rentcon	−0.154	−0.248	−0.206	−0.219	−0.252	−0.207	−0.227
	(0.081)	(0.117)	(0.119)	(0.119)	(0.118)	(0.119)	(0.120)
chat	−0.080						
	(0.174)						
diff_exante_pos		0.186		0.124			
		(0.078)		(0.084)			
diff_arg_score			0.157	0.124			
			(0.056)	(0.060)			
exante_pos_op					−0.002		−0.095
					(0.147)		(0.161)
exante_pos_al					−0.380		−0.333
					(0.150)		(0.156)
arg_score_op						0.165	0.159
						(0.080)	(0.088)
arg_score_al						−0.146	−0.081
						(0.107)	(0.108)
female	0.065	0.060	0.144	0.116	0.083	0.141	0.127
	(0.184)	(0.269)	(0.269)	(0.271)	(0.270)	(0.270)	(0.272)
age	−0.027	−0.092	−0.075	−0.085	−0.089	−0.075	−0.085
	(0.059)	(0.091)	(0.091)	(0.091)	(0.091)	(0.091)	(0.091)
renting_out	0.135	0.074	0.242	0.182	0.076	0.242	0.184
	(0.271)	(0.365)	(0.367)	(0.370)	(0.366)	(0.367)	(0.372)
info_no_camp	−0.034	0.162	0.156	0.167	0.186	0.153	0.176
	(0.209)	(0.293)	(0.294)	(0.294)	(0.296)	(0.295)	(0.299)
info_yes_camp	0.070	0.144	0.120	0.118	0.152	0.123	0.136
	(0.205)	(0.295)	(0.297)	(0.298)	(0.297)	(0.298)	(0.301)
chat_length		0.054	0.039	0.045	0.054	0.038	0.040
		(0.039)	(0.038)	(0.039)	(0.039)	(0.040)	(0.041)
Constant	−1.459	−1.621	−1.664	−1.634	−1.268	−1.667	−1.243
	(0.330)	(0.633)	(0.633)	(0.637)	(0.669)	(0.633)	(0.676)
Obs.	1,039	518	518	518	518	518	518
Akaike Inf. Crit.	905.551	440.111	437.816	437.664	439.760	439.800	439.045

Notes: The table reports results of binomial regressions with *opinion_change_bin* as the dependent variable. All independent variables are as described in Table 4. Log odds are reported as coefficients and standard errors are in parentheses.

Table A5: Opinion change (Multinomial, controls shown)

	All	All	Chat	Chat	Chat	Chat
	Ch_to_No	*Ch_to_Yes*	*Ch_to_No*	*Ch_to_Yes*	*Ch_to_No*	*Ch_to_Yes*
value_opinionchange	−0.149	0.066	−0.493	−0.407	−0.465	−0.376
	(0.252)	(0.402)	(0.364)	(0.588)	(0.367)	(0.593)
diff_understand_rentcon	−0.192	−0.041	−0.232	−0.160	−0.237	−0.142
	(0.094)	(0.143)	(0.135)	(0.224)	(0.137)	(0.228)
chat	0.001	−0.288				
	(0.202)	(0.309)				
diff_exante_pos			0.167	−0.001		
			(0.109)	(0.182)		
diff_arg_score			0.170	0.018		
			(0.069)	(0.107)		
exante_pos_op					−0.126	−0.308
					(0.186)	(0.328)
exante_pos_al					−0.411	−0.707
					(0.194)	(0.342)
arg_score_op					0.220	0.069
					(0.100)	(0.181)
arg_score_al					−0.141	0.159
					(0.138)	(0.161)
female	−0.081	0.450	0.072	0.289	0.100	0.394
	(0.211)	(0.342)	(0.304)	(0.525)	(0.307)	(0.538)
age	−0.121	0.194	−0.149	0.093	−0.124	0.135
	(0.070)	(0.101)	(0.106)	(0.168)	(0.105)	(0.168)
renting_out	0.240	−0.229	0.177	0.181	0.204	0.059
	(0.303)	(0.544)	(0.426)	(0.674)	(0.427)	(0.681)
info_no_camp	−0.289	0.608	−0.091	0.985	−0.059	1.255
	(0.242)	(0.367)	(0.335)	(0.562)	(0.341)	(0.598)
info_yes_camp	0.370	−0.728	0.295	−0.503	0.332	−0.606
	(0.236)	(0.373)	(0.339)	(0.550)	(0.344)	(0.572)
chat_length			0.027	0.085	0.021	0.079
			(0.043)	(0.077)	(0.045)	(0.081)
Constant	−1.413	−3.798	−1.434	−4.474	−1.286	−4.365
	(0.378)	(0.622)	(0.715)	(1.269)	(0.727)	(1.287)
Obs.	1039	1039	518	518	518	518
Akaike Inf. Crit.	1,096.652	1,096.652	535.991	535.991	535.674	535.674

Notes: The table reports results of multinomial regressions with *opinion_change_cat* as the dependent variable. All independent variables are as described in Table 4. Log odds are reported as coefficients and standard errors are in parentheses.

Table A6: Opinion change (Distance to prior, controls shown)

	All	Chat	Chat	Chat
value_opinionchange	0.009	0.017	0.017	0.018
	(0.030)	(0.046)	(0.046)	(0.046)
diff_understand_rentcon	−0.005	0.004	0.006	0.006
	(0.010)	(0.016)	(0.016)	(0.016)
chat	−0.010			
	(0.022)			
exante_pos_avg		0.030		0.025
		(0.012)		(0.013)
arg_score_avg			0.013	0.007
			(0.006)	(0.007)
female	0.025	0.026	0.023	0.026
	(0.023)	(0.033)	(0.033)	(0.033)
age	0.002	−0.003	−0.001	−0.002
	(0.007)	(0.011)	(0.011)	(0.011)
renting_out	−0.046	−0.030	−0.029	−0.027
	(0.037)	(0.049)	(0.050)	(0.049)
info_no_camp	0.014	0.020	0.027	0.024
	(0.026)	(0.036)	(0.036)	(0.036)
info_yes_camp	−0.036	−0.043	−0.048	−0.045
	(0.026)	(0.035)	(0.035)	(0.035)
chat_length		0.001	0.002	0.002
		(0.005)	(0.005)	(0.005)
Constant	−0.043	−0.080	−0.072	−0.084
	(0.043)	(0.078)	(0.076)	(0.077)
Obs.	1,170	569	569	569
R^2	0.005	0.018	0.013	0.019
F Statistic	0.789	1.118	0.799	1.098

Notes: The table reports results of OLS regressions with *opinion_change_dist* as the dependent variable. All independent variables are as described in Table 4. Furthermore, the variable *exante_pos_avg* denotes the chat composition of chat-partners only (positions in favor of minus positions against rent control) and *arg_score_avg* denotes the argumentative positions of chat partners only (arguments in favor of minus arguments against rent control). Heteroskedasticity-consistent standard errors are reported in parentheses.

Table A7: Opinion change (Binary, Chat subjects that did not participate in W2 are added)

	All	Chat	Chat	Chat	Chat	Chat	Chat
chat	−0.090						
	(0.175)						
diff_exante_pos		0.192		0.130			
		(0.078)		(0.084)			
diff_arg_score			0.159	0.124			
			(0.056)	(0.060)			
exante_pos_op					0.013		−0.082
					(0.147)		(0.161)
exante_pos_al					−0.377		−0.332
					(0.150)		(0.156)
arg_score_op						0.174	0.165
						(0.080)	(0.089)
arg_score_al						−0.136	−0.072
						(0.108)	(0.108)
Constant	−1.142	−1.380	−1.415	−1.382	−1.024	−1.423	−0.986
	(0.298)	(0.601)	(0.599)	(0.603)	(0.644)	(0.600)	(0.649)
Obs.	1,272	751	751	751	751	751	751
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Akaike Inf. Crit.	915.135	443.517	441.549	441.162	443.389	443.482	442.680

Notes: The table reports results of binomial regressions with *opinion_change_bin* as the dependent variable. All independent variables are as described in Table 4. Log odds are reported as coefficients and standard errors are in parentheses.

Table A8: Opinion change (Multinomial, Chat subjects that did not participate in W2 are added)

	All *Ch_to_No*	All *Ch_to_Yes*	Chat *Ch_to_No*	Chat *Ch_to_Yes*	Chat *Ch_to_No*	Chat *Ch_to_Yes*
chat	−0.005	−0.312				
	(0.203)	(0.311)				
diff_exante_pos			0.171	0.0004		
			(0.109)	(0.182)		
diff_arg_score			0.171	0.017		
			(0.070)	(0.108)		
exante_pos_op					−0.126	−0.309
					(0.187)	(0.328)
exante_pos_al					−0.413	−0.706
					(0.193)	(0.342)
arg_score_op					0.229	0.076
					(0.100)	(0.181)
arg_score_al					−0.131	0.169
					(0.139)	(0.161)
Constant	−1.121	−3.431	−1.178	−4.274	−1.013	−4.139
	(0.343)	(0.561)	(0.674)	(1.239)	(0.686)	(1.254)
Obs.	1272	1272	751	751	751	751
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Akaike Inf. Crit.	1,106.163	1,106.163	539.508	539.508	539.215	539.215

Notes: The table reports results of multinomial regressions with *opinion_change_cat* as the dependent variable. All independent variables are as described in Table 4. Log odds are reported as coefficients and standard errors are in parentheses.

Table A9: Opinion change (Distance to prior, Chat subjects that did not participate in W2 are added)

	All	Chat	Chat	Chat
chat	−0.009			
	(0.022)			
exante_pos_avg		0.020		0.017
		(0.008)		(0.009)
arg_score_avg			0.008	0.004
			(0.004)	(0.004)
Constant	−0.040	−0.062	−0.055	−0.063
	(0.032)	(0.047)	(0.046)	(0.047)
Obs.	1,450	849	849	849
Controls	Yes	Yes	Yes	Yes
R^2	0.006	0.016	0.012	0.017
F Statistic	1.090	1.505	1.156	1.421

Notes: The table reports results of OLS regressions with *opinion_change_dist* as the dependent variable. The variable *exante_pos_avg* denotes the chat composition of chat-partners only (positions in favor of minus positions against rent control) and *arg_coders_avg* denotes the argumentative positions of chat partners only (arguments in favor of minus arguments against rent control). Heteroskedasticity-consistent standard errors are reported in parentheses.

Table A10: Opinion change (Full sample, NoChat subjects that did not participate in W2 are added)

	Binary	Multinomial		OLS
		Ch_to_No	*Ch_to_Yes*	
chat	0.048	0.130	−0.169	−0.026
	(0.173)	(0.201)	(0.309)	(0.017)
Constant	−1.198	−1.183	−3.481	−0.014
	(0.298)	(0.342)	(0.562)	(0.018)
Obs.	1,610	1,610	1,610	2,134
Controls	Yes	Yes	Yes	Yes
Akaike Inf. Crit.	925.430	1,116.440	1,116.440	
R^2				0.006
F Statistic				1.552

Notes: The table reports results of binomial, multinomial and OLS regressions with *opinion_change_bin*, *opinion_change_cat* and *opinion_change_dist* as the dependent variables. All independent variables are as described in Table 4 to Table 8. Log odds are reported as coefficients and standard errors are in parentheses.

Table A11: Opinion change (Binary, Subjects that did not participate in W2 are added with ML predictions)

	All	Chat	Chat	Chat	Chat	Chat	Chat
chat	0.176						
	(0.121)						
diff_exante_pos		0.228		0.144			
		(0.065)		(0.072)			
diff_arg_score			0.167	0.129			
			(0.041)	(0.045)			
exante_pos_op					0.074		−0.027
					(0.108)		(0.118)
exante_pos_al					−0.516		−0.424
					(0.118)		(0.124)
arg_score_op						0.136	0.134
						(0.062)	(0.068)
arg_score_al						−0.212	−0.118
						(0.080)	(0.082)
Constant	−1.028	−1.144	−1.279	−1.179	−1.090	−1.268	−1.096
	(0.221)	(0.471)	(0.472)	(0.475)	(0.475)	(0.472)	(0.480)
Obs.	1,842	751	751	751	751	751	751
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Akaike Inf. Crit.	1,795.198	747.605	743.186	741.235	739.713	744.729	735.902

Notes: The table reports results of binomial regressions with *opinion_change_bin* as the dependent variable. All independent variables are as described in Table 4. Log odds are reported as coefficients and standard errors are in parentheses.

Table A12: Opinion change (Multinomial, Subjects that did not participate in W2 are added with ML predictions)

	All *Ch_to_No*	All *Ch_to_Yes*	Chat *Ch_to_No*	Chat *Ch_to_Yes*	Chat *Ch_to_No*	Chat *Ch_to_Yes*
chat	0.420	−0.311				
	(0.146)	(0.197)				
diff_exante_pos			0.110	0.175		
			(0.085)	(0.121)		
diff_arg_score			0.200	−0.027		
			(0.054)	(0.073)		
exante_pos_op					−0.100	0.029
					(0.139)	(0.202)
exante_pos_al					−0.311	−0.795
					(0.140)	(0.246)
arg_score_op					0.229	−0.174
					(0.075)	(0.144)
arg_score_al					−0.180	0.006
					(0.102)	(0.122)
Constant	−1.078	−2.735	−0.961	−3.480	−0.882	−3.410
	(0.267)	(0.354)	(0.545)	(0.846)	(0.549)	(0.860)
Obs.	1842	1842	751	751	751	751
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Akaike Inf. Crit.	2,234.649	2,234.649	921.186	921.186	913.053	913.053

Notes: The table reports results of multinomial regressions with *opinion_change_cat* as the dependent variable. All independent variables are as described in Table 4. Log odds are reported as coefficients and standard errors are in parentheses.

Table A13: Opinion change (Distance to prior, Subjects that did not participate in W2 are added with ML predictions)

	All	Chat	Chat	Chat
chat	−0.090			
	(0.018)			
exante_pos_avg		0.038		0.032
		(0.010)		(0.011)
arg_score_avg			0.015	0.006
			(0.005)	(0.006)
Constant	−0.022	−0.135	−0.124	−0.140
	(0.034)	(0.068)	(0.067)	(0.067)
Obs.	2,140	849	849	849
Controls	Yes	Yes	Yes	Yes
R^2	0.013	0.021	0.013	0.022
F Statistic	3.535	1.982	1.264	1.892

Notes: The table reports results of OLS regressions with *opinion_change_dist* as the dependent variable. The variable *exante_pos_avg* denotes the chat composition of chat-partners only (positions in favor of minus positions against rent control) and *arg_coders_avg* denotes the argumentative positions of chat partners only (arguments in favor of minus arguments against rent control). Heteroskedasticity-consistent standard errors are reported in parentheses.

Table A14: Opinion change (Binary, manual annotations)

	All	Chat	Chat	Chat	Chat	Chat	Chat
chat	−0.080						
	(0.174)						
diff_exante_pos		0.186		0.105			
		(0.078)		(0.090)			
diff_arg_coders			0.141	0.106			
			(0.051)	(0.059)			
exante_pos_op					−0.002		−0.110
					(0.147)		(0.167)
exante_pos_al					−0.380		−0.307
					(0.150)		(0.158)
arg_coders_op						0.115	0.112
						(0.070)	(0.081)
arg_coders_al						−0.184	−0.121
						(0.094)	(0.097)
Constant	−1.459	−1.621	−1.706	−1.670	−1.268	−1.694	−1.287
	(0.330)	(0.633)	(0.634)	(0.637)	(0.669)	(0.635)	(0.674)
Obs.	1,039	518	518	518	518	518	518
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Akaike Inf. Crit.	905.551	440.111	438.243	438.895	439.760	439.941	440.078

Notes: The table reports results of binomial regressions with *opinion_change_bin* as the dependent variable. All independent variables are as described in Table 4. Log odds are reported as coefficients and standard errors are in parentheses.

Table A15: Opinion change (Multinomial, manual annotations)

	All *Ch_to_No*	All *Ch_to_Yes*	Chat *Ch_to_No*	Chat *Ch_to_Yes*	Chat *Ch_to_No*	Chat *Ch_to_Yes*
chat	0.001	−0.288				
	(0.202)	(0.309)				
diff_exante_pos			0.150	−0.030		
			(0.114)	(0.197)		
diff_arg_coders			0.135	0.043		
			(0.067)	(0.110)		
exante_pos_op					−0.171	−0.084
					(0.191)	(0.349)
exante_pos_al					−0.422	−0.560
					(0.200)	(0.352)
arg_coders_op					0.207	−0.197
					(0.087)	(0.203)
arg_coders_al					−0.060	−0.105
					(0.112)	(0.188)
Constant	−1.413	−3.798	−1.495	−4.497	−1.375	−4.351
	(0.378)	(0.622)	(0.716)	(1.275)	(0.726)	(1.295)
Obs.	1039	1039	518	518	518	518
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Akaike Inf. Crit.	1,096.652	1,096.652	538.069	538.069	536.197	536.197

Notes: The table reports results of multinomial regressions with *opinion_change_cat* as the dependent variable. All independent variables are as described in Table 4. Log odds are reported as coefficients and standard errors are in parentheses.

Table A16: Opinion change (Distance to prior, manual annotations)

	All	Chat	Chat	Chat
chat	−0.010			
	(0.022)			
exante_pos_avg		0.030		0.019
		(0.012)		(0.013)
arg_coders_avg			0.015	0.010
			(0.006)	(0.007)
Constant	−0.043	−0.080	−0.075	−0.084
	(0.043)	(0.078)	(0.077)	(0.078)
Obs.	1,170	569	569	569
Controls	Yes	Yes	Yes	Yes
R^2	0.005	0.018	0.018	0.022
F Statistic	0.789	1.118	1.161	1.240

Notes: The table reports results of OLS regressions with *opinion_change_dist* as the dependent variable. The variable *exante_pos_avg* denotes the chat composition of chat-partners only (positions in favor of minus positions against rent control) and *arg_coders_avg* denotes the argumentative positions of chat partners only (arguments in favor of minus arguments against rent control). Heteroskedasticity-consistent standard errors are reported in parentheses.

Table A17: Opinion change (Binary, only groups of four and five)

	All	Chat	Chat	Chat	Chat	Chat	Chat
chat	−0.080						
	(0.174)						
diff_exante_pos		0.222		0.160			
		(0.083)		(0.090)			
diff_arg_score			0.164	0.122			
			(0.059)	(0.063)			
exante_pos_op					0.094		0.022
					(0.161)		(0.173)
exante_pos_al					−0.353		−0.296
					(0.164)		(0.170)
arg_score_op						0.163	0.132
						(0.085)	(0.093)
arg_score_al						−0.166	−0.111
						(0.117)	(0.118)
Constant	−1.459	−1.549	−1.562	−1.546	−1.269	−1.561	−1.245
	(0.330)	(0.708)	(0.706)	(0.713)	(0.766)	(0.707)	(0.777)
Obs.	1,039	464	464	464	464	464	464
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Akaike Inf. Crit.	905.551	383.115	382.406	381.219	384.245	384.406	384.305

Notes: The table reports results of binomial regressions with *opinion_change_bin* as the dependent variable. All independent variables are as described in Table 4. Log odds are reported as coefficients and standard errors are in parentheses.

Table A18: Opinion change (Multinomial, only groups of four and five)

	All *Ch_to_No*	All *Ch_to_Yes*	Chat *Ch_to_No*	Chat *Ch_to_Yes*	Chat *Ch_to_No*	Chat *Ch_to_Yes*
chat	0.001	−0.288				
	(0.202)	(0.309)				
diff_exante_pos			0.212	−0.020		
			(0.116)	(0.191)		
diff_arg_score			0.166	0.036		
			(0.073)	(0.110)		
exante_pos_op					−0.013	−0.276
					(0.191)	(0.336)
exante_pos_al					−0.393	−0.700
					(0.207)	(0.362)
arg_score_op					0.187	0.093
					(0.107)	(0.186)
arg_score_al					−0.194	0.155
					(0.157)	(0.168)
Constant	−1.413	−3.798	−1.223	−4.663	−1.088	−4.534
	(0.378)	(0.622)	(0.807)	(1.419)	(0.819)	(1.435)
Obs.	1039	1039	464	464	464	464
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Akaike Inf. Crit.	1,096.652	1,096.652	468.079	468.079	470.561	470.561

Notes: The table reports results of multinomial regressions with *opinion_change_cat* as the dependent variable. All independent variables are as described in Table 4. Log odds are reported as coefficients and standard errors are in parentheses.

Table A19: Opinion change (Distance to prior, only groups of four and five)

	All	Chat	Chat	Chat
chat	−0.010			
	(0.022)			
exante_pos_avg		0.035		0.029
		(0.012)		(0.013)
arg_score_avg			0.015	0.008
			(0.006)	(0.007)
Constant	−0.043	−0.107	−0.085	−0.108
	(0.043)	(0.085)	(0.083)	(0.085)
Obs.	1,170	509	509	509
Controls	Yes	Yes	Yes	Yes
R^2	0.005	0.028	0.021	0.031
F Statistic	0.789	1.625	1.187	1.583

Notes: The table reports results of OLS regressions with *opinion_change_dist* as the dependent variable. The variable *exante_pos_avg* denotes the chat composition of chat-partners only (positions in favor of minus positions against rent control) and *arg_score_avg* denotes the argumentative positions of chat partners only (arguments in favor of minus arguments against rent control). Heteroskedasticity-consistent standard errors are reported in parentheses.

Table A20: Opinion change (Binary, without fastest 10%)

	All	Chat	Chat	Chat	Chat	Chat	Chat
chat	−0.008						
	(0.184)						
diff_exante_pos		0.160		0.101			
		(0.081)		(0.088)			
diff_arg_score			0.140	0.112			
			(0.058)	(0.063)			
exante_pos_op					0.010		−0.083
					(0.153)		(0.169)
exante_pos_al					−0.313		−0.269
					(0.156)		(0.162)
arg_score_op						0.162	0.157
						(0.088)	(0.098)
arg_score_al						−0.110	−0.061
						(0.107)	(0.108)
Constant	−1.478	−1.322	−1.339	−1.328	−1.048	−1.343	−1.006
	(0.355)	(0.666)	(0.668)	(0.670)	(0.703)	(0.668)	(0.711)
Obs.	935	466	466	466	466	466	466
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Akaike Inf. Crit.	813.266	404.653	402.669	403.360	405.297	404.560	405.692

Notes: The table reports results of binomial regressions with *opinion_change_bin* as the dependent variable. All independent variables are as described in Table 4. Log odds are reported as coefficients and standard errors are in parentheses.

Table A21: Opinion change (Binary, without slowest 10%)

	All	Chat	Chat	Chat	Chat	Chat	Chat
chat	−0.200						
	(0.184)						
diff_exante_pos		0.151		0.095			
		(0.082)		(0.090)			
diff_arg_score			0.132	0.105			
			(0.059)	(0.064)			
exante_pos_op					−0.097		−0.183
					(0.159)		(0.175)
exante_pos_al					−0.395		−0.351
					(0.157)		(0.164)
arg_score_op						0.138	0.147
						(0.089)	(0.098)
arg_score_al						−0.123	−0.050
						(0.110)	(0.111)
Constant	−1.412	−1.968	−1.981	−1.962	−1.516	−1.981	−1.459
	(0.346)	(0.683)	(0.684)	(0.687)	(0.719)	(0.684)	(0.729)
Obs.	935	466	466	466	466	466	466
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Akaike Inf. Crit.	816.064	392.030	390.397	391.289	390.577	392.388	391.656

Notes: The table reports results of binomial regressions with *opinion_change_bin* as the dependent variable. All independent variables are as described in Table 4. Log odds are reported as coefficients and standard errors are in parentheses.

Table A22: Opinion change (Multinomial, without fastest 10%)

	All	All	Chat	Chat	Chat	Chat
	Ch_to_No	*Ch_to_Yes*	*Ch_to_No*	*Ch_to_Yes*	*Ch_to_No*	*Ch_to_Yes*
chat	0.118	−0.306				
	(0.216)	(0.319)				
diff_exante_pos			0.155	−0.019		
			(0.112)	(0.193)		
diff_arg_score			0.141	0.036		
			(0.072)	(0.114)		
exante_pos_op					−0.094	−0.273
					(0.191)	(0.341)
exante_pos_al					−0.405	−0.582
					(0.204)	(0.355)
arg_score_op					0.190	0.130
					(0.110)	(0.194)
arg_score_al					−0.093	0.156
					(0.136)	(0.168)
Constant	−1.482	−3.673	−1.138	−4.292	−1.023	−4.192
	(0.413)	(0.648)	(0.750)	(1.371)	(0.760)	(1.381)
Obs.	935	935	466	466	466	466
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Akaike Inf. Crit.	986.296	986.296	494.026	494.026	496.124	496.124

Notes: The table reports results of multinomial regressions with *opinion_change_cat* as the dependent variable. All independent variables are as described in Table 4. Log odds are reported as coefficients and standard errors are in parentheses.

Table A23: Opinion change (Multinomial, without slowest 10%)

	All	All	Chat	Chat	Chat	Chat
	Ch_to_No	*Ch_to_Yes*	*Ch_to_No*	*Ch_to_Yes*	*Ch_to_No*	*Ch_to_Yes*
chat	−0.164	−0.302				
	(0.212)	(0.332)				
diff_exante_pos			0.089	0.032		
			(0.118)	(0.198)		
diff_arg_score			0.150	0.026		
			(0.075)	(0.111)		
exante_pos_op					−0.320	−0.290
					(0.218)	(0.346)
exante_pos_al					−0.383	−0.724
					(0.203)	(0.363)
arg_score_op					0.215	0.089
					(0.113)	(0.191)
arg_score_al					−0.094	0.163
					(0.141)	(0.168)
Constant	−1.377	−3.745	−1.744	−5.042	−1.551	−4.964
	(0.395)	(0.655)	(0.770)	(1.385)	(0.787)	(1.407)
Obs.	935	935	466	466	466	466
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Akaike Inf. Crit.	988.939	988.939	481.365	481.365	478.614	478.614

Notes: The table reports results of multinomial regressions with *opinion_change_cat* as the dependent variable. All independent variables are as described in Table 4. Log odds are reported as coefficients and standard errors are in parentheses.

Table A24: Opinion change (Distance to prior, without fastest 10%)

	All	Chat	Chat	Chat
chat	−0.015			
	(0.024)			
exante_pos_avg		0.030		0.025
		(0.012)		(0.013)
arg_score_avg			0.012	0.006
			(0.007)	(0.007)
Constant	−0.040	−0.071	−0.067	−0.075
	(0.046)	(0.085)	(0.083)	(0.084)
Obs.	1,053	512	512	512
Controls	Yes	Yes	Yes	Yes
R^2	0.007	0.019	0.013	0.020
F Statistic	0.882	1.063	0.739	1.016

Notes: The table reports results of OLS regressions with *opinion_change_dist* as the dependent variable. The variable *exante_pos_avg* denotes the chat composition of chat-partners only (positions in favor of minus positions against rent control) and *arg_score_avg* denotes the argumentative positions of chat partners only (arguments in favor of minus arguments against rent control). Heteroskedasticity-consistent standard errors are reported in parentheses.

Table A25: Opinion change (Distance to prior, without slowest 10%)

	All	Chat	Chat	Chat
chat	−0.002			
	(0.024)			
exante_pos_avg		0.019		0.011
		(0.012)		(0.013)
arg_score_avg			0.011	0.008
			(0.007)	(0.008)
Constant	−0.034	−0.056	−0.055	−0.061
	(0.046)	(0.081)	(0.079)	(0.080)
Obs.	1,053	512	512	512
Controls	Yes	Yes	Yes	Yes
R^2	0.005	0.011	0.012	0.013
F Statistic	0.601	0.600	0.671	0.671

Notes: The table reports results of OLS regressions with *opinion_change_dist* as the dependent variable. The variable *exante_pos_avg* denotes the chat composition of chat-partners only (positions in favor of minus positions against rent control) and *arg_score_avg* denotes the argumentative positions of chat partners only (arguments in favor of minus arguments against rent control). Heteroskedasticity-consistent standard errors are reported in parentheses.

Table A26: Opinion change (Distance to prior, Ordered Logit)

	All	Chat	Chat	Chat
chat	−0.034			
	(0.108)			
exante_pos_avg		0.115		0.080
		(0.050)		(0.055)
arg_score_avg			0.072	0.051
			(0.031)	(0.034)
Obs.	1,170	569	569	569
Controls	Yes	Yes	Yes	Yes
Akaike Inf. Crit.	3947.789	1843.038	1842.917	1842.808

Notes: The table reports results of OLS regressions with *opinion_change_dist* as the dependent variable. All independent variables are as described in Table 4. Furthermore, the variable *exante_pos_avg* denotes the chat composition of chat-partners only (positions in favor of minus positions against rent control) and *arg_score_avg* denotes the argumentative positions of chat-partners only (arguments in favor of minus arguments against rent control). Heteroskedasticity-consistent standard errors are reported in parentheses.

Appendix B

In this section, we investigate the potential mediation effect from our exogenous treatment variation within chats, i.e. the chat composition in ex-ante positions on rent control (*diff_exante_pos*), through the potential mediator of argument composition (*diff_arg_score*) on the outcome *opinion_change_bin*. In other words, the chat composition in ex-ante positions on rent control might not only affect opinion change directly but also through the argument composition in the chat. Chats that are homogeneous with regard to ex-ante positions might have a different argument structure than those chats with heterogeneous views which in turn affects opinion change.

Since the chat composition (*diff_exante_pos*) can take a lot of different combinations with regard to positions on rent control, e.g. three versus two or five versus none etc., as a first step, we simplify the chat composition to a dummy variable that is equal to one (treated) for chats that contain more opposing than aligned views for a subject and zero otherwise (control). We denote this dummy as *majority_opp*. Thus, we test if a subject is more likely to change opinion if opposing views are the majority, regardless of the specific composition of the chat. See Figure B1 for an illustration of the *direct* and *indirect* effect of the exogenous treatment variation on the outcome, i.e. opinion change.

$$\begin{tikzcd}
 & \textit{argument composition} \arrow[dr, Rightarrow] & \\
\textit{majority opposed} \arrow[ur, Rightarrow] \arrow[rr, Rightarrow] & & \textit{opinion change}
\end{tikzcd}$$

Figure B1: Potential Mediation

We identify a causal mediation effect with the following four assumptions that are also known as sequential ignorability or sequential independence assumptions (e.g. Huber 2020): There must not be confounders between treatment and outcome relationship (Assumption 1). There must not be confounders between mediator and outcome relationship (Assumption 2). There must not be confounders between treatment and mediator relationship (Assumption 3). There must not be confounders affected by the treatment between mediator and outcome relationship (Assumption 4).

Assumption 1 and 3 are met because our treatment, i.e the chat composition, is randomized. For Assumption 2 we have to carefully think of all post-treatment potential confounders that affect the path from the mediator to the outcome. In our case, where the mediator is the argument compositions of subject i's chat partners, it is hard to imagine a post-chat confounder that affects chat partners argumentation and at the same time subject i's opinion change decision. The same is true for Assumption 4, which states that there should be no treatment-induced confounders between i's chat partners argumentation and i's opinion change. Another argument is put forward by VanderWeele (2016): Assumption 4 is more plausible, the less time is elapsed between the treatment and the mediator. In our case, the argument composition within chats directly follows the exogenous chat composition in positions on rent control, leaving little room for potential confounders. Under these sequential ignorability assumptions, a causal mediation from the treatment to the outcome variable can be established. In the following, we empirically investigate if such a mediation effect exists.

For the empirical analysis of this potential mediation effect, we use the methods proposed in Imai et al. (2010a) and Imai et al. (2010b) that are implemented in the R-package "mediation" ((Tingley et al. 2014). The advantage of these methods are (a) that they allow high flexibility with regard to the type of regression model used for the outcome and mediator model and (b) it implements a sensitivity analysis that allows investigating "how strongly" the assumptions need to be violated in order to reach different conclusions from the mediation analysis. This identification strategy for the causal effect is also called "partial identification based on sensitivity checks" (Huber 2020).

As a first step, we formulate the outcome and mediator model as

$$opinion\_change_i = \alpha + \beta majority\_opp_i + \delta diff\_arg\_score_i + \gamma X_i + \varepsilon_i \quad (1)$$

$$diff\_arg\_score_i = \lambda + \theta majority\_opp_i + \phi X_i + \eta_i, \quad (2)$$

where X is a matrix that contains all covariates that are used as control variables in our opinion change regressions. Regression models (1) and (2) are subsequently used to estimate if there is a indirect causal effect from the chat composition on opinion changes through the argument structure in the chats.

Results of the mediation analysis are provided in Table B1. The ACME (Average causal mediation effect) is significant for those that face more opposing than aligned views and the chat (treated) as well as for those that do not (control). This means, that the chat composition with regard to a majority of opposing views exhibits a significant indirect effect on opinion change via the mediator *diff_arg_score*. The ADE (Average direct effect), however, is not significant, i.e. there is no direct effect of the majority on a subject's opinion change. Thus, the majority does not per se affect changes in opinions among subjects but only via the argument composition of chat partners. This is consistent with our finding in column 4 of Table 4: Adding the chats argument composition makes the direct effect of the chat composition disappear. Finally, we perform a sensitivity analysis, which allows us to assess how robust

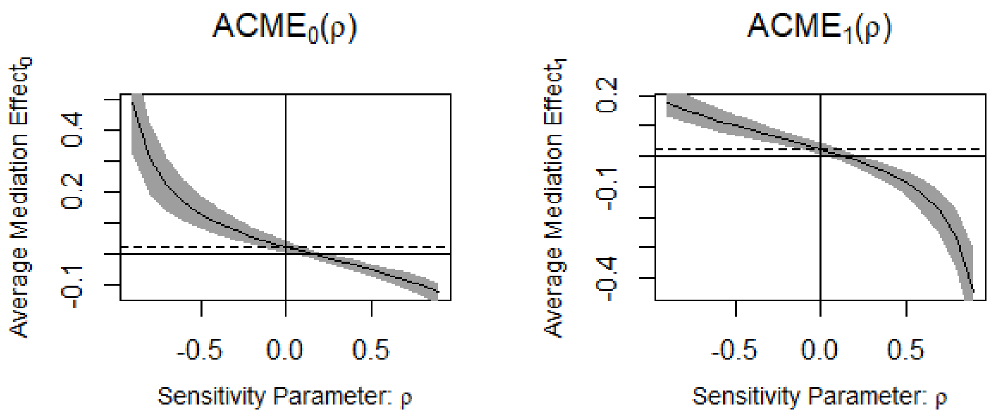
Table B1: Causal Mediation Analysis

Effect	Estimate	CI lower	CI upper	p-value
ACME (control)	0.025	0.003	0.05	0.022
ACME (treated)	0.029	0.004	0.05	0.022
ADE (control)	0.030	-0.031	0.10	0.334
ADE (treated)	0.034	-0.035	0.11	0.334
Total Effect	0.059	-0.005	0.13	0.068
Prop. Mediated (control)	0.423	-1.265	3.73	0.086
Prop. Mediated (treated)	0.485	-1.073	3.50	0.086
ACME (average)	0.027	0.003	0.05	0.022
ADE (average)	0.032	-0.033	0.10	0.334
Prop. Mediated (average)	0.454	-1.177	3.62	0.086

Notes: Confidence intervals are obtained with nonparametric bootstrap using the percentile method. Sample size used 518. Simulations: 1000. * indicates significance at the 10% level, ** at the 5% level and *** at the 1% level.

our direct and indirect effect estimates are to a potential violation of the sequential ignorability assumptions and how substantial a violation in the assumptions would have to be in order to considerably alter our inferences about direct and indirect effects.

The basic idea of the sensitivity analysis is to study the correlation ρ of the errors of the outcome and mediatior models (ε and η). Under sequential ignorability, ρ is equal to zero. If important confounders are omitted that affect both our mediator *diff_arg_score* and the outcome opinion change, ε and η are either positively or negatively correlated. Thus the magnitude of the correlation coefficient ρ represents the departure from the ignorability assumption. Results are summarized in Figure B2. We see that for those being treated with a majority of opposing views ($ACME_1$), only ρ in the higher positive domain would result in a different sign of the estimated mediation effect. Overall, ρ needs to be 0.1 to have a ACME of zero and would need to be larger to draw different conclusions about the mediation effect.



(a) For ACME (control)

(b) For ACME (treated)

Figure B2: Sensitivity Analysis

Appendix C

In this section, we report the results of a Heckman selection model. We model the selection of participants into our sample with regard to our dependent variable, i.e. opinion change. We define the variable *selection* that is equal to zero if a participant did not take part in wave 2. Moreover, *selection* is equal to zero if a participant took part in both waves but a) the voting intention is missing, b) the vote in wave 2 is missing or c) both are missing. The variable is equal to one if both are available and therefore the participants selected into our sample. Table C1 reports a summary of these selection effects.

Table C1: Construction of selection variable

Type	Frequency	*selection*
No W2 participation	1374	0
W1&W2 and no voting intention	89	0
W1&W2 and no voting info	54	0
W1&W2 and no info at all	162	0
W1&W2 and info available	1201	1

Notes: Note that those participants that participated in both waves but where information is missing sum up to 89+54+162=305, as we reported in Table 3.

We correct for these selection effects with the two-step procedure proposed by Heckman (1979). We summarize the first stage probit estimation modeling the selection effect as follows (we drop the subscript for participant i for convenience, standard errors are reported in parentheses):

$$\begin{aligned} selection = &\underset{(0.097)}{-0.247} - \underset{(0.003)}{0.007} predict\_ballot\_diff + \underset{(0.060)}{0.525} chat \\ &+ \underset{(0.020)}{0.076} age + \underset{(0.061)}{0.102} info\_no\_camp \end{aligned} \tag{3}$$

We choose this model from a series of models according to the Akaike information criterion (AIC). The variable *predict_ballot_diff* is defined as $abs(predict\_ballot - 50)$, where *predict_ballot* is a participant's belief about the share of Yes-votes on the ballot (in percent). It serves as our exclusion restriction and does not appear in the second stage estimation. It models a participant's expectation in wave 1 of the election outcome. The higher its value, the more a participant expects the ballot to be already decided. As we can see from Equation 3, the more a participant expects the ballot to be already decided, the less likely she will select into our sample. Results of the second stage estimations reported in Table C2 to Table C4 show that our results are robust to the correction of these selection effects.

Table C2: Opinion change (Binary, Heckman selection)

	All	Chat	Chat	Chat	Chat	Chat	Chat
chat	0.141						
	(0.090)						
diff_exante_pos		0.024		0.015			
		(0.011)		(0.012)			
diff_arg_score			0.020	0.017			
			(0.007)	(0.007)			
exante_pos_op					−0.005		−0.022
					(0.019)		(0.020)
exante_pos_al					−0.059		−0.050
					(0.017)		(0.018)
arg_score_op						0.025	0.027
						(0.011)	(0.012)
arg_score_al						−0.015	−0.004
						(0.012)	(0.012)
Constant	−0.287	−0.127	−0.191	−0.157	−0.127	−0.190	−0.148
	(0.279)	(0.247)	(0.258)	(0.253)	(0.250)	(0.258)	(0.253)
Obs. (both stages)	1,874	1,354	1,354	1,354	1,354	1,354	1,354
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
ρ	0.936	1.037	1.093	1.066	1.062	1.092	1.081
Inverse Mills Ratio	0.480 (0.274)	0.543 (0.386)	0.614 (0.403)	0.576 (0.394)	0.569 (0.391)	0.612 (0.403)	0.591 (0.396)

Notes: The table reports results of the second stage of a Heckman selection model with *opinion_change_bin* as the dependent variable of a linear model. All independent variables are as described in Table 4. Standard errors in parentheses are corrected for the two-step procedure using the *sampleSelection* package in R (Toomet and Henningsen 2008).

Table C3: Opinion change (Directional, Heckman selection)

	All	All	Chat	Chat	Chat	Chat
	Ch_to_No	*Ch_to_Yes*	*Ch_to_No*	*Ch_to_Yes*	*Ch_to_No*	*Ch_to_Yes*
chat	0.087	0.054				
	(0.070)	(0.047)				
diff_exante_pos			0.015	−0.0002		
			(0.010)	(0.006)		
diff_arg_score			0.017	−0.0002		
			(0.006)	(0.004)		
exante_pos_op					−0.014	−0.008
					(0.018)	(0.011)
exante_pos_al					−0.031	−0.020
					(0.016)	(0.010)
arg_score_op					0.026	0.002
					(0.011)	(0.006)
arg_score_al					−0.010	0.006
					(0.011)	(0.006)
Constant	−0.095	−0.192	−0.162	0.004	−0.170	0.022
	(0.218)	(0.146)	(0.237)	(0.105)	(0.241)	(0.104)
Obs. (both stages)	1,874	1,874	1,354	1,354	1,354	1,354
Controls	Yes	Yes	Yes	Yes	Yes	Yes
ρ	0.732	0.803	1.125	−0.048	1.142	−0.101
IMR	0.274 (0.215)	0.205 (0.144)	0.585 (0.369)	−0.009 (0.166)	0.610 (0.375)	−0.019 (0.165)

Notes: The table reports results of the second stage of a Heckman selection model with *Chat_to_No* and *Change_to_Yes* as the dependent variables of a linear model. The variable *Chat_to_No* (*Change_to_Yes*) is equal to one if a participant changed to a No-vote (Yes-vote) and zero otherwise. All independent variables are as described in Table 4. Standard errors in parentheses are corrected for the two-step procedure using the *sampleSelection* package in R (Toomet and Henningsen 2008).

Table C4: Opinion change (Distance to prior, Heckman selection)

	All	Chat	Chat	Chat
chat	0.097			
	(0.081)			
exante_pos_avg		0.030		0.024
		(0.011)		(0.013)
arg_score_avg			0.013	0.007
			(0.006)	(0.007)
Constant	−0.574	−0.171	−0.186	−0.193
	(0.389)	(0.579)	(0.582)	(0.579)
Obs. (both stages)	2,746	2,146	2,146	2,146
Controls	Yes	Yes	Yes	Yes
ρ	0.751	0.207	0.256	0.247
IMR	0.356 (0.258)	0.078 (0.491)	0.097 (0.493)	0.093 (0.491)

Notes: The table reports results of the second stage of a Heckman selection model with *opinion_change_dist* as the dependent variable of a linear model. All independent variables are as described in Table 4. Furthermore, the variable *exante_pos_avg* denotes the chat composition of chat-partners only (positions in favor of minus positions against rent control) and *arg_score_avg* denotes the argumentative positions of chat-partners only (arguments in favor of minus arguments against rent control). Standard errors in parentheses are corrected for the two-step procedure using the *sampleSelection* package in R (Toomet and Henningsen 2008).

Online Appendix

In this section, we detail the argument mining procedures that result in the explanatory variable *diff_arg_score* we employ to account for the heterogeneity of argumentative positions of an individual's chat partners. The goal is to arrive at an average argumentative position for each subject that is used in the construction of *diff_arg_score*.

In a first step, a random forest classification model with features extracted from the language model BERT (Devlin et al. 2018) is trained to distinguish argumentative messages from those that are not. We use the claim-premise model as the underlying theory of argumentation (Toulmin 1958, Walton 2009), where an argument consists of a claim and a premise. A premise, which is also referred to as evidence, gives justification for the claim. As Rinott et al. (2015) point out: "Needless to say, evidence plays a critical role in a persuasive argument".

For the training and testing phase of the classification exercise, 3933 textbox- and chat messages are manually labelled as either containing such a justification, i.e. a premise, for an underlying claim or not. The labeling scheme is outlined in Table D1. A claim like *Rent control is not a good idea* does not contain any justification and is therefore labeled as *NoPremise*. The same is true for introductory messages such as *Hi there, how are you?*. Sometimes, justifications are provided without the claim being explicitly stated (premise plus implicit claim). In fact, this frequently occurs in our chat data, where a claim might be stated at the beginning of the discussion and justifications are given later on without referring to the underlying claim again. As we label our data for premises on rent control, we perform the argument mining task of context-specific premise detection. Overall, 1415 (44%) of messages were labeled as containing a premise and 1778 (56%) as not containing a premise. Three trained coders annotated the data set independently. Unweighted Cohen's kappa and Krippendorff's alpha for the labeling procedure are 0.75 and 0.75 respectively, indicating substantial agreement among coders. We discarded the 740 messages where coders disagreed.

In addition to our manual labels, each message is represented by a numerical vector that represents its semantic meaning using the language model BERT (Devlin et al. 2018). These vectors contain nondimensional numbers that numerically represent the meaning of the message. Similarities and differences in meaning across messages can be analyzed by the distance of the vectors in the vector space. These vectors representations are fed into a random forest classification model (Breiman 2001) together with the manually obtained labels, i.e. *NoPremise* and *Premise*, to train the algorithm to automatically detect messages with and without argumentative reasoning. In other words, the labels help the algorithm to concentrate on those dimensions of the vectors that clearly distinguish argumentative from non-argumentative messages. Vector representations from BERT together with the random forest classification model are used because this results in one of the best algorithms with regard to overall performance (Hüning et al. (2021) perform a horse-race of different classification models and NLP techniques to detect arguments in the very same chat data). Moreover, the random forest classifier stands out compared to more sophisticated classification models such as multi-layer neural networks because it does not need extensive prior calibration, is easy to implement and shows little overfitting in applications (Varian 2014, Penczynski 2019).

Results of the Machine Learning exercise are obtained by performing stratified 10-fold cross validation. In each fold, 20 randomly drawn hyper-parameter were tested with regard to the number of variables randomly sampled as candidates at each split of a decision tree. The hyper-parameter value that resulted in the best overall accuracy was 85. The random forest estimated 504 decision trees. Overall the system can distinguish non-argumentative messages from argumentative messages with an accuracy of 91%. Precision, and recall of detecting premises are 89% and 90%, respectively. The F1-value, the harmonic mean of precision and recall, is 89%. As a result, the trained classification model predicts (out-of-sample) for each message the probability of an argument being present or not.

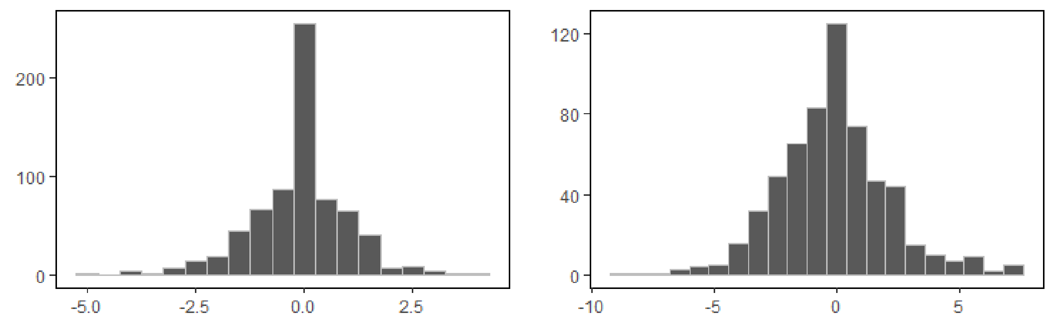
Table D1: Labeling Scheme

Example	Type	Label
"Hi there, how are you?"	None	NoPremise
"Rent control is not a good idea"	Claim only	NoPremise
"Rent control is good because it will lead to affordable housing."	Claim plus premise	Premise
"It would lead to higher rental prices in the long run."	Premise with implicit claim	Premise

Notes: Three trained coders annotated the data set independently. Unweighted Cohen's kappa and Krippendorff's alpha for the labeling procedure are 0.75 and 0.75 respectively, indicating substantial agreement among coders. We discarded the 740 messages where coders disagreed.

In a second step, a second random forest is trained to predict the position for each argumentative message from the first step i.e. if it is in favor of or against rent control. For this, all arguments from the 3933 manually labelled messages are labelled as in favor of or against rent control. Again, vector representations from BERT and these labels are fed into a random forest to predict the position of the argument. The random forest estimated 504 decision trees. In each fold, 20 randomly drawn hyper-parameter were tested with regard to the number of variables randomly sampled as candidates at each split of a decision tree. The value that resulted in the best overall accuracy was 160. The accuracy of this second algorithm is 78%. Precision, recall and F1-value are 80%, 76% and 78%. The raw probabilities of this prediction are used as a proxy for the argumentative persuasiveness of each message in the domains of being in favor of and being against rent control, respectively. Probabilities of arguments against rent control are multiplied by -1. Following this first two steps allows the first algorithm to concentrate on argumentative structure regardless of the position of the argument and the second algorithm to concentrate on what distinguishes pro versus con arguments of rent control.

In a third step, the average argumentative position of each subject is calculated as the sum of the message-level argument scores. For instance, an individual expressed three arguments during the chat discussion that were detected by the algorithm. One in favour of rent control assigned with probability 0.8 and two against assigned with probabilities 0.7 and 0.6. The average argumentative position of this individual is $0.8-0.7-0.6=-0.5$ (Remember that argument probabilities against rent control are multiplied by -1). Measuring the position of each argument, i.e. in favor of and against, on the message level has the following advantage: An individual might be engaged both in argumentation in favor of and against rent control. Only the sum of all arguments measures an individual's overall argumentative position on rent control. The left panel of Figure D1 illustrates the distribution of argument scores across all chat participants.



(a) Individual argument scores

(b) chat-partners' heterogeneity in scores

Figure D1: Distributions of individual argument scores and chat-partners' heterogeneity in scores

Finally, the heterogeneity of argumentative positions of an individual's chat partners is summarized as follows. Formally, for an individual i that is a priori against rent control, we calculate $\sum_{j=1}^{n} ArgPosition_j$, with $j \neq i$, while n is the number of subjects in $i's$ chat group without i. Contrastingly, for an individual that is a priori in favor of rent control, we calculate $(-1) * \sum_{j=1}^{n} ArgPosition_j$. We refer to this variable as *diff_arg_scores*. We thus calculate for each individual the strength of arguments that are opposing her minus the strength of arguments that align with her prior voting intention. The right panel of Figure D1 illustrates the distribution of *diff_arg_scores* across all chat participants.