

RESEARCH ARTICLE

# Introducing ReChat: A Lab-in-the-Cloud for Text Discussions

Xiaoxiao Shen<sup>\*,1</sup> and William Small Schulz<sup>2</sup>

<sup>1</sup>Postdoctoral Associate at Macmillan Center and Lecturer in Political Science, Yale University, New Haven, CT 06511, USA

<sup>2</sup>Postdoctoral Scholar, Cyber Policy Center, Stanford University, Stanford, CA 94305, USA

\*Corresponding author. Email: xiaoxiao.shen@yale.edu

## Abstract

Text is a major medium of contemporary interpersonal communication, but is difficult for social scientists to study unless they have significant resources, or the skills to build their own research platform. In this paper, we introduce a cloud-based software solution to this problem: ReChat, an online research platform for conducting experimental and observational studies of live text conversations. We demonstrate ReChat by applying it to a specific phenomenon of interest to political scientists: conversations among co-partisans. We present results from two studies, focusing on (1) self-selection factors that make chat participants systematically unrepresentative, and (2) a pre-registered analysis of loquaciousness that finds a significant association between speakers' ideological extremity and the amount they write in the chat. We conclude by discussing practical implications and advice for future practitioners of chat studies.

**Keywords:** software, communication, discussion, chat, experiment

Word Count: 4994

## 1. Studying Political Conversations: The State of the Field

Interpersonal conversation is widely considered an important venue for attitude expression and other key political behaviors, but its study is often impeded by logistical and methodological challenges. Recently, scholars have started building their own bespoke data-collection tools for studying textual conversations in controlled experiments conducted over the internet, but these tools are not available to most researchers. This paper therefore introduces a general-purpose tool for studying online text-based chats, in order to fill this gap. Below, we review how text-based chats have arisen as a prominent modality of contemporary communication research, and how we hope to advance and expand the availability of this approach to a larger community of scholars.

Interpersonal conversation has long been an elusive fascination for political scientists. This has been true since the earliest studies of political communication by the Columbia scholars, who argued that interpersonal discussions influenced political outcomes, at the same time that they lamented the difficulty of collecting the “systematic inventory” of conversations necessary to estimate this influence (Lazarsfeld, Berelson, and Gaudet 1948, p. 150). Over the years, scholars have developed creative ways to characterize political discussions observationally, using methods like cross-sectional surveys (e.g. Mutz 2002) and participant observation (e.g. Cramer Walsh 2004). But, though these methods can describe conversations, they cannot directly test claims about their causes and effects.

To this end, lab and field experiments have proven invaluable, if costly. For example, Klar (2014) and Druckman, Levendusky, and McLain (2018) used lab experiments to identify the effects of group homogeneity and the 2-step flow of partisan news exposure in group discussions, respectively. Lab discussions have also been used to study conformity (Carlson and Settle 2016; Levitan and Verhulst 2016) and polarization (Strandberg, Himmelroos, and Grönlund 2019; Grönlund, Herne, and Setälä 2015), as well as the effects of discussion itself (e.g. Simon and Sulkin 2002). Field experiments, meanwhile, have been used to great effect in testing perspective-taking and persuasive interventions (e.g. Broockman and Kalla 2016; Kalla, Levine, and Broockman 2022). The number of topics and outcomes that can be studied with such experiments is limitless, but in practice their application is limited by cost and logistical considerations: field experiments are very expensive, and lab experiments pose challenges in regard to participant scheduling, and the physical capacity of the laboratory space. Yet, interest in political conversation has only continued to grow in response to technological developments over the past decade.

The rise of social media, in particular, has increased interest in text-based conversations, and the availability of free APIs for data collection has facilitated many observational studies of public speech on social platforms (e.g. Brady et al. 2017) – though platforms have increasingly restricted this access, making data collection opportunities more scarce (Freelon 2018). Likewise, although scholars have conducted experiments on major platforms (e.g. Munger 2017; Bail et al. 2018), they struggle to account for platforms’ “black box” features, and many are not replicable (Munger 2019). Also, much contemporary communication occurs in private messaging apps, and although it is possible to run experiments in these apps (e.g. Vermeer et al. 2021), they are designed for privacy rather than data collection, which limits their usefulness for research purposes.

So, researchers have increasingly turned to developing their own data collection tools, in the form of researcher-constructed text communication platforms. These platforms permit lab-like research designs that can probe the effects of political conversation in a digital context. For example, Combs et al. (2023) developed “DicussIt,” an asynchronous chat platform that they paid participants to use (telling them they were testing “a new social media platform,” p. 1455), in an experiment to measure effects of cross-partisan conversations. Rossiter (2023) developed “Chatter,” a real-time chat platform, to compare the effects of political and non-political conversations on affective polarization. Jaidka et al. (2021) created a “custom social media platform” (p. 7) composed of chatrooms, which the authors compared to Discord. Rather than developing their own platform, Santoro and Broockman (2022) were granted special access to the “AllSides” video call platform to conduct a study on conversation effects’ durability. Argyle et al. (2023), meanwhile, built a chatroom with an AI tool that rephrased participants’ messages to improve conversations. Each of these studies makes important scientific contributions, but to our knowledge none of the chat platforms developed by these scholars is currently available for other researchers to use. Although this is an understandable consequence of the expense of software development, these researchers are needlessly reinventing a fundamentally similar technology in parallel, while others are precluded from conducting such studies entirely. This is regrettable, since textual communication is a major feature of modern social life, generating a wide variety of research questions that merit study from a diversity of scholarly perspectives.

Most academics cannot afford to build a bespoke platform. While general-purpose tools like oTree (Chen, Schonger, and Wickens 2016) and Empirica (Almaatouq et al. 2021) can be used to create interactive chatrooms, these still require substantial coding. This makes chat studies inaccessible to

scholars who lack large budgets or extensive technical skills. As a result, many potentially-informative chat studies are never conducted. We consider the recent glut of chat-based research (exemplified by the papers cited in the preceding paragraph) as evidence of *de facto* demand for such designs among political scientists, and so we assert that unnecessary technological and financial barriers are likely hindering valuable social science.

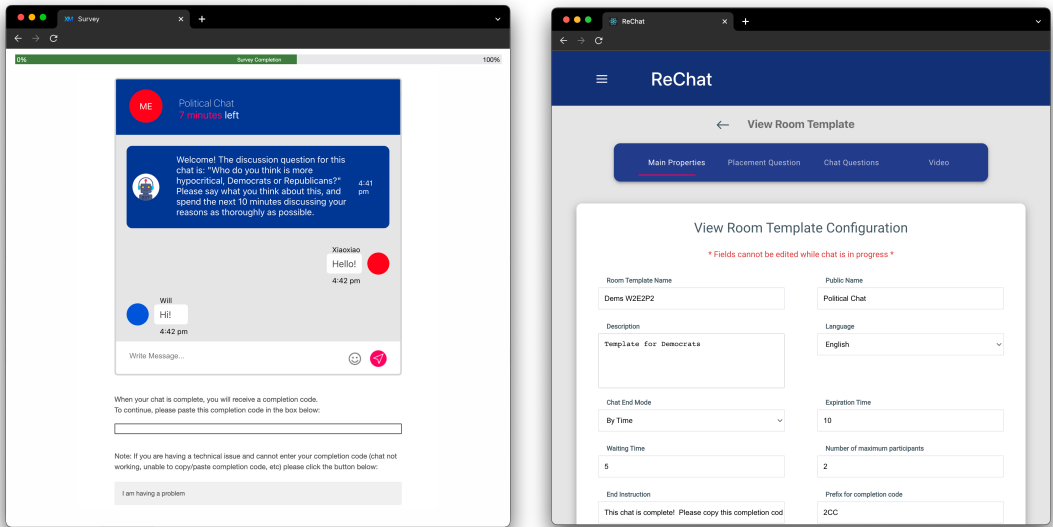
## 2. Introducing ReChat

This paper introduces a solution, developed by one of the authors (Shen), named ReChat. ReChat (<https://reso.chat/>) is a general-purpose chat platform for researchers to observe live online interpersonal communications at a large scale. In contrast to the bespoke platforms described above, it is designed to provide a core toolkit to facilitate most chat studies, and to be easily integrated with Qualtrics surveys and R data analysis pipelines. In order to make ReChat affordable for researchers, while ensuring financial viability to maintain the tool in the long run, ReChat is available to academic users on a pay-what-you-can donation basis. Prospective users can sign up online (<https://reso.chat/signup>), or contact the corresponding author.

We designed ReChat to be easily integrated into an online survey (although it can also be used as a stand-alone interface), and we provide a Qualtrics Survey File (QSF) to help new users embed a ReChat interface into a Qualtrics survey. This allows scholars to conduct chat experiments using their existing knowledge of survey experiment design and cloudworker recruitment (from online panels such as Amazon Mechanical Turk).

Researchers own all data they generate using ReChat, and can download chat transcripts to their personal computer for quality checks and content analyses (meanwhile, survey responses can be downloaded from Qualtrics as in a typical survey). We also provide an R package, `rechat` (<https://github.com/willschulz/rechat>), that eases the initial steps of parsing chat transcripts and linking them to participants' survey responses.

ReChat allows researchers to recruit participants into chatrooms (see Figure 1a for participant view) that can be manipulated by the researcher (see Figure 1b for back-end view), allowing efficient management of large-scale chat studies, without needing to write any code. Furthermore, ReChat lets researchers define instructions and questions to guide the chat, and configure additional options for the conversation. Researchers can create multiple chatroom templates to conduct multiple studies



(a) Chat window embedded in Qualtrics survey.

(b) Chatroom template settings panel.

**Figure 1.** ReChat back-end as seen by researcher (b) and front end as seen by participant (a).

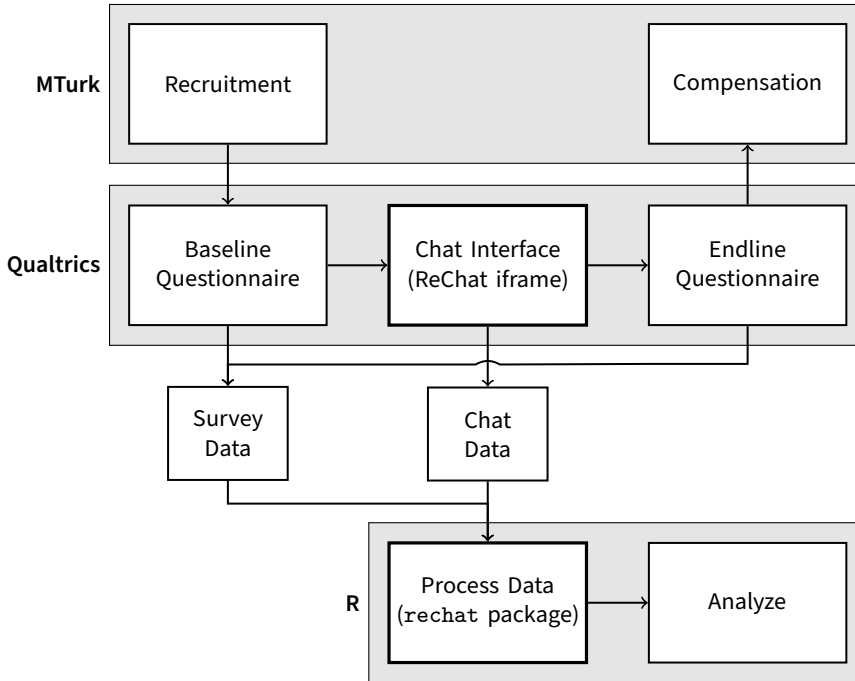
simultaneously. Moreover, ReChat allows significant aspects of the research process to be fully automated, requiring no research assistants to moderate discussions, and employs a waiting room system to place participants into chatrooms as they are recruited, obviating the need to schedule chat sessions in advance (reducing costly attrition).

ReChat is ideal for scholars who wish to conduct chat studies without the significant up-front costs of developing original software. As demonstrated in the following section, it can facilitate a variety of chat research designs, and with its integration into Qualtrics, behaviorists can easily enhance a survey experiment by incorporating a live social interaction to measure their dependent variable. We also hope that ReChat will be useful to early-career researchers who wish to gain experience with conversational data. Ultimately, we hope that ReChat will open up chat studies to a more diverse range of scholars, and unlock their potential to contribute to this research area, while also advancing their own agendas.

### 3. How to Use ReChat: Demonstration Studies

To run a study with ReChat, the researcher creates a ReChat chat template, embeds a ReChat chat window in a Qualtrics survey, rapidly recruits participants from Amazon Mechanical Turk (or a similar crowdwork platform), downloads both the chat and survey data (as CSV files), links these

data (we provide an R package, `rechat`, that simplifies this step), and conducts their analyses. Figure 2 illustrates the overall workflow of a ReChat study, wherein the ReChat chat interface and `rechat` R package are integrated into existing toolkits for data collection and analysis.



**Figure 2.** Summary of technical components, flow of participants through a typical ReChat study, and flow of data into analyses. Bolded boxes highlight our contributions: the ReChat platform and `rechat` R package.

To demonstrate, we present a pair of studies<sup>1</sup> that we fielded on Amazon Mechanical Turk (MTurk) in Spring 2022. Study 1 (N = 483) was fielded in March 2022, and withheld details about the chat activity from initial recruitment, in order to observe selection effects (see Section 3.3). Study 2 (N=629) was fielded in May 2022, with two main differences from Study 1: chat duration was doubled to 10 minutes in order to study loquaciousness, and was explicitly advertised as a chat study. Both studies convened Democratic-identifying US adults to participate in dyadic (that is, two-person) chats about the relative hypocrisy of Democrats and Republicans, and aimed, respectively, to characterize predictors of self-selection into participation in a chat study, and to test predictions about differences in loquaciousness amongst chat participants. Our main analyses are intentionally simple, and we relegate certain details to the appendices (for example, Appendix 9 conceptualizes loquaciousness in greater detail), in order to focus our main text on practical advice for scholars seeking to implement

1. Replication data and code available at: <https://doi.org/10.7910/DVN/PWHPTG> (Schulz and Shen 2024)

their own chat studies. In the sections that follow, we present our procedure, and the insights gleaned at each stage, as an instructional example for future scholars to use and build upon.

### **3.1 *Designing Chatroom Templates in ReChat***

The first step in implementing a study with ReChat is to create a chatroom template (see Figure 1b), which specifies the conditions of the chat that are repeated each time a new chatroom is convened: size (the number of people per chatroom), duration (participants see a countdown timer showing the time remaining in the chat), and a schedule of messages to be posted by the “moderator bot.” This is a simple bot that delivers messages – instructions, discussion prompts, and multiple-choice questions – at set time intervals during the chat. The moderator bot does not respond to the participants, but instead guides the discussion according to a fixed script, providing a consistent discussion structure that is held constant across all chats convened under a given template.

For our demonstration studies, we simply used the moderator bot to deliver the following instruction (see also Figure 1a) at the very start of the chat:

Welcome! The discussion question for this chat is: “Who do you think is more hypocritical, Democrats or Republicans?” Please say what you think about this, and spend the next 10 minutes discussing your reasons as thoroughly as possible.

We designed this prompt to provoke conversations about partisanship, without demanding extensive political knowledge (see Appendix 2). A researcher can specify more elaborate sequences of instructions and multiple-choice questions in their template, according to their research design.

Templates specify several other chat parameters (see Appendix 1 for a full list), including the participant-facing title of the chat (we simply used “Political Chat”), the language of the chat (currently English and Chinese are supported), and the maximum number of minutes that incoming participants can be kept waiting in ReChat’s waiting-room system. Each template is associated with an entry URL that takes participants into a waiting room, where they remain until either the set  $N$  of participants arrives (and they are placed into a chat together), or until the waiting time expires (and they are debriefed).

Every participant who enters a chat via a given URL participates under the parameters of the associated chat template. This allows researchers to place all participants into chats under identical

conditions, or alternatively, to randomize participants into different templates, to observe the effects of alternative chats. The entire process of running a chat study can be fully automated, so that a single researcher can run dozens of chats in tandem with no manual intervention (although a researcher may optionally join any live chat as a human moderator).

### 3.2 *Embedding ReChat in a Survey*

ReChat can be used as a stand-alone chat interface, or it can be integrated into a standard online survey platform, such as Qualtrics. To facilitate the latter use-case, we provide a Qualtrics Survey File (QSF) where a ReChat window is embedded as an `iframe`<sup>2</sup> on one of the pages of the survey. Researchers can edit this QSF in the standard online survey-building interface, to add pre- and post-chat survey questions, implement certain kinds of experimental treatments, and even assign participants to chatrooms conditional on their survey responses. This means that researchers familiar with online questionnaire design can implement a sophisticated chat study by applying their pre-existing expertise in survey design and participant recruitment (see Section 3.3).

To integrate a chat into a survey, the researcher creates a new survey (using the provided QSF file), adds the entry URL for their ReChat template as an embedded data variable (to point the `iFrame` to their template), and creates additional survey pages before and after the chat page to obtain consent and pose questions that can be used as independent or dependent (if placed after the chat) variables.

In our demonstration studies, we measured key demographics and other variables such as political interest, news and social media use, partisanship strength (see Appendix 3), ideology, and strength of social identification as a liberal or conservative (using the measure developed by Huddy, Mason, and Aarøe 2015). We also measured extroversion (adapted from Ashton and Lee 2009) and self-monitoring<sup>3</sup> (Berinsky and Lavine 2011). We used two “feeling thermometers” to collect participants’ subjective feelings of warmth towards Democrats and Republicans (we subtracted warmth-towards-Republicans from warmth-towards-Democrats to derive a measure of participants’ affective favoritism of Democrats over Republicans) both before and after the chat, to observe changes. We also measured key demographic variables after the chat, including age, gender, and education. Participants then exited the survey and received compensation (see Figure 2).

2. [https://www.w3schools.com/tags/tag\\_iframe.ASP](https://www.w3schools.com/tags/tag_iframe.ASP)

3. Self-monitoring is a psychological trait that reflects an individual’s sensitivity to social cues, their desire for social acceptability, and their consequent inclination to adjust their behavior to fit in with their social environment.



Finally, the QSF file also includes a “completion code” field that we use to ascertain chat completion: when a chat reaches its set duration, ReChat provides each participant with a unique completion code to paste into a text box in the Qualtrics survey in which ReChat is embedded, allowing the participant to proceed to the end of the survey, and allowing the researcher to link their survey responses to their chat data (see Section 3.5) and disburse the appropriate payment. The code also records whether a participant timed out in the waiting room (we still paid these participants \$1 for their time). Thus we link chat and survey data collection, facilitating data quality audits, technical error handling, and downstream analyses, using a practice very familiar<sup>4</sup> to crowd-workers. We now turn to discuss recruitment procedures, and patterns of self-selection that researchers should anticipate in chat studies.

### 3.3 *Crowdworker Recruitment, Self-Selection, and Failure Rate*

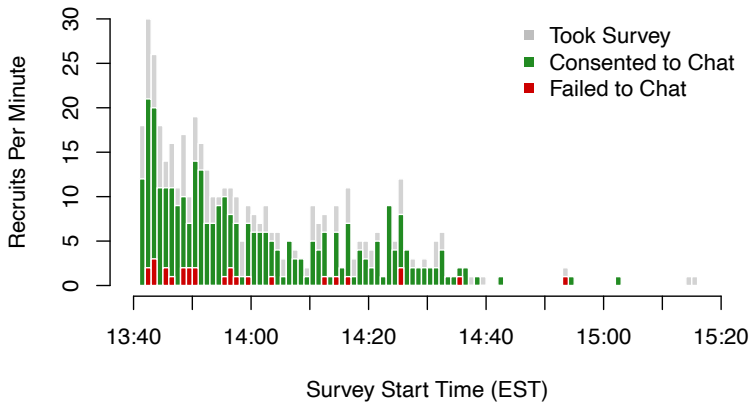
MTurk is an ideal recruitment source for chat studies, because it facilitates rapid recruitment and it allows follow-up “bonus” payments to participants. Rapid recruitment is necessary in order to bring participants into the chat waiting rooms simultaneously and thus place people into live chats together; we therefore launched our studies in early afternoon, when many workers are active, and used tools provided by CloudResearch<sup>5</sup> to accelerate recruitment. Bonus payments can be a useful tool for incentivizing chat participation: in Study 1, for example, participants earned an automatic \$0.75 for completing our survey, plus an additional \$1.00 bonus if they completed a (5-minute) chat, plus a *further* \$1.00 bonus if the “thoroughness” of their chat was above average (this last incentive was designed as an experimental treatment, which we discuss further in Section 3.4). MTurk thus permitted rapid recruitment into our survey, as well as non-coercive incentives to both enter and engage in a chat within the survey.

We advertised Study 1 as a generic survey with an unspecified optional “5-minute activity” (see Appendix 4). This allowed us to start with a typical MTurk survey sample, and observe self-selection into chat participation (which makes chat samples distinctive), as well as establish a failure rate amongst individuals who agreed to chat (which is an important logistical consideration).

Figure 3 shows the number of participants recruited per minute in Study 1, including total

4. Indeed, the main problem we encountered was that respondents were so accustomed to pasting their MTurk Worker IDs into such text boxes for other MTurk studies that they erroneously did so in our study – so we clarified the instructions.

5. In particular, we used the “HyperBatch” feature. See <https://www.cloudfiresearch.com/> for details. We thank Kokil Jaidka, Alvin Zhou, and Yphrach Lelkes for recommending this tool.



**Figure 3.** Study 1 Recruitment.  
See Appendix 7 for corresponding plots for Study 2.

**Table 1.** Self-Selection into Chat Participation (Study 1)

	<i>Dependent variable:</i>				
	Chat Consent				
	(1)	(2)	(3)	(4)	(5)
Extroversion	0.198* (0.102)		0.088 (0.115)	0.139 (0.122)	0.136 (0.123)
Self-Monitoring		0.222** (0.107)	0.143 (0.120)	0.123 (0.128)	0.082 (0.130)
Political Interest			0.246** (0.112)	0.289** (0.121)	0.267** (0.121)
Partisanship Strength			-0.037 (0.219)	0.031 (0.227)	0.029 (0.230)
Affective Polarization			-0.007 (0.005)	-0.006 (0.005)	-0.006 (0.005)
7-Point Ideology			-0.044 (0.140)	0.033 (0.146)	0.020 (0.148)
Ideological Identity Strength			0.105 (0.144)	0.112 (0.150)	0.055 (0.153)
Media Consumption Scale			-0.023 (0.121)	-0.039 (0.126)	-0.079 (0.127)
Male				-0.330 (0.226)	-0.347 (0.228)
Age				-0.015* (0.009)	-0.013 (0.009)
College				0.823** (0.370)	0.894** (0.375)
Social Media Expressor					0.586** (0.232)
Constant	0.966*** (0.102)	0.968*** (0.103)	0.570 (0.743)	0.079 (0.886)	-0.364 (0.905)
Observations	483	483	482	464	464
Log Likelihood	-283.326	-282.966	-278.133	-266.747	-263.575
Akaike Inf. Crit.	570.652	569.933	574.265	557.495	553.150

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

number of survey-takers (gray), the subset who consented to participate in a chat (green), and the subset of those who consented but failed to complete a chat (red). Though we lack precise data on participants' waiting times, Figure 3 shows that most participants were successfully paired with a partner within the 5-minute maximum waiting time, and since only one of each pair waits at all, the mean waiting time was less than 2.5 minutes. We set a target N of 500 for Study 1, and, as seen in Figure 3, the study was mostly complete within 1 hour of launch. Faster recruitment is possible, and theoretically depends only on the number of active crowdworkers at the time of a study, since ReChat does not have a set capacity. However, we recommend that the researcher monitor recruitment in real time, to guard against the risk of waiting-room timeouts in the event of a recruitment slowdown, and in practice this implies running recruitment for larger studies in modest-sized batches.

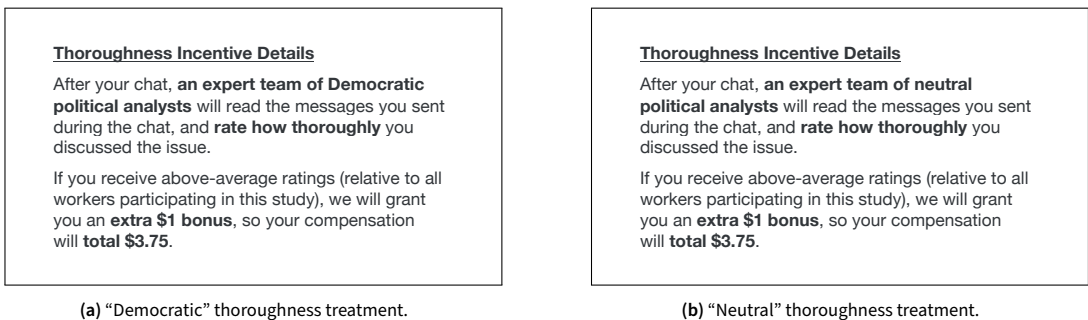
Overall, 72% of subjects recruited to the Study 1 survey consented to chat. We analyzed predictors of chat-consent with logistic regression (see Table 1), and found that extroversion ( $p = .052$ ) and self-monitoring ( $p = .037$ ) both individually predicted chat participation (see Table 1, columns 1 and 2). Neither retained significance in more saturated models (columns 3-5), possibly because extroversion and political interest were correlated (see Appendix 5). Still, the importance of extroversion was qualitatively evident in free-text explanations offered by non-participants (see Appendix 6). Among the political variables we measured, only political interest was a significant ( $p = .028$ ) predictor of chat participation, and of the demographic characteristics we measured, college education was the largest and most robustly significant ( $p = .094$ ) predictor of chat participation. Age was predictive in some models, however this was not robust to alternative specifications. Finally, we found a large and significant ( $p = .017$ ) relationship between chat participation and using social media to express one's political views.

We found that 92% of respondents who consented to participate in a chat actually completed one. Of the remaining 8%, only one participant timed out in the waiting room. It should be noted that some of these non-completes may reflect participant non-compliance (e.g. entering a waiting room and then neglecting to participate in the rest of the study), so this gives a conservative measure of ReChat's technical performance. Further analyses can be found in Appendix 8, which compares mobile and desktop users, and Appendix 13, which provides a random sample of actual chat transcripts.

### 3.4 Conducting Experiments with ReChat

Researchers can easily conduct a variety of experiments with ReChat. One category of experiments involves creating multiple ReChat templates and randomizing which template each participant is assigned to, in order to compare the effects of conversations with different topics, durations, or group sizes (see Appendix 1 for a complete list of adjustable variables in ReChat). Another class of experiments applies an informational, priming, framing, vignette, or incentive treatment, as in a typical survey experiment, and uses the chat to collect a behavioral outcome measure.

We demonstrate this latter type of experiment through a manipulation we applied to the thoroughness incentive bonuses. Recall that participants could receive a bonus payment for being more “thorough” than the average participant. As an experimental treatment, we informed participants that their chat messages would be judged for their “thoroughness” and that participants with above-average thoroughness would receive an additional monetary bonus, and we manipulated this text<sup>6</sup> (see Figure 4) to state either that participants’ messages would be judged by “Democratic political analysts” or “neutral political analysts.” This was intended to simulate accountability to different “imagined audiences” (Marwick and boyd 2011): accountability to a co-partisan audience (since all participants were Democrats), or accountability to a neutral audience. We then tested for treatment effects on the length of participants’ messages, as a proxy for their thoroughness.



**Figure 4.** Comparison of (a) “Democratic” thoroughness judge and (b) “neutral” thoroughness judge treatment conditions, as displayed to participants in Study 2.

### 3.5 Data Processing & Analysis in R

We provide an open-source software package, `rechat`, to process the chat data that is exported from ReChat into a format convenient for analysis in R. We provide `rechat` in a GitHub repository <sup>7</sup>

6. Treatment occurred after deciding whether to participate, in order to avoid contaminating self-selection in Study 1.

7. <https://doi.org/10.7910/DVN/PWHPTG>

with instructions for its use. We also provide replication materials<sup>8</sup> that offer examples new users can follow. Here, we describe our analyses at a tutorial level of detail.

First, we downloaded the chat data from ReChat (as a CSV). Then, using the `parseChat()` function, we read all data into R as a list of chat dataframes, and summarized each message's length using the `featurizeChat()` function. `featurizeChat()` can apply any function (that takes a string as its argument) to each chat message, and appends the values returned as a new column in the chat dataframes. This allows researchers to apply functions from existing text analysis packages. This can be used to create sentiment scores, add semantic embeddings, or even apply modern language models to generate highly sophisticated content analyses. We, however, wanted to conduct a very simple analysis for demonstration purposes, and so summarized message character length using the base R function `nchar()`.

Then, we used `summarizeChat()` to aggregate from the message to the participant level, and bind participant-total-character-counts to the survey data downloaded from Qualtrics. This links the survey data to the participant outcomes measured from the chat data using participants' completion codes (see Section 3.2). For example, we added a column "char\_count" to the survey dataframe, which we used as a proxy for thoroughness. By a similar process, we also constructed measures of message count, word count, and unique word count. We then analyzed predictors of these loquaciousness proxies, in a linear<sup>9</sup> regression framework. The follow pseudocode summarizes these steps:

```
library(rechat)

#Parse Chat Data Downloaded from ReChat
chat_data <- parseChat("path/to/downloaded_file.csv")

#Featurize Transcripts with Character Counts
chat_data <- featurizeChat(chat_data,
                           featurization_function = nchar)

#Bind Character Sums to Survey Data
survey_data <- qualtrics::fetch_survey(...)
survey_data <- summarizeChat(survey_data, chat_data,
                             chat_feature_name = "nchar",
                             summary_function = sum,
                             summary_feature_name = "char_count")
```

---

8. <https://github.com/willschulz/introducing-rechat>

9. As a robustness check, Appendix 12 implements log-linear models, which do not change our results.

```
#Analyze Predictors
mod <- lm(char_count ~ t + ideo_extremity + male, data = survey_data)
```

Exploratory analyses of the Study 1 data (see Appendix 10) indicated participants' loquaciousness was significantly associated with male gender and liberal political ideology (which, in an all-democrat sample, is equivalent to ideological extremity). We also found that loquaciousness was significantly affected by the experimental treatment described in Section 3.4, such that participants in the neutral accountability condition were more thorough than those in the co-partisan accountability treatment.

We implemented Study 2 as a pre-registered<sup>10</sup> replication of these analyses. Our planned analyses of character count (see Table 8, Model 1) did not replicate the Study 1 findings regarding gender and the thoroughness treatment, but did replicate the finding regarding ideology. In this all-Democrat sample, more liberal participants were (again) significantly more loquacious than more moderate participants. Exploratory analyses of *message* count (Table 2, Model 2) offer suggestive evidence of a neutral accountability treatment effect on loquaciousness, and an association with male gender, but message count was not the pre-registered outcome variable (see Appendix 11 for detailed analyses).

**Table 2.** Loquaciousness (Study 2)

	<i>Dependent variable:</i>			
	# Characters (Pre-Reg)	# Messages	# Words	# Unique Words
	(1)	(2)	(3)	(4)
Treatment				
(Neutral Accountability)	14.124 (37.953)	1.253** (0.528)	4.730 (7.018)	1.350 (3.985)
Ideological Extremity	49.128** (21.520)	0.516* (0.266)	8.588** (3.922)	5.887*** (2.249)
Male	-19.456 (39.386)	1.128** (0.513)	-5.114 (7.155)	-1.284 (4.128)
Constant	803.724*** (51.802)	7.022*** (0.610)	147.349*** (9.485)	99.950*** (5.484)
Observations	390	390	390	390
R <sup>2</sup>	0.012	0.033	0.013	0.015
Adjusted R <sup>2</sup>	0.004	0.025	0.005	0.007
Residual Std. Error (df = 386)	375.307	4.888	68.442	39.000
F Statistic (df = 3; 386)	1.574	4.364***	1.667	1.965

Note:

\* p<0.1; \*\* p<0.05; \*\*\* p<0.01

10. <https://osf.io/atfyq>. We deviated from our pre-registration in a conservative direction, using 2-tailed instead of 1-tailed tests, and collected less than our target sample size due to budgetary and recruitment constraints (see Appendix 7).

#### 4. Conclusion

This paper has introduced tools to lower the start-up costs of studying online textual conversations in a controlled setting. We hope this will help advance scholarship by removing barriers that prevent a broader set of researchers from pursuing such studies.

ReChat is designed to serve the needs of the typical social scientist, rather than conversation specialists. It is not well suited for studying fine-grained aspects of platform design or user interfaces (since these are not adjustable), for which custom-built platforms are inherently more appropriate. Instead, ReChat is designed to expand the space of feasible study designs for the typical academic who is skilled at designing online surveys. For example, it is an ideal tool for behaviorists who wish to augment their survey experiments with a chat component. It is also an excellent entry-level tool for early-career researchers to explore their interest in conversation as a treatment or an outcome. We believe this best serves the field at large, by making it easier for more social scientists to study live social interactions, using their existing research skills.

Our demonstration studies illustrate how ReChat can be integrated into a researcher's existing infrastructure for running online surveys in Qualtrics and analyzing data in R. Although ReChat can be used on its own, integrating it with existing tools lowers start-up costs for new conversation researchers. Our demonstration studies also illustrate two important features of chats that researchers should be mindful of: not all survey participants are equally willing to participate in a chat, and not all chat participants contribute equally to the content of the chat. In particular, we found that chat participants were significantly older, more interested in politics, and more likely to be college-educated than non-participants (with some evidence that extroversion and self-monitoring predicted participation as well), and that (at least in chats between Democrats) ideological leftists were more loquacious than moderates.

These analyses were intentionally simple. Conversational transcripts are amenable to a wide variety of analyses, encompassing longstanding and diverse qualitative approaches (e.g. Schiffrin, Tannen, and Hamilton 2001), as well as a burgeoning contemporary literature on computational methods for analyzing conversations (e.g. Rossiter 2022), but reviewing or demonstrating these methods would be beyond the scope of this article. However, we hope that ReChat will contribute indirectly to the methodological advancement of conversation analysis, by enabling the collection of conversation data and thereby enlarging the community of scholars with an interest in its analysis.

We also hope that expanding the availability of live chats as a component of online surveys will expand the range of research applications to which such chats are put. There are certainly many applications pertaining to social media: for example, we find that ideological extremists were particularly loquacious – a pattern that has also been identified on social media platforms, as a contributor to the polarization of online discourse writ large (e.g. Hughes 2019; Bail 2021, see also Appendix 11.2). Further research should investigate the reasons for this, along with other features of textual interactions that contribute to polarization and other concerns about social media. But a much broader range of applications is available, especially when one remembers that the survey itself was first developed as a means of quantifying “public opinion,” which was originally conceptualized to include conversations and other interactive forms of political expression (Delli Carpini 2011). Surveys flatten the social aspects of public opinion to resemble something like a secret ballot (Blumer 1948), in the interest of being more systematic and scaleable (compared to, say, focus groups). But ReChat creates a systematic and scaleable way to reintroduce social interaction to the study of public opinion. By putting respondents in conversation, we can observe how people advocate for their positions in practice, whether they engage in self-censorship or preference falsification when faced with resistance, and how they iteratively update their opinions and integrate new ideas into their worldview. By rendering these behavioral manifestations of opinion observable through social interaction, chats can enhance surveys’ fulfillment of their original purpose.

We therefore encourage scholars to consider how their experiments on attitudinal outcomes might be adapted to a chat context. This is easy to implement due to the integration of ReChat with Qualtrics, and opens innumerable possibilities. For example, scholars of political identity might treat their participants with different identity primes, and observe how these treatments affect live inter-partisan interactions, rather than using attitudinal proxy measures of partisan animosity. Researchers concerned with misinformation can test interventions to attenuate their spread, by observing whether or not participants repeat false claims to chat partners. Similarly, to study political information transmission, researchers might have some participants read various news articles prior to the chat, and measure knowledge of pertinent facts afterwards. By directing participants to chat templates according to variables measured or randomized in Qualtrics, a practically infinite set of research designs is possible, in addition to those furnished by the variables, such as size, duration, and topic, that can be manipulated within a ReChat template itself (see Appendix 1 for full list).



Of course, scholars should think carefully about how the chat setting both mimics and differs from the kinds of conversation to which they wish to generalize. For example, it is difficult to convene research chats between real-world friends and relations – our most common interlocutors in daily life. Chats between strangers necessarily lack the comfort, mutual knowledge, and expectations that come from a pre-existing relationship. Strangers do sometimes converse on social media platforms (which may explain why this is such a prevalent framing for chat studies), but even these interactions are attended by a host of performative and reputational considerations, identity signals, and platform-specific norms that are not necessarily present in a research chat. Researchers should carefully consider and acknowledge how such divergences affect the external validity of their designs.

However, this “blank slate” nature of research chats can be an asset, since it allows the researcher to intervene in aspects of conversation that are normally fixed. For example, a chat experiment can manipulate participants’ knowledge of their interlocutors’ partisanship, views, and social identities. Experimenters can also simulate various social motivations, as in our application of “thoroughness” bonuses that sought to simulate accountability to a neutral or co-partisan audience. So, even when they do not perfectly simulate real-world interactions, research chats can support a basic science of the mechanisms of political conversation.

By making it easier for more people to run such studies, we hope that ReChat will help support a new generation of survey experiments that allow social scientists to study live social interactions, at a large scale and at a modest cost. We hope to foster this work, and look forward to seeing how it can complement the designs that have already been implemented, explore the design space we have mapped above, and discover new and innovative ways of using chats to advance the study of opinion and political science writ large.

**Acknowledgments** The authors gratefully acknowledge advice and input from Andy Guess, Tali Mendelberg, Rory Truex, Yph Lelkes, Kokil Jaidka, Alvin Zhou, participants at SPSA 2023, and participants in the American Political Behavior Research Group at Princeton University.

**Funding Statement** This research was supported by a grant from the Princeton University Data-Driven Social Science Initiative.

**Data Availability** The data, code, and survey instruments required to replicate all analyses in this article are available at the Journal of Experimental Political Science Dataverse within the Harvard Dataverse Network, at: <https://doi.org/10.7910/DVN/PWHPTG>.

**Conflicts of Interest** Xiaoxiao Shen holds the patent<sup>11</sup> for ReChat, however ReChat is available to academic researchers on a pay-what-you-can donation basis. The demonstration studies presented in this article were designed, implemented, and analyzed by William Small Schulz, and were in no way influenced by the financial interests of either author.

**Ethics Statement** This research was approved by the Princeton University IRB (Protocol #13887). This research adheres to APSA's Principles and Guidance for Human Subjects Research, as discussed in further detail in Appendix 14.

## References

- Abrams, Dominic, Margaret Wetherell, Sandra Cochrane, Michael A. Hogg, and John C. Turner. 1990. Knowing what to think by knowing who you are: Self-categorization and the nature of norm formation, conformity and group polarization\*. *British Journal of Social Psychology* 29 (2): 97–119.
- Almaatouq, Abdullah, Joshua Becker, James P. Houghton, Nicolas Paton, Duncan J. Watts, and Mark E. Whiting. 2021. Empirica: a virtual lab for high-throughput macro-level experiments. *Behavior Research Methods* 53 (5): 2158–2171.
- Argyle, Lisa P., Christopher A. Bail, Ethan C. Busby, Joshua R. Gubler, Thomas Howe, Christopher Rytting, Taylor Sorensen, and David Wingate. 2023. Leveraging AI for democratic discourse: Chat interventions can improve online political conversations at scale. *Proceedings of the National Academy of Sciences* 120 (41): e2311627120.
- Ashton, Michael, and Kibeom Lee. 2009. The HEXACO–60: A Short Measure of the Major Dimensions of Personality. *Journal of Personality Assessment* 91 (4): 340–345.
- Bail, Christopher A. 2021. *Breaking the social media prism: how to make our platforms less polarizing*. 1st. Princeton: Princeton University Press.
- Bail, Christopher A., Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, M. B. Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. 2018. Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences* 115 (37): 9216–9221.

---

11. *Permanent U.S. Patent No. 11,843,470*. (<https://patents.google.com/patent/US11843470B2/en?inventor=xiaoxiao+shen&dq=xiaoxiao+shen>)

- Berinsky, Adam J., and Howard Lavine. 2011. Self-monitoring and political attitudes. In *Improving Public Opinion Surveys: Interdisciplinary Innovation and the American National Election Studies*, edited by John H. Aldrich and Kathleen M. McGraw, 29–45. Princeton University Press.
- Blumer, Herbert. 1948. Public Opinion and Public Opinion Polling. *American Sociological Review* 13 (5): 542.
- Bor, Alexander, and Michael Bang Petersen. 2021. The Psychology of Online Political Hostility: A Comprehensive, Cross-National Test of the Mismatch Hypothesis. *American Political Science Review*, 1–18.
- Brady, William J., Julian A. Wills, John T. Jost, Joshua A. Tucker, and Jay J. Van Bavel. 2017. Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences* 114 (28): 7313–7318.
- Broockman, David, and Joshua Kalla. 2016. Durably reducing transphobia: A field experiment on door-to-door canvassing. *Science* 352 (6282): 220–224.
- Broockman, David E, Joshua L Kalla, and Sean J Westwood. 2023. Does Affective Polarization Undermine Democratic Norms or Accountability? Maybe Not.
- Burnstein, Eugene, and Amiram Vinokur. 1977. Persuasive argumentation and social comparison as determinants of attitude polarization. *Journal of Experimental Social Psychology* 13 (4): 315–332.
- Carlson, Taylor N., and Jaime E. Settle. 2016. Political Chameleons: An Exploration of Conformity in Political Discussions. Publisher: Springer US, *Political Behavior* 38 (4): 817–859.
- Chen, Daniel L., Martin Schonger, and Chris Wickens. 2016. oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance* 9:88–97.
- Combs, Aidan, Graham Tierney, Brian Guay, Friedolin Merhout, Christopher A. Bail, D. Sunshine Hillygus, and Alexander Volfovsky. 2023. Reducing political polarization in the United States with a mobile chat platform. *Nature Human Behaviour* 7 (9): 1454–1461.
- Cramer Walsh, Katherine. 2004. *Talking about politics: Informal groups and social identity in american life*. Chicago: University of Chicago Press.
- D'Agostino, R.B. 1970. Transformation to Normality of the Null Distribution of G1. *Biometrika* 57 (3): 679–681.
- Delli Carpini, Michael X. 2011. Constructing Public Opinion: A Brief History of Survey Research. In *The Oxford Handbook of American Public Opinion and the Media*, 1st ed., edited by George C. Edwards, Lawrence R. Jacobs, and Robert Y. Shapiro, 284–301. Oxford University Press.
- Druckman, James N., Samara Klar, Yanna Krupnikov, Matthew Levendusky, and John Barry Ryan. 2022. (Mis-)Estimating Affective Polarization. *The Journal of Politics*, 1106–1117.
- Druckman, James N., Matthew S. Levendusky, and Audrey McLain. 2018. No Need to Watch: How the Effects of Partisan Media Can Spread via Interpersonal Discussions. *American Journal of Political Science* 62 (1): 99–112.
- Freelon, Deen. 2018. Computational Research in the Post-API Age. *Political Communication* 35 (4): 665–668.

- Groenendyk, Eric, and Yanna Krupnikov. 2021. What Motivates Reasoning? A Theory of Goal-Dependent Political Evaluation. *American Journal of Political Science* 65 (1): 180–196.
- Grönlund, Kimmo, Kaisa Herne, and Maija Serälä. 2015. Does Enclave Deliberation Polarize Opinions? *Political Behavior* 37 (4): 995–1020.
- Huddy, Leonie, Lilliana Mason, and Lene Aarøe. 2015. Expressive Partisanship: Campaign Involvement, Political Emotion, and Partisan Identity. *American Political Science Review* 109 (1): 1–17.
- Hughes, Adam. 2019. *A small group of prolific users account for a majority of political tweets sent by U.S. adults.*
- Insenberg, D. J. 1986. Group polarization: A critical review and meta-analysis. *Journal of Personality and Social Psychology* 50 (6): 1141–1151.
- Jaidka, Kokil, Alvin Zhou, Yphtach Lelkes, Jana Egelhofer, and Sophie Lecheler. 2021. Beyond Anonymity: Network Affordances, Under Deindividuation, Improve Social Media Discussion Quality. *Journal of Computer-Mediated Communication*, zmab019.
- Kalla, Joshua L., Adam Seth Levine, and David E. Broockman. 2022. Personalizing Moral Reframing in Interpersonal Conversation: A Field Experiment. *The Journal of Politics* 84 (2): 1239–1243.
- Karpowitz, Christopher, and Tali Mendelberg. 2014. *The Silent Sex: Gender, Deliberation, and Institutions*. Princeton: Princeton University Press.
- Klar, Samara. 2014. Partisanship in a social setting. *American Journal of Political Science* 58 (3): 687–704.
- Lazarsfeld, Paul F., Bernard Berelson, and Hazel Gaudet. 1948. *The People's Choice: How the Voter Makes Up His Mind in a Presidential Campaign*. 2nd. New York: Columbia University Press.
- Lee, David S. 2009. Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. *REVIEW OF ECONOMIC STUDIES*.
- Leviton, Lindsey C., and Brad Verhulst. 2016. Conformity in Groups: The Effects of Others' Views on Expressed Attitudes and Attitude Change. *Political Behavior* 38 (2): 277–315.
- Marwick, Alice E., and Danah boyd. 2011. I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society* 13 (1): 114–133.
- Mendelberg, Tali, Christopher F. Karpowitz, and J. Baxter Oliphant. 2014. Gender inequality in deliberation: Unpacking the black box of interaction. *Perspectives on Politics* 12 (1): 18–44.
- Munger, Kevin. 2017. Experimentally Reducing Partisan Incivility on Twitter, 1–31.
- . 2019. The Limited Value of Non-Replicable Field Experiments in Contexts With Low Temporal Validity. *Social Media + Society* 5 (3): 205630511985929.
- Mutz, Diana C. 2002. Cross-cutting Social Networks: Testing Democratic Theory in Practice. *American Political Science Review* 96 (1): 111–126.

- Myers, David G., and Helmut Lamm. 1976. The group polarization phenomenon. *Psychological Bulletin* 83 (4): 602–627.
- Noelle-Neumann, Elisabeth. 1991. The Theory of Public Opinion: The Concept of the Spiral of Silence. In *Communication Yearbook*, 256–287. International Communication Association.
- Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David G. Rand. 2014. Structural Topic Models for Open-Ended Survey Responses: STRUCTURAL TOPIC MODELS FOR SURVEY RESPONSES. *American Journal of Political Science* 58 (4): 1064–1082.
- Rossiter, Erin. 2022. Measuring Agenda Setting in Interactive Political Communication. *American Journal of Political Science* 66 (2): 337–351.
- . 2023. *The Similar and Distinct Effects of Political and Non-Political Conversation on Affective Polarization*.
- Santoro, Erik, and David E. Broockman. 2022. The promise and pitfalls of cross-partisan conversations for reducing affective polarization: Evidence from randomized experiments. *Science Advances* 8 (25): eabn5515.
- Schiffrin, Deborah, Deborah Tannen, and Heidi E Hamilton, eds. 2001. *The Handbook of Discourse Analysis*. Massachusetts: Blackwell Publishers.
- Schulz, William, and Xiaoxiao Shen. 2024. *Replication Data for: Introducing ReChat: A Lab-in-the-Cloud for Text Discussions*. V. DRAFT VERSION. <https://doi.org/10.7910/DVN/PWHPTG>. <https://doi.org/10.7910/DVN/PWHPTG>.
- Simon, Adam F., and Tracy Sulkin. 2002. Discussion's Impact on Political Allocations: An Experimental Approach. *Political Analysis* 10 (4): 403–412.
- Strandberg, Kim, Staffan Himmelroos, and Kimmo Grönlund. 2019. Do discussions in like-minded groups necessarily lead to more extreme opinions? Deliberative democracy and group polarization. *International Political Science Review* 40 (1): 41–57.
- Tauchmann, Harald. 2014. Lee (2009) Treatment-Effect Bounds for Nonrandom Sample Selection. *The Stata Journal: Promoting communications on statistics and Stata* 14 (4): 884–894.
- Turner, John C. 1991. *Social influence*. Bristol, PA: Open University Press.
- Vermeer, Susan A. M., Sanne Kruike-meier, Damian Trilling, and Claes H. de Vreese. 2021. WhatsApp with Politics?: Examining the Effects of Interpersonal Political Discussion in Instant Messaging Apps. *The International Journal of Press/Politics* 26 (2): 410–437.
- Zaller, John. 1992. *The Nature and Origins of Mass Opinion*. Cambridge: Cambridge University Press.