Effects of Exposure to Political Protests on Political Discussion and Attitudes in Authoritarian Regimes: Evidence from Eight Natural Experiments with Chinese Social Media

Han Zhang[1]

[1]Ph.D. Candidate, Department of Sociology, Princeton University

ABSTRACT

This paper estimates the causal effect of eight protests in Hong Kong from 2012 to 14 on the frequency of political discussions and pro-democracy attitudes among witnesses from mainland China. Using the eight protests as natural experiments, I identified treatment groups that were in Hong Kong when protest occurred and control groups that left Hong Kong just before the protests. Using difference-in-differences estimators, I found that physical presence at the scene of the protests caused an enduring increase on witnesses' frequency of political discussions and reinforce preexisting attitudes. In contrast, just being in Hong Kong when protest occurred did not increase the frequency of political discussions and decreased pro-democracy attitudes. The results were consistent across four robustness checks. This paper has implications for the use of social media data to study the consequences of protests.

1 INTRODUCTION

Popular protest pursuing democracy in one region may spread to another geographically and culturally proximate authoritarian region. Such diffusion occurred in many major waves of revolution, from the 1848 European revolutions to the collapse of the former Soviet bloc and the Arab Spring (Weyland, 2009). Even if the diffusion does not happen, scholars rarely disagree that protests in one country at least increase the mobilization potential of nearby countries. It is possible because protests raise the expectations of the masses for similar events in their own countries (Hale, 2013; Kuran, 1991). Under appropriate conditions, the increased potential can facilitate cross-region protests.

The law-like assertion that exposure to political protest increase mobilization potential of a nearby authoritarian regime has several problems. First, empirical studies on protest diffusion are often only at the macro-level. Rarely do scholars examine the individual-level basis of diffusion. Part of the reason is that individual-level data on exposure to political protests are difficult to obtain. Furthermore, it is not easy to disentangle the impact of political protests from the impact of political predispositions, which makes particular subgroups, especially opposition activists, overrepresented in protest surveys (Welsh, 1981). The lack of individual-level data on exposure to protests and strong selection biases make the assumed link between exposure to protest and increased mobilization potential in non-democracies popular but largely untested.

In this paper, I take eight political protests in Hong Kong from 2012 to 2014 as the strategic site and study how these protests affected Chinese visitors who were in Hong Kong when the protest occurred. China is an ideal case for testing the link between exposure to protests and increased mobilization potential. At first sight, there is reason enough to believe that the protests in Hong Kong could lead people in China to increasingly expect something similar to be likely, given cultural similarities and geographical proximity. These favorable conditions are strengthened by the fact that large political protests have been almost extinct

in China since the Tiananmen Square protest in 1989, despite a tremendous amount of social discontents. Political protests in Hong Kong thus could serve as a signal that triggers political protests in China from seemingly nowhere (Kuran, 1997). Not surprisingly, media such as the New York Times and the National Public Radio (NPR), and scholarly work such as White (2016), all predict that the cycle of protests in Hong Kong may have a liberalizing effect on China. Due to data limitation, however, these predictions have not been tested.

To disentangle potential selection biases from the effect of seeing a protest, I use the eight protests in Hong Kong as exogenous variations that randomly expose political protests to Chinese visitors. I leverage precise spatial and temporal information from geolocated posts on social media to construct comparison groups that are similar otherwise, although some were in Hong Kong during the protests and may have been exposed to political protests while others had already left Hong Kong just before the protests and were not exposed to the protests. I use machine learning techniques to measure the frequency of political discussions and their attitudes toward political democracy. I collect pre- and post-treatment political discussions of both the treatment and control groups, which enables me to use difference-in-differences models to estimate the causal impact of exposure to political protests.

I find that the effect of exposure depends on the channel of exposure, and varies across individuals by their level of political discussion and pro-democracy tendency. Physical presence at political protests has an enduring impact on promoting witnesses' frequency of political discussions, and the effect is higher among users who already discussed politics frequently beforehand. Physical presence produces a polarizing effect by reinforcing the preexisting political views toward democracy. On the other hand, if Chinese visitors were in Hong Kong during the date of a protest but were not at the scene, the protest had no impact on their frequency of discussing politics after they returned China, and decreased their support for democracy. The results are consistent under four different types of robustness checks.

These findings highlight a more complex empirical understanding of the literature of social movements and authoritarian politics. First, this paper challenges the often assumed

2

link between exposure to protest and increased mobilization potential in the protest diffusion literature. Second, this paper extends the literature on the consequence of social movements, by revealing the heterogeneous effect of protest on its audience by the channel of exposure, by time, and by preexisting political behaviors and predispositions. Last, the research offers a concrete example of how "big data", combined with appropriate research designs and computational methods, can inform scholars on questions that would otherwise be hard to examine with traditional data.

2   Protests in Hong Kong and Their Potential Impact on China

Hong Kong has entered a major cycle of contention in recent years. I focus on eight large political protests in Hong Kong from 2012 to 14 (Cheung, 2014; Davis, 2015; Chan, 2014).[1] Each protest had more than 1,500 participants.[2] The last protest was the Umbrella Movement, which caught international attention in September 2014.

The Chinese government regained sovereignty over Hong Kong from the UK in 1997. Under the "One Country, Two Systems" rule, Hong Kong has kept its capitalist economic system, legal traditions and some level of political rights. However, the Chinese government indirectly control the city politics by designing an election scheme that favors businessmen (Fong, 2014, 2015; Ortmann, 2016). The government-business collusion breeds grievances against rising social inequality, especially among the city's youth (Cheng and Chan, 2017). These structural precedents are accompanied by the cultural tensions between Chinese tourists and local residents (Rowen, 2016; Wassler et al., 2018). The restricted po-

---

[1]There are other large political protests during the time period, such as the Annual New Year's March and the Annual Memorial of the Tiananmen Square Protest which takes a sit-in form. Because these protests were recurring, even though they have thousands of participants, I did not include them in my study.

[2]I do not include some smaller protests because 1) I cannot identify many witnesses from these protests and 2) there will be treatment interferences if I include too many small protests such that the dates of protests have overlaps.

litical freedom, economic inequality, and cultural conflicts were three major issues of the protests. Appendix A.1 provides details of the eight protests in Hong Kong.

While understanding the mobilization process and consequences of the protests are important for their own sake (e.g., Chan (2014); Ortmann (2016); Cai (2016)), this paper focuses on how these protests influenced visitors from China who do not live or work in Hong Kong. With political protests extremely rare in China post-1989, they shocked witnesses, according to the New York Times and others.[3] Furthermore, the mass media in China rarely report political protests due to government pressure (Stockmann and Gallagher, 2011; Qin et al., 2018).[4] Therefore, observing political collective action during their travels to other regions is one feasible venue for Chinese visitors to experience or know about political protests.

The chance of seeing a protest abroad is not small: over 40 million Chinese people visited Hong Kong for tourism in 2014, and over 100 million visited foreign countries in total (Wassler et al., 2018). These travelers belong to the rising Chinese middle class, spending more than $128 billion on foreign travel. Moreover, protesters in Hong Kong strategically choose to protest in the central business regions, near government offices and major parks, which are also the tourist sites for most Chinese visitors (Cheng and Chan, 2017). Some of them try to persuade tourists that "it's important that we stay here to sway their hearts and minds since they'll go back to China afterward"(Rowen, 2016), while others try to pressure Chinese tourists who are thought to be one source of cultural conflicts. In sum, the frequency

---

[3]Gough, Neil and Ramzy, Austin, "Mainland Chinese Tourists Get a Glimpse of Rebellion", New York Times `http://www.nytimes.com/2014/10/02/world/asia/for-mainland-tourists-in-hong-kong-standoff-is-both-inconvenience-and-inspiration.html?emc=eta1&_r=0,`
Jacobs, Andrew, "Support for Protesters Is Hard to Find on the Streets of Beijing", New York Times, `http://www.nytimes.com/2014/10/10/world/asia/in-beijing-young-chinese-see-little-to-cheer-in-hong-kong-protests.html?_r=0,`
Kuhn, Anthony, "Hong Kong Protests Offer A Revelation To Mainland Chinese", NPR, `http://www.npr.org/2014/10/01/353041486/hong-kong-protests-offer-a-revelation-to-mainland-chinese`

[4]I used one of the largest electronic searchable newspaper databases in China, the WiseNews, to filter reports about the Umbrella Movement in Hong Kong. The website is `http://wisesearch6.wisers.net`. The WiseNews has 770 sources, including both national and regional level sources. Among these newspapers, there was only one report about the protest on the first day of the event, and there were only 30 reports within the next 3 days about the protest, all using very similar titles criticizing the protests. The critical reports may have been coordinated by government propaganda offices. The scarcity of reports sharply contrasts the media coverage of Umbrella Movements worldwide.

of events, the overlap between protest sites and tourists sites, and the enormous number of visitors create many opportunities for the Chinese middle-class visitors to be exposed to a political protest during their travel to Hong Kong.

2.1   Consequences of Protest on Exposed Populations

This paper focuses on two outcomes from exposure to protests: changes in the frequency of political discussions and political attitudes toward democracy. Everyday political talk has attracted attention for its ability to promote a healthy democracy (Gamson, 1992; Eliasoph, 1998). Frequent political discussions are associated with regular political participation and propensities to accept and participate in political protests (Schussman and Soule, 2005; Verba et al., 1995; Kwak et al., 2005). In authoritarian regimes, political talk is even more important, because much of everyday political resistance is performed as *talking* rather than *doing* (Johnston, 2005, p.108), give the risk of collective action, and limited institutional channels of political participation.

Traditional political talk in non-democracies occurs in private or "free spaces", restricted among trusted family members or friends (Polletta, 1999). In the age of the Internet, online political talks, such as posting to social media, has been theorized as a new, independent type of political participation (Gibson and Cantijoch, 2013). Moreover, the Internet provides arguably the only public sphere in which citizens can discuss dissident contents in non-democracies (Tufekci and Wilson, 2012; Shirky, 2011, p.366). Empirical research consistently finds that political discussions on social media predict participation in collective action (Tufekci and Wilson, 2012; Valenzuela et al., 2012; Zúñiga et al., 2009).

Turning now to political attitudes, theorists and empiricists agree that pro-democracy attitudes held at the individual level are conducive to the installation of democracy (Inglehart, 2003; Stevens et al., 2006). The existing literature finds a wide range of explanations for pro-democracy attitudes, including macro-level cultural values (Almond and Verba, 1963), social structure (Lipset, 1994; Rueschemeyer et al., 1992), instrumental considerations re-

garding economic performances (Wang, 2007; Dalton, 1994), and political learning though media, civic associations and education (Mattes and Bratton, 2007; Finkel and Smith, 2011).

While the role of online political talk in mobilization has been theorized and empirically examined, the reverse story, namely how exposure to protests may encourage political discussions, has rarely been studied. Similarly, the literature has little examined the relationship between exposure to political protests and pro-democracy attitudes. The next section develops a theoretical framework toward understanding how and why spatial exposure to political protests may induce changes in political discussion and attitudes.

3   THEORY AND HYPOTHESES

3.1   Spatial Diffusion of Protests

Spatial diffusion of collective action has been an important and intensively studied area. While spatial proximity is not a sufficient condition for diffusion, it certainly increases the likelihood of diffusion (Biggs, 2005). Researchers often assume that seeing political protests in one country can lead people in other countries to increasingly expect something similar to be likely in their own countries (Hale, 2013). This mechanism is dubbed as the "attribution of similarity" (Tarrow, 2005). Even though a mobilization may not succeed, scholars often attribute its failure to increased government repression or co-optation (Lynch, 2014), instead of decreased mobilization potential among the masses (Klandermans and Oegema, 1987).

However, most empirical evidence on this widely held assumption is based at the event-level. Not many have zoomed in to study the individual basis of diffusion. This approach might be explained by the "mobilization turn" in the study of contentious politics, which often assumes that there is always enough discontent in any society and as a result focuses on the mobilization process instead. Recent critiques have urged a reexamination of the formation of individuals' political opinions, which in turn affect their responses to different mobilization attempts (Walder, 2009; McAdam and Boudet, 2012).

Burgeoning studies in democracies have found that spatial proximity to political protests shapes people's political opinions about the movement itself, and more general political attitudes toward related social issues. Wallace et al. (2014) find that during the 2006 Immigrant Rights Marches, spatial proximity (within 50 miles) to multiple small marches had a positive impact on political efficacy, whereas proximity to large-scale marches led to lower feelings of efficacy. Using the same data, Branton et al. (2015) find that residing in the county of a march altered Latinos' immigration policy preferences. Andrews et al. (2016) find that living in a county where a sit-in occurred make white Southerners more likely to support for the sit-in, while such effect does not hold for counties near a sit-in county.

These studies are the major building blocks of this article, but they are also limited in several ways. First, these studies are limited in their causal power. Andrews et al. (2016) recognizes that the result could be explained by self-selection mechanisms, namely that civil rights protest was more likely to emerge in sympathetic areas. To reduce the concern of self-selection, Wallace et al. (2014) and Branton et al. (2015) used the opportunity that the 2006 Latino National Survey was performed during the Immigrant Rights Marches, thus creating comparison groups between respondents that were surveyed before or after local events. Because each respondent was only surveyed once, however, other time-varying covariates that are causally prior to the event could also have produced the changes as observed in these studies.[5]

Second, these studies conflate different *channels* through which spatial proximity produces changes in the exposed audience's political behaviors and attitudes. Spatial proximity to protest can mean increased chances of physical presence at the protests, increased local media coverage, information spreading through friendship networks, and more opportunities to be contacted by social movement organizations (Andrews and Biggs, 2006). Third, because the data they utilize are drawn from a single point in time, it is impossible to examine the duration of the effect. Last, while existing studies explored variations of the effects

---

[5]See (Legewie, 2013, pp. 9 - 10) and McKenzie (2012) for discussions of the threat to the causality of similar design.

along basic sociodemographic characteristics, none of the existing studies can collect data on pre-treatment measures of their outcomes. Therefore, they do not theorize how protest may influence an audience differently through their pre-treatment outcome distributions.

Next, I explicate how the channel of exposure matters, whether the effect will endure, and how the effect may vary across subgroups with different levels of political discussions and pro-democracy attitudes. A discussion of the causal effects is left to Section 4.2.

3.2 Physical and Mediated Exposure to Protest

I focus on two particularly important channels: direct, physical exposure to political protests and mediated exposure.[6] In most collective action, there is physical co-presence of a large number of people within a limited space, which exposes the event to bystanders at the scene who are neither protesters nor their targets. Although neglected by movement scholars, bystanders who witness an event are crucial to any collective action. Movement theorists have conceptualized contentious politics as "dramas in which protagonists and antagonists compete to affect audiences' interpretations of power relations" (Benford and Hunt, 1992). In this dramaturgical view, an audience is required for any protest (Tilly, 1993). A major staging task of protest actors, who generally have relatively fewer resources, involves promotion and publicity activities to attract audiences. Physical co-presence, especially at places with symbolic meanings or high population densities, is a natural way to command public attention and create audiences whom the protesters want to persuade (Sewell Jr, 2001; Tilly, 2000). Yet, as Benford and Hunt (1992) note, physical co-presence can create a double-edged sword, namely "audience segregation", in the sense that the unintended audience also appears at the scene and interprets the protest in a way movement actors do not expect.

Alternatively, mediated channels, including newspapers, radio, television and the Inter-

---

[6]Beyond these two, other important channels include social networks and contacts with social movement organizations (Andrews and Biggs, 2006)

net, indirectly exposure the movement to audiences.[7] The chance that the media can actually reach its target audience is smaller: one can hardly not notice a protest if you pass by it, but one can easily skip reports about protests in the media. Furthermore, media reports often diminish the emotional energy of a protest by neutralizing their reports (Smith et al., 2001; McCarthy et al., 1996), which may be unfavorable for mobilization. However, mass media naturally reach a much larger audience pool (Andrews and Caren, 2010). It makes Koopmans and Olzak (2004) to think that "it is no longer the co-present public that counts most, but the mass audience that sits at home and watches or reads the media coverage of the demonstration".

Research on the role of media in collective action is overwhelming, but estimates of the impact of collective action on audiences at the scene are scant (Gamson, 2004). Among the pioneering studies on the spatial impact of protests, Wallace et al. (2014) and Andrews et al. (2016) hint that spatial proximity to political protests matters because it increases local media coverage, but they ignore physical presence as a noteworthy channel. Their model also cannot distinguish whether the effect is from mediated, physical, or other channels.

The failure to separate physical exposure from mediated exposures can be especially problematic in authoritarian regimes. This is because authoritarian regimes often impose strict controls on media (Egorov and Sonin, 2011; Mcmillan and Zoido, 2004; Qin et al., 2018), such that the proposed role of local media needs to be reconsidered. The exception is Schwartz (2016), whose field experiment in Mexico shows that protest is a double-edged sword: people will become more enthusiastic when they hear of protests through the news, but they disengage when faced with an actual protest because the protests interrupt their daily lives.

How does physical exposure to political protests differ from mediated exposure? I argue that physical exposure generates a much stronger impact than mediated exposure at the individual level. First, physical witnesses share the most important ingredients of generat-

---

[7]Tarrow (2005) denotes the diffusion of protest through media as *non-relational* diffusion, and calls the diffusion through brokers as *mediated* diffusion.

ing "emotional energy"—the physical assembly of people with high density, clearly bounded relations (between participants and targets), and a shared focus of attention (Collins, 2001, 1993a)—which may have a life-long impact on participants (McAdam, 1989, 1999; Yang, 2000). While the original theory of Collins focused on how emotional energy can foster solidarity among participants, it is reasonable to expect that emotional energies can also spread to bystanders. This is because spatial concentration often fosters mobilization, by creating an atmosphere of excitement and anger that can often elicit spontaneous and individualistic responses to events (Zhao, 1998, 2000; Snow and Moss, 2014).

On the flip side, the study of "media effects" in sociology and communication research has yet to reach a consensus on whether media can indeed persuade people, or if correlations between mediated exposure and behavior changes are due to self-selection into media sources (Stroud, 2008). Specifically, communication scholars have found that face-to-face communications are more effective than mediated exposure regarding political persuasion (Carpini et al., 2004; Gastil, 2000).

Based on these results, I hypothesize that:

Hypothesis 1. *Physical exposure leads to a much larger effect than mediated exposure at the individual level.*

3.3 Temporal patterns of effect

If protest exerts any influence on bystanders, how long does the influence endure? Existing studies only provide a point estimate on the impact of protests on bystanders (Andrews et al., 2016; Wallace et al., 2014), but there are two groups of studies that answered the questions of the duration of effect of protests on participants. The emotional energy theory of (Collins, 1993b) will predict a short-term effect unless the bystanders are successfully mobilized into participating the political actions. In contrast, the second group show that exposure to certain social movements may lead to life-long consequences on the participating individuals (McAdam, 1989, 1999; Yang, 2000; Mazumder, 2018). The empirical evidence of

this group, however, often comes from some of the biggest mass popular movements, such as the Civil Rights Movement of the Cultural Revolution. It is not clear whether their findings can be generalized into the protests studied in this paper, which is big but has yet to proven to be as profound. In sum, existing studies on the consequences of movements either suggests ultra-short or an ultra-long effect, but the scope condition of these protests is not clear.

In my case, I posit that the duration of the effect of protests on bystanders in should sit somewhere in between, consider some characteristics of the protests in Hong Kong. First, it should not be a one-time reactionary response, because the exposure to political protests provides audience resources for talking politics (Gamson, 1992; Walsh, 2010; Perrin, 2006). Resources are the "evidence" people use to support their political discussions. Exposure to protests provides valuable experiences about political protests, such as personal stories or anecdotes Gamson (1992). Furthermore, simply being in Hong Kong allow Chinese visitors to access various media sources, which they may have troubles accessing when they returned to China. Last, real-world events may give people narrative resources with which they can use to compare their local political atmosphere, which allows expressions of political concern while avoiding the risk of incurring conflict (Eliasoph, 1998; Perrin, 2006). Once gained, the experiential, media and narrative resources can be regarded as "toolkits" that accompany witnesses for a long time, instead of disappearing directly after protests (Swidler, 1986). The resources can be activated during daily everyday political talks when an audience returns to China.

But the effect should also not be extremely long. It is because sustained political discussions often rely on institutional frameworks, such as volunteer association groups, civic education, and relatively free news media (Schofer and Fourcade-Gourinchas, 2001; Skocpol and Fiorina, 2004). Once the witnesses return to the authoritarian regimes, there are obstacles to them expressing their civic engagement in an institutional channel, which would potentially exhaust witnesses' enthusiasm. Therefore, the long-term effect on citizens from authoritarian regimes will decrease quicker than the effect on citizens in democratic regions.

Hypothesis 2. *The effect of protests endures after protests, much longer than the duration of the event if there is an effect on the first date.*

3.4 Heterogenous treatment effects

Do protest trigger more political discussions among users who already actively discussed politics beforehand, or is the opposite true? Should the former be true, these active users have the potential to become what Tarrow (2005) calls "rooted cosmopolitans", who start as locally politically active citizens, are better educated, well connected, and speak more languages. Frequent global travels, which bring with them exposure to political protests, then equip opportunities for these locally engaged citizens to link their local issues with global political activities. On the flip side, it suggests that protests have limited impact on promoting political discussions among the masses. By contrast, if the least politically active users increase their discussions the most, protests will increase the general level of political discussions among Chinese visitors.

I argue that the most politically active users may have the highest increase in their discussion frequency. As discussed previously, exposure to political protests supplies the audiences resources for political talks. Nevertheless, to comprehend the political messages during the protests and use them as resources, it requires preexisting experiences of political discussions. Studies in political deliberation find that preexisting political discussions provide citizens with a valuable "rehearsal" function to develop and practice arguments and enhance feelings of political competence as qualified political actors (Conover et al., 2002). Without the preparation and the feeling of competence, it is common for citizens to avoid talking about politics.

Moreover, previous experiences of political discussions often enhance the level of political knowledge, because discussions transmit new knowledge, force the discussants to invest more cognitive energy in understanding what they will be talking about, elaborate their understandings after the discussions (Eveland et al., 2005; Kwak et al., 2005). Knowledge is

required to understand the messages of the protests; without basic political knowledge, new information remains meaningless (Price and Zaller, 1993; Bennett et al., 2000). Frequent political discussions thus enhance political knowledge, which in turn helps frequent discussants better understand the messages of protests and helps them use the messages as resources. In sum, frequent discussion not only helps citizens prepare for future political discussions but it also indirectly promotes political knowledge, which in turn facilitates political discussions. Therefore, I hypothesize:

Hypothesis 3. *Individuals who discuss politics more frequently will have higher treatment effects than individuals who discuss politics less frequently.*

Similarly, I argue that the protests with democratic appeals are unlikely to trigger pro-democracy attitudes uniformly across the study population. Protest is often polarizing, challenging people to choose sides. While previous research in contentious politics has often focused on how such polarizing creates countermovements against the movement itself (Meyer and Staggenborg, 1996), there is reason to believe that the protests can also produce polarizing effects on general political opinions. Research on public opinion consistently finds that people tend to resist messages that oppose their preexisting political beliefs, but recognize information consistent with their prior beliefs (Zaller, 1992; Taber and Lodge, 2006). Yet, people often selectively approach or avoid information in the first place (Prior, 2007). It creates a "reinforcing spiral" in which the information selection and effect process reinforce each other to create polarization in ideological spaces (Slater, 2007; Baldassarri and Bearman, 2007). Other factors such as levels of political awareness can also contribute to the amount of information each individual has in the first place (Converse, 2006; McGuire et al., 1969; Zaller, 1992).

Drawing on these previous research, I make one distinction. For my study, the strength of selection differs by the channel of exposure. Physical witnesses do not have much room to select or avoid information: the sudden occurrence of protests directly exposes them to information that they may or may not want to assess. Their reaction to the protests thus

could be thought of as the effect of exposure alone, net of other selection processes. In contrast, the mediated exposure group has more freedom to selectively approach or avoid political protests. Avoidance should be especially strong for bystanders who were anti-democracy or moderates before the protests. The avoidance makes the polarizing effect of the protests weaker because anti-democracy and moderate individuals will just avoid protest such that the messages about pro-democracy protests will not further ignite their opposition.

Hypothesis 4. *Physical exposure will create a polarizing effect to strengthen pre-existing beliefs; mediated exposure will produce a smaller effect in absolute scale.*

4 RESEARCH DESIGN

4.1 Data

This research falls into the emerging scholarship that uses social media to understand political behaviors. I collected data from Sina Weibo (hereafter Weibo), China's largest microblogging platform.[8] Weibo is functionally very similar to Twitter, which cannot be accessed within China. On Weibo, users can post messages up to 140 characters, mention and follow other users, and repost (similar to retweets on Twitter).[9] Like Twitter, Weibo is an open platform—users do not need to be mutual friends to read others' posts.

Social media data are huge in scale and have low time latency to collect, but their non-representativeness is also evident (Salganik, 2017). Many social scientists thus treat social media as a convenient but imperfect data source (Diaz et al., 2016a). While fully acknowledging these biases, conventional ways of data collection hardly answer the questions in this research. Specifically, three things are difficult to accomplish without using social media: identifying witnesses and collect pre-protest data, estimating the effect at finer spatiotem-

---

[8]Weibo literally means micro-blog in Chinese. As of September 2016, Weibo had 132 million active daily users, and 297 million active monthly users (see `http://data.weibo.com`).

[9] One difference between Weibo and Twitter is that Weibo allows users to comment on a post without retweeting (similar to comments on Facebook).

poral resolution, and drawing causal conclusions.

First, protest surveys cannot begin before unexpected collective action, which not only makes it hard to design systematic surveys to capture witnesses at the scene but also limits their ability to capture pre-treatment audience information (Bail, 2015; Budak and Watts, 2015). The latter is especially relevant for measuring attitudes, since people may change their self-reports based on event outcomes. Second, even if protest surveys incorporate questions about audiences, they are often limited in temporal variations[10] and cannot precisely measure a bystander's distance from a protest. Third, conventional protest surveys do not have counterfactual control groups—in this case, Chinese people who were similar to witnesses in all other aspects but did not see the same protest by chance—and thus have trouble identifying causal effects. Beyond these general problems, in non-democracies, surveys about protests and political behaviors are especially difficult due to the pressure of repression, resulting in questionable data quality (Welsh, 1981; Kwiatkowski, 1992).

Using social media addresses all three problems. I identified witnesses by finding Weibo users who were in Hong Kong during the time of protests. Weibo has a function called "check-in", which adds precise GPS information to a regular post. The spatial resolution of a check-in post is within 10 meters and the temporal resolution at the level of seconds. The precise spatiotemporal information of geolocation posts allows me to identify users who were close to protests as they were happening. Based on their check-in records and using the publicly available Application Programming Interface (API) of Weibo (see Appendix B for technical details), I collected a list of 42,280 users who lived in mainland China and who checked in in Hong Kong from 2012 to 2014.[11] For all 42,280 users, I also collected their posts

---

[10]Most surveys are drawn from a single point in time, either from surveys undertaken during the demonstrations or after a protest wave has subsided. See a similar critique by (Wallace et al., 2014).

[11]Users residing in Hong Kong were excluded from the population because it is difficult to distinguish whether they actively joined protests by self-selection, or witnessed it by chance. Users living outside mainland China in Taiwan, Macau or any foreign country were also excluded since they may have witnessed political protests from their local residences. Users who were in Hong Kong for multiple months were deleted from the population since they might be living there without having declared their residence. I excluded users who were in Hong Kong on the dates of more than one protest since it indicates self-selection to participate in protests.

within six months before and after her check-in in Hong Kong. The total number of posts was 61,475,120. As a non-intrusive measure, political discussions on Weibo are less subject to intentional misreporting as is often seen in surveys in authoritarian regimes. With these posts, I can compare changes in political discussion within subjects, while existing studies have only individuals' attitudinal measure either before or after events.

4.2 Design

I use the protests as natural experiments to draw causal conclusions (Dunning, 2012). The eight protests happened due to some unexpected events. Furthermore, media reports on the protests in Hong Kong has been rare. It makes it difficult for Chinese visitors, especially those that are not politically active, to access information about the protests when they were in China, and thus unlikely to self-select into participating in the events (this point will be empirically validated in Section 5.1). Therefore, these protests provide exogenous variations on the timing when a Chinese tourist visited Hong Kong.

I constructed treatment and control groups from the 42,280 users, using the GPS and time information of their check-ins. The key idea was to identify two groups for each protest: one treatment group that was in Hong Kong when a protest occurred, and another control group that had just left Hong Kong when a protest occurred. These two groups should be similar enough on other characteristics but only differ on their time of visit to Hong Kong. Comparing the treatment and the control groups thus will reduce selection biases.

A simple design divides users by their dates of check-in: users who checked in on the dates of protests belong to the treatment group, and who checked in before protests to belong to the control group. This strategy is problematic for the control group because check-ins are not an "always-on" measure and users opt-in to add their geolocation when they post (Salganik, 2017). It is possible that a user who did not check in on the date of protests was still in Hong Kong on that day and thus could possibly be exposed. To make sure that the control group already left Hong Kong before the protest, I leveraged a specific

16

visa requirement: each Chinese visitor is allowed to stay up to seven days in Hong Kong. Therefore, if a user's check-in records in Hong Kong were before seven days of a protest, she should already have left Hong Kong when protests began. My control groups thus were users whose check-in records were within 7 - 15 days before a protest[12].

Within treatment groups, I further separated users who were exposed to the protests through physical or mediated channels. If a treated user's distance to the center of a protest was less than 1 km, I regarded her as being present at the protest.[13] The control group for the physically-present treatment group also checked in within 1km to the protest, hence having a similar probability to visit the place of protests as the treated users. If a treated user's distance was within 1 - 5 km of a protest, I regarded her as accessing protests through mediated channels. Similarly, the control group for the mediated treatment group also checked in within 1 - 5 km of the protest. I discarded users whose check-in records were more than 5 km away, making sure that the check-in location was not in mainland China. Figure 1 visualizes the construction of treatment group by the temporal and spatial distance to protest for a single protest. I repeated the process for all eight protests in my study.

[Figure 1 about here.]

Importantly, I acknowledge that spatial proximity to protests did not guarantee that treated users actually received the messages of protests either through physical or mediated channels. Spatial proximity provided an environment for them to be exposed through physical or mediated channels, compared with other people in China who could not access the protest through either channel. Specifically, the chance that a Weibo user whose distance was within 1 km of a protest indeed noticed the protest is higher because the protest itself demanded attention. The chance that a Weibo user whose distance is more than 1 km accessed the protests through media is lower; they might not pay attention to it during their travels, or they have chosen to either observe or avoid the protests. Nevertheless, this limitation is

---

[12]I choose a one-week interval to increase the size of control groups and thus improving statistical precision.

[13]This user's actual distance to the protests may be shorter, considering that all protests have more than 1,500 participants and often spans several streets.

common among the other studies in the spatial impact of political protests (Wallace et al., 2014; Branton et al., 2015; Andrews et al., 2016), none of which could measure actual reception. By using more granular spatial distance measures enabled by the GPS information, my design better approximates actual reception.

4.3   Outcome Measures

As explicated in Section 2.1, this article seeks two main outcomes: individual-level changes in the frequency of political discussion and attitude toward democracy. Given the sheer number of posts, I used machine learning methods to estimate the outcomes from social media discussions, based on human-coded training examples. Measuring political discussions and attitudes with texts is not easy. Concerning China, so do many other authoritarian regimes, the task is even harder, due to the heavy use of political sarcasm. The sarcasm makes the context-dependency of political discussions on Chinese social media is particularly strong. The necessity to consider context often at the level of sentence nullifies the common "bag-of-word" assumption used in conventional automated text analysis, which discards word orders (Grimmer and Stewart, 2013).

To address the concerns of the context dependency, I leverage the "word embedding" techniques in computer science (Mikolov et al., 2013; Pennington et al., 2014). Embedding algorithms seek to represent each word into a dense numeric vector (with often dozens to hundreds of dimensions), while at the same time per-serving the information the word's context into the vector representation.[14] Modern embeddings are obtained by finding the optimal numeric vector representation of each word, which best predict its contexts (neighboring words) in a corpus, using neural networks (**?**). Because it is designed to learn each word's representation from its contexts, word embeddings are regarded as better capturing

---

[14] Another common way to represent a word into a numeric vector is through the document-term matrix W, in which each cell W_{im} is the number of times the m-th word occurs in the i-th document.(Grimmer and Stewart, 2013). The column vector of W thus represents the word in a given corpus. This representation is a sparse representation because most words occur in only a small proportion of documents, thus making the vector contains zeros for most of the entries. Embedding is a dense representation because it has fewer dimensions with the document-term vector representation.

semantic relationships. It has achieved tremendous success in recent years and has been widely used in various natural language processing tasks (Joulin et al., 2017). In the social sciences, it has been used to evaluate word meaning changes and to map cultural networks (Garg et al., 2018; Kozlowski et al., 2018).

The second advantage of embedding techniques is that they can borrow information from external knowledge. In practice, researchers often start from pre-trained embeddings from massive datasets such as the entire Wikipedia, instead of from scratch with their own corpus, which is often tiny in scale. The general representations can then be tuned to incorporate domain-specific knowledge. Specifically, I used embeddings from Li et al. (2018), which considers the unique morphological[15] and semantic relations between Chinese words. Furthermore, it incorporates character-level information in constructing word vectors, which not only better address unknown words, but also reduces the common concern of the accuracy of word segmentation in Chinese.[16] The word embeddings are then used to perform classifications of posts into ones that are discussing politics or not and to evaluate changes in sentiments, as will be explained below.

Frequency of political discussions I operationalized the frequency of political discussions as the percentage of each user's posts about political issues among all her posts within a given timeframe. Measuring frequency as a relative quantity ensures that a user's propensity to talk on Weibo does not impact the results.

For each post, I used FastText, a tool developed by Facebook, to classify whether it is discussing politics or not (Joulin et al., 2017). To do that, research assistants collected 10000 posts that discussed politics as positive examples. Then I sample 990000 random Weibo posts, remove posts that contain politics-related words, and use them as negative examples(Bojanowski et al., 2017).[17] FastText took the pre-trained Chinese word vectors

---

[15]E.g., the lack of apparent distinctions between roots and affixes.

[16]There is no space between Chinese words such that most automated algorithms need to segment a sentence into words first, which could lead to errors.

[17]I oversample negative training data, to match the ratio between posts that are discussing politics on

from (Li et al., 2018) and tuned them based on the Weibo posts. FastText then fit a logistic regression based on the training data's vector representation to predict whether the post is discussing politics or not. The precision of the classifier is 90.3% while the recall is 70.7% based on human codings (Appendix C.2).

Attitudes toward democracy   To examine attitudinal changes, I restrict my analysis to posts that already discussed politics. I used a semi-supervised machine learning technique called label propagation (Zhu et al., 2003). Research assistants first coded a random sample of 1000 posts that discussed political issues as pro-democracy, neutral or anti-democracy (numerized as 1,0 and -1 respectively). The scores of human-coded posts were then "propagated" to uncoded posts based on their similarities. To calculate similarities between posts, following standard practices, I first took the average of embeddings of all words in a post to create a vector representation of each post (Kenter et al., 2016). For an uncoded post, its sentiment score was then the average cosine similarities between itself and all other coded posts. Last, I standardized the sentiment scores to have mean 0 and variance 1. If sentiment scores are positive, it suggests positive attitudes toward democracy compared with the average. If sentiment scores are negative, it suggests negative attitudes toward democracy compared with the average.

Alternative measures based on dictionaries   To increase the validity of my measure, I measure both outcomes in a second way using a dictionary-based method. Simply put it, I use the method proposed in the state-of-art computer science literature (Hamilton et al., 2016) to build domain-specific lexicon dictionaries containing words about politics or pro-democracy attitudes. Then I counted the appearances of lexicons in each post and transformed the count to a dummy variable (0/1) as a measure of whether the post is discussing politics (the

_____

Weibo or those that are not: discussing politics is rare on Weibo. If I use an equal number of positive and negative examples, the classifier will overestimate the proportion of positive examples. I removed posts that contain politics-related words in a customized dictionary, from the 999000 posts, as will be explained later. In this way, these random posts should have a low probability of discussing politics.

same for the pro-democracy measure). The detailed description of dictionary construction is provided in Appendix C.1.

4.4 Identification of Causal Effect

Formally, let $y_i(0)$ be individual i's outcome if she is not treated, and $y_i(1)$ her outcome if she is treated. Let D_i be the treatment status (1 if treated and 0 otherwise). The individual-level treatment effect is thus $y_i(1) - y_i(0)$. The fundamental problem of causal inference is that exactly one of these outcomes is measured at any given time. Formally, define y_i as the observed outcome for individual i. We have:

$$
y_i = \begin{cases} y_i(1), & \text{if } D_i = 1 \\ y_i(0), & \text{if } D_i = 0 \end{cases}
$$

Methodologists have proposed numerous ways to estimate $y_i(1) - y_i(0)$ in its aggregate form (for a review, see Imbens and Wooldridge (2009)). In this study, I used the difference-in-differences design. The key assumption in the difference-in-difference design is called the "parallel trend" assumption: if treated users were not treated, they would change their outcome at the same trend as the control users (Angrist and Pischke, 2008). Therefore, the changes of control users can be used to approximate the counterfactual distribution of treated users $(y_i(0))$, and the comparison between the changes of the treatment group and the control group will eliminate biases caused by unobserved covariates.

Average Treatment Effects For both outcomes, I measure the Average Treatment Effect on the Treated (ATT) for different channels of exposure and post-treatment time period:

$$ATT_{jt} = E[y_i(1) - y_i(0)|D_i = 1] \tag{1}$$

$$= \left(\sum_{i \in T_j} \frac{y_{ia}^t}{|T_j|} - \sum_{i \in T_j} \frac{y_{ib}}{|T_j|}\right) - \left(\sum_{i \in C_j} \frac{y_{ia}^t}{|C_j|} - \sum_{i \in C_j} \frac{y_{ib}}{|C_j|}\right) \tag{2}$$

Where y_{ia}^t is the post-treatment outcome during timeframe t of person i; y_{ib} is the pre-treatment outcome within 180 days. T, C are indexes for treatment and control groups respectively. j is an index for physical or mediated exposure and their corresponding control groups. ATT_{jt} can vary by the channel of exposure and time after protest, thus revealing the heterogeneities in treatment effects along space and time.

Quantile Treatment Effect   Based on Hypotheses 3 and 4, treatment effects may be concentrated on subgroups with different frequency of political discussions and attitudes toward democracy. To understand the heterogeneity of treatment effect by the distribution of pre-treatment outcomes, I utilized *Quantile Treatment Effect on the Treated (QTT)*.

Let $F_{y_i(1)}(y)$ and $F_{y_i(0)}(y)$ be the distributions of $y_i(1)$ and $y_i(0)$, respectively. Then $F_{y_i(1)}^{-1}(\tau)$ and $F_{y_i(0)}^{-1}(\tau)$ are called the *quantile* function. An example when $\tau = 0.5$ is the median of $y_i(1)$ and $y_i(0)$, respectively. $QTT(\tau)$ is defined as the difference between the potential outcome distribution of the treatment group and the potential outcome distribution of the control group at the τth quantile (Athey and Imbens, 2006). Formally:

$$QTT(\tau) = F_{y_i(1)|D_i=1}^{-1}(\tau) - F_{y_i(0)|D_i=0}^{-1}(\tau) \tag{3}$$

In plain language, for instance, the QTT at the 0.5 quantiles is the difference between the median change of the treatment group and the median change of the control group. I used the estimator proposed by Callaway and Li (2015), which allows estimating the potential outcome distributions with a difference-in-differences design. The authors also assume "parallel trend" between treatment and control groups. The detail of the estimator can be

22

found in Callaway and Li (2015); I use the authors' R package to implement the estimator (Callaway, 2018).

5 RESULTS

5.1 Descriptive analysis

Table 1 shows the sizes of the treatment and control groups and other basic demographic covariates. The covariates are obtained from users' self-reported profile pages on Weibo.[18] To put these comparisons in context, I also compare the study population with a set of random users ($n = 10000$) that have at least 10 Weibo posts between 2012 and 2014.

Table 1 shows that the study population was 64% female, which either suggests that female users are more likely to check-in on Weibo or they are more likely to travel.[19] Both the treatment and the control group were better educated than random Weibo users. Most people were less than 25 years old, suggesting that they were born during or before the Tiananmen Square Protest in 1989 and are unlikely to have any memory of the last major political protest in China. In the Appendix D.1, I show that all provinces of China contributed to the study population, while wealthier, coastal regions were overrepresented.[20] While there is no good way to measure the socioeconomic status of Weibo users in our study, Chinese people who travel to Hong Kong tend to be wealthier. Therefore, our result should not be understood as applicable to the entire Chinese population. However, the impact on this set of young, highly educated social media users is important since they could become the future middle class, which have been known critical in fostering civil society (Lipset, 1994).

Users who went to Hong Kong during protests were not politically active before the events. For the treatment group, the distribution of the pre-protest frequency of political

---

[18]Some users did not report their gender, education or age. The reporting rate, however, is similar between the treatment and control groups.

[19]The World Tourism Organization's report suggests that Chinese females are more likely to travel aboard than males, with a ratio of 53:47. https://www.e-unwto.org/doi/book/10.18111/9789284414307

[20]This finding is not surprising since wealthier regions of China are also over-represented on Weibo.

discussions was skewed toward zero. Similarly, the treatment group was predominantly politically moderate instead of pro-democracy in the first place (Figure 2). The experimental design did *not* select a group of active political discussants, who might have better ability to circumvent the Chinese media control and Internet censorship to obtain information about where the protests might occur, which would contaminate the treatment effect (Chen and Yang, 2018).

[Table 1 about here.]

[Figure 2 about here.]

Pre-treatment balance In natural experiments, the control and treatment groups should be similar in their pre-treatment outcomes and other characteristics. The pre-treatment balance can be observed from Table 1 and Figure 2. Table 1 shows that the treatment and control groups were alike regarding their mean number of pre-treatment posts and their demographic characteristics. Figure 2 shows that the treatment and the control groups were similar in their distributions of the frequency of pre-treatment political discussions, and their sentiment toward democracy. To add more confidence, both the treatment and the control groups were similar to the random Weibo users regarding their political discussion patterns. It suggests that although users in our study population were more likely to be young, urban, and highly educated than random Weibo users, their frequency of political discussions and attitudes toward democracy was much the same.

5.2 Frequency of social and political discussions

Physical exposure vs. mediated exposure Figure 3 shows the treatment effects across time by the type of exposure. The effect was consistently higher among Chinese tourists who were closer to the protests as they were happening. The effect of physical exposure was

24

statistically significant for most timeframes. On the other hand, the treatment effect for users exposed through mediated channels could not be statistically distinguished from zero except for the first day after the protests. The differences between physical and mediate exposure are statistically significant for half of the timeframes. Therefore, confirming Hypothesis 1, physical presence at protests induced a stronger increase in users's political discussions than that of mediated exposure. In fact, Figure 3 implies that *just being at a place of protest is not enough*; it requires a strong condition— physical presence—to generate statistically significant increases in the frequency of political discussions.

Temporal duration of the effect   Chinese visitors who were at the scene of protests doubled their frequency of political discussions on the first day of the protests. The effect on the mediated users was also statistically significant but smaller in scale. After seven days, all of the Chinese tourists who were in Hong Kong during protests must have returned to China due to the visa requirement, as explained in Section 4.2. Did the treatment effect endure after they returned to China? For witnesses who were at the scene of the protests, their increases remained at a high level (40 - 50% increase) after they returned to China. Confirming Hypothesis 2, the increase lasted for about four months but did not sustain longer. The duration of the effect is a mid-range, between an ultra-short and a lifetime consequence as having been suggested in the literature.

In contrast, for the treatment group who were exposed through mediated channels, the effect decreased to be not statistically significant after they returned to China. Therefore, the duration of effect also depended heavily on the channel of exposure: physical presence created a sustained effect, while mediated exposure failed to elicit changes in the long run.

[Figure 3 about here.]

Quantile treatment effect on political discussions Figure 4 shows the estimates of QTT on the frequency of political discussions.[21] For both channels, the increases were higher in the highest quantiles, namely users who already discussed political issues the most often. This observation confirms Hypothesis 3 that the exposure to political protests provides resources and opportunity to politically active users to further express concern over political issues. The difference between two channels was statistically significant in lower and middle quantiles but was not significant in the highest quantile, which may suggest that people who discussed politics the most often, regardless of whether they checked-in nearby the protests, will actively seek information about the protests while they were in Hong Kong. One thing to note is that the lower and middle quantiles of the mediated-treatment group did not increase their frequency of political discussions; the observed increase in ATT is mostly driven by a small fraction of the politically active users.

[Figure 4 about here.]

5.3 Changes in political attitudes

Physical exposure vs. mediated exposure Figure 5 shows the ATT on the attitudes toward political democracy. Physical exposure did not induce witnesses to become more positive toward democracy; the effect could not be statistically distinguished from zero. On the other hand, mediated exposure decreases the support for democracy. It suggests that if Chinese tourists went to Hong Kong but did not see the protest personally, they would turn against democracy, instead of supporting it.

Temporal duration of the effect Figure 5 also shows that the effects of physical exposure on pro-democracy attitudes are not statistically significant for almost all timeframes (expect

---

[21]This estimate is pooled over 180 days after protests, due to data sparsity consideration, since discussing politics is not an everyday behavior for my study population. QTT would be higher if it were estimated within the first day after the protest.

the first month after the protests). On the other hand, treated users who were exposed through mediated channels showed an enduring decrease in their support for democracy, even after they returned to China.

[Figure 5 about here.]

Quantile treatment effect on political attitudes Last, I examine how pre-treatment political predispositions affect attitudinal changes. Figure 6 shows the quantile treatment effects with respect to the support for democracy. If users were physically present, the protests had a polarizing effect. Users who were already anti-democracy further decreased their pro-democracy attitudes because they were exposed to information that they would otherwise avoid. In contrast, users who were already pro-democracy became more supportive after being exposed to the protests. Confirming Hypothesis 4, the polarizing effect under mediated exposure is consistently smaller than that from physical exposure. Compared with mediated exposure, physical exposure produces greater support for democracy among the pro-democracy and greater opposition for democracy among the anti-democracy.

[Figure 6 about here.]

What is more interesting is the change in pro-democracy attitudes of the politically moderate. Under physical exposure, there was no effect for the moderates. Under mediated exposure, the moderates decreased their support for democracy.

Why would stay in a city with protests but not physically present lead to less support for democracy among moderates? One possibility is the cultural conflicts between Chinese tourists and Hong Kong local residents make visitors hostile toward city residents and hence their appeals. If so, we would expect that the protests against Chinese tourists will exhibit a higher level of decrease compared to other protests. Among the eight protests, a demonstration on February 19, 2012, in particular, protested the effects of Chinese tourists on Hong Kong. I analyze the QTT for this specific protest and find that it was not the case

(Appendix D.5). Therefore, the cultural clashes are not the primary reason why we observe a decrease among the mediated-treatment group.

Another explanation, as articulated in forming Hypothesis 4, is that the moderate visitors chose to avoid the protests, not because of their preexisting ideology, but because the protests interrupted their trips. Because of the interruption, they did not have a chance to hear the appeals of the protests. Instead, the interruption might induce negative impressions of the protests, which leads to a negative evaluation of the associated democratic appeals. Theorists such as Eliasoph (1998) have presented similar arguments about democracies but in everyday life settings instead of an atypical experience of seeing protests in another region. Further research needs to test the credibility of this explanation directly.

5.4   Robustness check

To ensure that the results are not confined to my choice of Hong Kong, I performed four groups of robustness checks. First, I used the same data collection methods to collect Chinese visitors who were exposed to one of the five recent protests in Taipei and control groups who just left Taipei before the protests (the details of the protests in Taipei is discussed in Appendix A.2). Then I replicated the analysis for this set of users. The results are qualitatively similar (Appendix D.2). Second, I rerun each analysis using alternative dictionary-based outcome measures. The results were also similar (Appendix D.3). Third, the study population should remain unchanged in their frequency discussion regarding topics that were irrelevant to politics. I found that the treatment groups (in either channel) indeed did not discuss more sports, an irrelevant topic (Appendix D.4). Last, there might be other unknown factors that make Weibo users who did not visit democracies have an increase in their frequency of online political discussions. To rule out this possibility, I examined ordinary Weibo users who traveled to other big cities of China— Beijing or Shanghai—on the same protest dates (Appendix D.4). I find that these users did not discuss more politics after the protests. The four robustness checks add further confidence to the main results.

6 CONCLUSION

This study causally identifies the effect of eight political protests in Hong Kong on the exposed Chinese tourists. Using the protests as natural experiments, I collected Chinese visitors who were in Hong Kong during the protests and compared them with other Chinese visitors who also went to Hong Kong but left just before the protests. Using difference-in-differences models, I estimated both the average and quantile treatment effects. I found that *physical presence* at the scene of the protests let the treated users discuss political issues more often, especially among users who already discussed politics frequently. It also reinforced preexisting political attitudes. On the other hand, mediated exposure did not produce any sizable impact on the frequency of political discussions, and it decreased the supportive attitudes toward democracy, especially among the politically moderate.

For the literature on protest diffusion in authoritarian regimes, this research does not support their widely assumed but rarely tested link: the exposure to political protests pursuing democracy increases mobilization potential in a nearby regime. On the other hand, this paper echoes recent research that shows that exposure to democracy through media and education may have an unexpected outcome to promote trust for the authoritarian governments, in countries such as East Germany, Croatia, and China (Kern and Hainmueller, 2009; Kern, 2011; Huang, 2015; Tai, 2015; DellaVigna et al., 2014; Bursztyn and Cantoni, 2015). Therefore, this paper calls for a reexamination on the basic assumptions of the protest diffusion literature and calls for a deeper understanding of authoritarian resilience (Nathan, 2003).

For social movement scholars, this paper extends the burgeoning studies in consequence of social movement literature by revealing the heterogeneous effects of protests alongside the channel of exposure, time, and pre-protest political discussions and attitudes. First, this study shows that mediated exposure to the protests may lead to an adversarial outcome by leading to lower support for political democracy, the appeals of all the protests used in this

study. This finding contrasts most existing studies that identify that exposure to protest through local media increased approval of the tactic and supportive attitudes toward the protests (Wallace et al., 2014; Andrews et al., 2016; Branton et al., 2015). We need to think about the scope condition of the direction in which protests can change public opinions. This study also shows that physical presence at the scene of protests can lead to different effects that mediated exposure because it demands attention from the witnesses. This difference has not noted widely among scholars (except (Schwartz, 2016)). Second, this study shows that the effect of protests on political discussions is longer than a short-term reactionary response but is shorter than the life-long consequence produced by the large social movements such as the Civil Rights Movement of the Cultural Revolution (McAdam, 1999; Yang, 2000). The duration of the effect of protest might impact how future researchers measure their outcome. For instance, if scholars rely on annual public opinion surveys, they may be unable to find any effect of protest simply because the time intervals between survey ways is too long. Last, following the plea of Walder (2009), this paper reveals that the preexisting level of political engagement and predisposition affect how individuals respond to the messages and appeals of political protests. I find that those who were already engaged in politics also had a higher increase in their frequency of discussions after protests. I also find that protests polarize bystanders based on their preexisting political attitudes. The effect, however, also differs by channels of exposures. We hope that other studies can invest in examining the generalizability of these findings in other protests.

There are limitations to the results. First, coverage and participation biases remained an issue for social media data, even when social media data is necessary for identifying bystanders. The coverage bias means that the specific platforms used in this paper—Weibo—does not capture all Chinese people who were exposed to the protests in Hong Kong. The participation biases mean that for people we captured in the study population, not all of them will discuss politics online (Diaz et al., 2016b). Thus there could be outcomes that we cannot observe. Second, this paper has not looked at how the two outcomes may influence each

other, as have been suggested in the literature (Mutz, 2002). Furthermore, the effect may spread across social networks on social media (Andrews and Biggs, 2006; González-Bailón et al., 2011). Last, there may be an effect beyond online political discussions, especially offline political participation (Enikolopov et al., 2017). These limitations and unanswered questions invite future research on the causal impact of political protests on political behaviors and attitudes.

Beyond these empirical findings and future directions, this research has its methodological implications. It illustrates how to use spatiotemporal information from social media data to find witnesses after protest ends, to collect their pre-event information, and to help to utilize natural experiments to estimate the impact of events. Conventional survey and interview techniques present difficulties obtaining data and formulating similar designs. This paper indicates that in social science research, social media data can go beyond its current use as a convenience sample compared to the "gold-standard" survey; its ability to construct control groups and obtain pre-event information is helpful in understanding the impact of unexpected events that cannot be answered by conventional data collection approaches.

REFERENCES

Almond, Gabriel Abraham and Sidney Verba. 1963. *The Civic Culture: Political Attitudes and Democracy in Five Nations.* Princeton University Press.

Andrews, Kenneth T., Kraig Beyerlein, and Tuneka Tucker Farnum. 2016. "The Legitimacy of Protest: Explaining White Southerners' Attitudes Toward the Civil Rights Movement." *Social Forces* 94:1021–1044.

Andrews, Kenneth T. and Michael Biggs. 2006. "The Dynamics of Protest Diffusion: Movement Organizations, Social Networks, and News Media in the 1960 Sit-Ins." *American Sociological Review* 71:752–777.

Andrews, Kenneth T. and Neal Caren. 2010. "Making the News Movement Organizations, Media Attention, and the Public Agenda." *American Sociological Review* 75:841–866.

Angrist, Joshua D. and JÃ¶rn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton university press.

Athey, Susan and Guido W. Imbens. 2006. "Identification and Inference in Nonlinear Difference-in-Differences Models." *Econometrica* 74:431–497.

Bail, Christopher A. 2015. "Lost in a Random Forest: Using Big Data to Study Rare Events." *Big Data & Society* 2:1–3.

Baldassarri, Delia and Peter Bearman. 2007. "Dynamics of Political Polarization." *American Sociological Review* 72:784–811.

Benford, Robert D. and Scott A. Hunt. 1992. "Dramaturgy and Social Movements: The Social Construction and Communication of Power." *Sociological Inquiry* 62:36–55.

Bennett, Stephen E., Richard S. Flickinger, and Staci L. Rhine. 2000. "Political Talk Over Here, Over There, Over Time." *British Journal of Political Science* 30:99–119.

Biggs, Michael. 2005. "Strikes as Forest Fires: Chicago and Paris in the Late Nineteenth Century." *American Journal of Sociology* 110:1684–1714.

Bojanowski, Piotr, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. "Enriching Word Vectors with Subword Information." *Transactions of the Association for Computational Linguistics* 5:135–146.

Branton, Regina, Valerie Martinez-Ebers, Tony E Carey, and Tetsuya Matsubayashi. 2015. "Social Protest and Policy Attitudes: The Case of the 2006 Immigrant Rallies." *American Journal of Political Science* 59:390–402.

Budak, Ceren and Duncan J Watts. 2015. "Dissecting the Spirit of Gezi: Influence vs. Selection in the Occupy Gezi Movement." *Sociological Science* pp. 370–397.

Bursztyn, Leonardo and Davide Cantoni. 2015. "A Tear in the Iron Curtain: The Impact of Western Television on Consumption Behavior." *The Review of Economics and Statistics* 98:25–41.

Cai, Yongshun. 2016. *The Occupy Movement in Hong Kong: Sustaining Decentralized Protest.* Taylor & Francis.

Callaway, Brantly. 2018. *qte: Quantile Treatment Effects.* R package version 1.2.2.

Callaway, Brantly and Tong Li. 2015. "Quantile Treatment Effects in Difference in Differences Models with Panel Data." .

Carpini, Michael X. Delli, Fay Lomax Cook, and Lawrence R. Jacobs. 2004. "Public Deliberation, Discursive Participation, and Citizen Engagement: A Review of the Empirical Literature." *Annual Review of Political Science* 7:315–344.

Chan, Johannes. 2014. "Hong Kong's Umbrella Movement." *The Round Table* 103:571–580.

Chen, Yuyu and David Y Yang. 2018. "The Impact of Media Censorship: Evidence from a Field Experiment in China."

Cheng, Edmund W. and Wai-Yin Chan. 2017. "Explaining Spontaneous Occupation: Antecedents, Contingencies and Spaces in the Umbrella Movement." *Social Movement Studies* 16:222–239.

Cheung, Alvin YH. 2014. "Burying the Joint Declaration: Beijing's International Law Reaction to the Umbrella Movement." *Available at SSRN 2540710* .

Collins, Randall. 1993a. "Emotional Energy as the Common Denominator of Rational Action." *Rationality and Society* 5:203–230.

Collins, Randall. 1993b. "Emotional Energy as the Common Denominator of Rational Action." *Rationality and Society* 5:203–230.

Collins, Randall. 2001. "Social Movements and the Focus of Emotional Attention." In *Passionate Politics: Emotions and Social Movements*, edited by Jeff Goodwin, James M. Jasper, and Francesca Polletta, pp. 27–44.

Conover, Pamela Johnston, Donald D. Searing, and Ivor M. Crewe. 2002. "The Deliberative Potential of Political Discussion." *British Journal of Political Science* 32:21–62.

Converse, Philip E. 2006. "The Nature of Belief Systems in Mass Publics (1964)." *Critical Review* 18:1–74.

Dalton, Russell J. 1994. "Communists and Democrats: Democratic Attitudes in the Two Germanies." *British Journal of Political Science* 24:469–493.

Davis, Michael C. 2015. "Hong Kong's Umbrella Movement and Beijing's Failure to Honor the Basic Law." *Available at SSRN 2544386* .

DellaVigna, Stefano, Ruben Enikolopov, Vera Mironova, Maria Petrova, and Ekaterina Zhuravskaya. 2014. "Cross-Border Media and Nationalism: Evidence from Serbian Radio in Croatia." *American Economic Journal: Applied Economics* 6:103–132.

Diaz, Fernando, Michael Gamon, Jake M. Hofman, Emre Kıcıman, and David Rothschild. 2016a. "Online and Social Media Data As an Imperfect Continuous Panel Survey." *PLOS ONE* 11:e0145406.

Diaz, Fernando, Michael Gamon, Jake M Hofman, Emre Kıcıman, and David Rothschild. 2016b. "Online and social media data as an imperfect continuous panel survey." *PloS one* 11:e0145406.

Dunning, Thad. 2012. *Natural Experiments in the Social Sciences: A Design-Based Approach.* Cambridge University Press.

Egorov, Georgy and Konstantin Sonin. 2011. "Dictators and Their Viziers: Endogenizing the

Loyalty-Competence Trade-Off." *Journal of the European Economic Association* 9:903–930.

Eliasoph, Nina. 1998. *Avoiding Politics: How Americans Produce Apathy in Everyday Life.* Cambridge University Press.

Enikolopov, Ruben, Alexey Makarin, and Maria Petrova. 2017. "Social Media and Protest Participation: Evidence from Russia." SSRN Scholarly Paper ID 2696236, Social Science Research Network, Rochester, NY.

Eveland, William P., Andrew F. Hayes, Dhavan V. Shah, and Nojin Kwak. 2005. "Understanding the Relationship Between Communication and Political Knowledge: A Model Comparison Approach Using Panel Data." *Political Communication* 22:423–446.

Finkel, Steven E. and Amy Erica Smith. 2011. "Civic Education, Political Discussion, and the Social Transmission of Democratic Knowledge and Values in a New Democracy: Kenya 2002." *American Journal of Political Science* 55:417–435.

Fong, Brian C. H. 2014. "Executive-Legislative Disconnection in Post-Colonial Hong Kong: The Dysfunction of the HKSAR's Executive-Dominant System, 1997-2012." *China Perspectives* pp. 5–14.

Fong, Brian C. H. 2015. *Hong Kong's Governance Under Chinese Sovereignty : The Failure of the State-Business Alliance After 1997.* New York, NY: Routledge.

Gamson, William A. 1992. *Talking Politics.* Cambridge University Press.

Gamson, William A. 2004. "Bystanders, Public opinion, and the Media." In *The Blackwell Companion to Social Movements*, edited by David A. Snow, Sarah A. Soule, and Hanspeter Kriesi, pp. 242–261. Blackwell Oxford, UK.

Garg, Nikhil, Londa Schiebinger, Dan Jurafsky, and James Zou. 2018. "Word Embeddings Quantify 100 Years of Gender and Ethnic Stereotypes." *Proceedings of the National Academy of Sciences* 115:E3635–E3644.

Gastil, John. 2000. "Is Face-to-Face Citizen Deliberation a Luxury or a Necessity?" *Political Communication* 17:357–361.

Gibson, Rachel and Marta Cantijoch. 2013. "Conceptualizing and Measuring Participation in the Age of the Internet: Is Online Political Engagement Really Different to Offline?" *The Journal of Politics* 75:701–716.

González-Bailón, Sandra, Javier Borge-Holthoefer, Alejandro Rivero, and Yamir Moreno. 2011. "The Dynamics of Protest Recruitment through an Online Network." *Scientific Reports* 1:197.

Grimmer, Justin and Brandon M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21:267–297.

Hale, Henry E. 2013. "Regime Change Cascades: What We Have Learned from the 1848 Revolutions to the 2011 Arab Uprisings." *Annual Review of Political Science* 16:331–353.

Hamilton, William L., Kevin Clark, Jure Leskovec, and Dan Jurafsky. 2016. "Inducing Domain-Specific Sentiment Lexicons from Unlabeled Corpora." In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 595–605. Association for Computational Linguistics.

Huang, Haifeng. 2015. "International Knowledge and Domestic Evaluations in a Changing Society: The Case of China." *American Political Science Review* 109:613–634.

Imbens, Guido W. and Jeffrey M. Wooldridge. 2009. "Recent Developments in the Econometrics of Program Evaluation." *Journal of Economic Literature* 47:5–86.

Inglehart, Ronald. 2003. "How Solid Is Mass Support for Democracy—And How Can We Measure It?" *PS: Political Science & Politics* 36:51–57.

Johnston, Hank. 2005. "Talking the Walk: Speech Acts and Resistance in Authoritarian Regimes." In *Repression and Mobilization*, edited by Christian Davenport, Hank Johnston, and Carol McClurg Mueller, pp. 108–137. University of Minnesota Press.

Joulin, Armand, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2017. "Bag of Tricks for Efficient Text Classification." In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers.*

Kenter, Tom, Alexey Borisov, and Maarten de Rijke. 2016. "Siamese CBOW: Optimizing Word Embeddings for Sentence Representations." In *Proceedings of the The 54th Annual Meeting of the Association for Computational Linguistics (ACL 2016)*, pp. 941–951. Association for Computational Linguistics.

Kern, Holger Lutz. 2011. "Foreign Media and Protest Diffusion in Authoritarian Regimes The Case of the 1989 East German Revolution." *Comparative Political Studies* 44:1179–1205.

Kern, Holger Lutz and Jens Hainmueller. 2009. "Opium for the Masses: How Foreign Media Can Stabilize Authoritarian Regimes." *Political Analysis* pp. 377–399.

Klandermans, Bert and Dirk Oegema. 1987. "Potentials, Networks, Motivations, and Barriers: Steps Towards Participation in Social Movements." *American Sociological Review* 52:519–531.

Koopmans, Ruud and Susan Olzak. 2004. "Discursive Opportunities and the Evolution of Right-Wing Violence in Germany." *American Journal of Sociology* 110:198–230.

Kozlowski, Austin C., Matt Taddy, and James A. Evans. 2018. "The Geometry of Culture: Analyzing Meaning through Word Embeddings." *arXiv:1803.09288 [cs]* .

Kuran, Timur. 1991. "Now out of Never: The Element of Surprise in the East European Revolution of 1989." *World Politics* 44:7–48.

Kuran, Timur. 1997. *Private Truths, Public Lies: The Social Consequences of Preference Falsification.* Harvard University Press.

Kwak, Nojin, Ann E. Williams, Xiaoru Wang, and Hoon Lee. 2005. "Talking Politics and Engaging Politics: An Examination of the Interactive Relationships Between Structural Features of Political Talk and Discussion Engagement." *Communication Research* 32:87–111.

Kwiatkowski, Piotr. 1992. "Opinion Research and the Fall of Communism: Poland 1981–1990." *International Journal of Public Opinion Research* 4:358–374.

Legewie, Joscha. 2013. "Terrorist Events and Attitudes toward Immigrants: A Natural Experiment." *American Journal of Sociology* 118:1199–1245.

Li, Shen, Zhe Zhao, Renfen Hu, Wensi Li, Tao Liu, and Xiaoyong Du. 2018. "Analogical Reasoning on Chinese Morphological and Semantic Relations." In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 138–143. Association for Computational Linguistics.

Lipset, Seymour Martin. 1994. "The Social Requisites of Democracy Revisited: 1993 Presidential Address." *American Sociological Review* 59:1–22.

Lynch, Marc. 2014. *The Arab Uprisings Explained: New Contentious Politics in the Middle East.* Columbia University Press.

Mattes, Robert and Michael Bratton. 2007. "Learning about Democracy in Africa: Awareness, Performance, and Experience." *American Journal of Political Science* 51:192–217.

Mazumder, Soumyajit. 2018. "The Persistent Effect of U.S. Civil Rights Protests on Political Attitudes." *American Journal of Political Science* 0.

McAdam, Doug. 1989. "The Biographical Consequences of Activism." *American Sociological Review* 54:744–760.

McAdam, Doug. 1999. "The Biographical Impact of Activism." In *How Social Movements Matter*, edited by Marco Giugni, Doug McAdam, and Charles Tilly. University of Minnesota Press.

McAdam, Doug and Hilary Boudet. 2012. *Putting Social Movements in Their Place: Explaining Opposition to Energy Projects in the United States, 2000–2005*. Cambridge University Press.

McCarthy, John D., Clark McPhail, and Jackie Smith. 1996. "Images of Protest: Dimensions of Selection Bias in Media Coverage of Washington Demonstrations, 1982 and 1991." *American Sociological Review* 61:478–499.

McGuire, William J., Gardner Lindzey, and Elliott Aronson. 1969. "Attitudes and Attitude Change." In *Handbook of Social Psychology*, volume 3. MA: Addison-Wesley.

McKenzie, David. 2012. "Beyond Baseline and Follow-Up: The Case for More T in Experiments." *Journal of Development Economics* 99:210–221.

Mcmillan, John and Pablo Zoido. 2004. "How to Subvert Democracy: Montesinos in Peru." *Journal of Economic Perspectives* 18:69–92.

Meyer, David S. and Suzanne Staggenborg. 1996. "Movements, Countermovements, and the Structure of Political Opportunity." *American Journal of Sociology* 101:1628–1660.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. "Distributed Representations of Words and Phrases and Their Compositionality." In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'13, pp. 3111–3119, USA. Curran Associates Inc.

Mutz, Diana C. 2002. "The Consequences of Cross-Cutting Networks for Political Partici-
pation." *American Journal of Political Science* 46:838–855.

Nathan, Andrew J. 2003. "Authoritarian Resilience." *Journal of Democracy* 14:6–17.

Ortmann, Stephan. 2016. "The Lack of Sovereignty, the Umbrella Movement, and
Democratisation in Hong Kong." *Asia Pacific Law Review; Hong Kong* 24:108–122.

Pennington, Jeffrey, Richard Socher, and Christopher Manning. 2014. "Glove: Global Vec-
tors for Word Representation." In *Proceedings of the 2014 Conference on Empirical Meth-
ods in Natural Language Processing (EMNLP)*, pp. 1532–1543. Association for Computa-
tional Linguistics.

Perrin, Andrew J. 2006. *Citizen Speak: The Democratic Imagination in American Life.*
Chicago: University of Chicago Press.

Polletta, Francesca. 1999. ""Free Spaces" in Collective Action." *Theory and Society* 28:1–38.

Price, Vincent and John Zaller. 1993. "Who Gets the News? Alternative Measures of News
Reception and Their Implications for Research." *The Public Opinion Quarterly* 57:133–
164.

Prior, Markus. 2007. *Post-Broadcast Democracy: How Media Choice Increases Inequality in
Political Involvement and Polarizes Elections.* Cambridge University Press.

Qin, Bei, David Strömberg, and Yanhui Wu. 2018. "Media Bias in China." *American
Economic Review* 108:2442–2476.

Rowen, Ian. 2016. "The Geopolitics of Tourism: Mobilities, Territory, and Protest in China,
Taiwan, and Hong Kong." *Annals of the American Association of Geographers* 106:385–393.

Rueschemeyer, Dietrich, Evelyne Huber Stephens, and John D. Stephens (eds.). 1992. *Cap-
italist Development and Democracy.* Chicago: University of Chicago Press.

Salganik, Matthew. 2017. *Bit by Bit: Social Research for the Digital Age.* Princeton University Press.

Schofer, Evan and Marion Fourcade-Gourinchas. 2001. "The Structural Contexts of Civic Engagement: Voluntary Association Membership in Comparative Perspective." *American Sociological Review* 66:806–828.

Schussman, Alan and Sarah A. Soule. 2005. "Process and Protest: Accounting for Individual Protest Participation." *Social Forces* 84:1083–1108.

Schwartz, Cassilde. 2016. *The Paradox of Confrontation: Experimental Evidence on the Audience Effects of Protest.* Ph.D. thesis, University of Pittsburgh.

Sewell Jr, William H. 2001. "Space in Contentious Politics." In *Silence and Voice in the Study of Contentious Politics*, edited by Sidney Tarrow, Ronald Aminzade, Jack Goldstone, Doug McAdam, Elizabeth Perry, William H. Sewell Jr, and Charles Tilly, volume 78, pp. 51–88.

Shirky, Clay. 2011. "The Political Power of Social Media: Technology, the Public Sphere, and Political Change." *Foreign Affairs* 90:28–41.

Skocpol, Theda and Morris P. Fiorina. 2004. *Civic Engagement in American Democracy.* Brookings Institution Press.

Slater, Michael D. 2007. "Reinforcing Spirals: The Mutual Influence of Media Selectivity and Media Effects and Their Impact on Individual Behavior and Social Identity." *Communication Theory* 17:281–303.

Smith, Jackie, John D. McCarthy, Clark McPhail, and Boguslaw Augustyn. 2001. "From Protest to Agenda Building: Description Bias in Media Coverage of Protest Events in Washington, D.C." *Social Forces* 79:1397–1423.

Snow, David A. and Dana M. Moss. 2014. "Protest on the Fly Toward a Theory of Spontaneity in the Dynamics of Protest and Social Movements." *American Sociological Review* 79:1122–1143.

Stevens, Daniel, Benjamin G. Bishin, and Robert R. Barr. 2006. "Authoritarian Attitudes, Democracy, and Policy Preferences among Latin American Elites." *American Journal of Political Science* 50:606–620.

Stockmann, Daniela and Mary E. Gallagher. 2011. "Remote Control: How the Media Sustain Authoritarian Rule in China." *Comparative Political Studies* p. 0010414010394773.

Stroud, Natalie Jomini. 2008. "Media Use and Political Predispositions: Revisiting the Concept of Selective Exposure." *Political Behavior* 30:341–366.

Swidler, Ann. 1986. "Culture in Action: Symbols and Strategies." *American Sociological Review* 51:273–286.

Taber, Charles S. and Milton Lodge. 2006. "Motivated Skepticism in the Evaluation of Political Beliefs." *American Journal of Political Science* 50:755–769.

Tai, Qiuqing. 2015. "Western Media Exposure and Chinese Immigrants' Political Perceptions." *Political Communication* pp. 1–20.

Tarrow, Sidney. 2005. *The New Transnational Activism.* Cambridge University Press.

Tilly, Charles. 1993. "Social Movements as Historically Specific Clusters of Political Performances." *Berkeley Journal of Sociology* 38:1–30.

Tilly, Charles. 2000. "Spaces of Contention." *Mobilization: An International Quarterly* 5:135–159.

Tufekci, Zeynep and Christopher Wilson. 2012. "Social media and the decision to participate in political protest: Observations from Tahrir Square." *Journal of Communication* 62:363–379.

Valenzuela, Sebastián, Yonghwan Kim, and Homero Gil de Zúñiga. 2012. "Social Networks that Matter: Exploring the Role of Political Discussion for Online Political Participation." *International Journal of Public Opinion Research* 24:163–184.

Verba, Sidney, Kay Lehman Schlozman, Henry E. Brady, and Henry E. Brady. 1995. *Voice and Equality: Civic Voluntarism in American Politics*, volume 4. Cambridge Univ Press.

Walder, Andrew G. 2009. "Political Sociology and Social Movements." *Annual Review of Sociology* 35:393–412.

Wallace, Sophia J, Chris Zepeda-Millán, and Michael Jones-Correa. 2014. "Spatial and Temporal Proximity: Examining the Effects of Protests on Political Attitudes." *American Journal of Political Science* 58:433–448.

Walsh, Katherine Cramer. 2010. *Talking about Politics: Informal Groups and Social Identity in American Life.* Chicago: University of Chicago Press.

Wang, Zhengxu. 2007. "Public Support for Democracy in China." *Journal of Contemporary China* 16:561–579.

Wassler, Philipp, Markus Schuckert, Kam Hung, and James F. Petrick. 2018. "You're Welcome? Hong Kong's Attitude towards the Individual Visit Scheme." *International Journal of Tourism Research* 20:637–649.

Welsh, William (ed.). 1981. *Survey Research and Public Attitudes in Eastern Europe and the Soviet Union.* Elsevier.

Weyland, Kurt. 2009. "The Diffusion of Revolution: '1848' in Europe and Latin America." *International Organization* 63:391–423.

White, Lynn. 2016. *Democratization in Hong Kong–and China?* Lynne Rienner Publishers.

Yang, Guobin. 2000. "The Liminal Effects of Social Movements: Red Guards and the Transformation of Identity." *Sociological Forum* 15:379–406.

Zaller, John. 1992. *The Nature and Origins of Mass Opinion.* Cambridge university press.

Zhao, Dingxin. 1998. "Ecologies of Social Movements: Student Mobilization during the 1989 Prodemocracy Movement in Beijing." *American Journal of Sociology* 103:1493–1529.

Zhao, Dingxin. 2000. "State-Society Relations and the Discourses and Activities of the 1989 Beijing Student Movement." *American Journal of Sociology* 105:1592–1632.

Zhu, Xiaojin, Zoubin Ghahramani, and John Lafferty. 2003. "Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions." In *ICML'03 Proceedings of the Twentieth International Conference on International Conference on Machine Learning*, pp. 912–919.

Zúñiga, Homero Gil De, Eulàlia Puig-I-Abril, and Hernando Rojas. 2009. "Weblogs, traditional sources online and political participation: an assessment of how the internet is changing the political environment." *New Media & Society* 11:553–574.

List of Figures

Distance to Protest (meters)

Control Group:
Media exposure

Control Group:
Physical exposure

5000

Treatment Group:
Media Exposure

1500 Treatment Group:
Physical
Exposure

Check-in Dates
to Protest Date
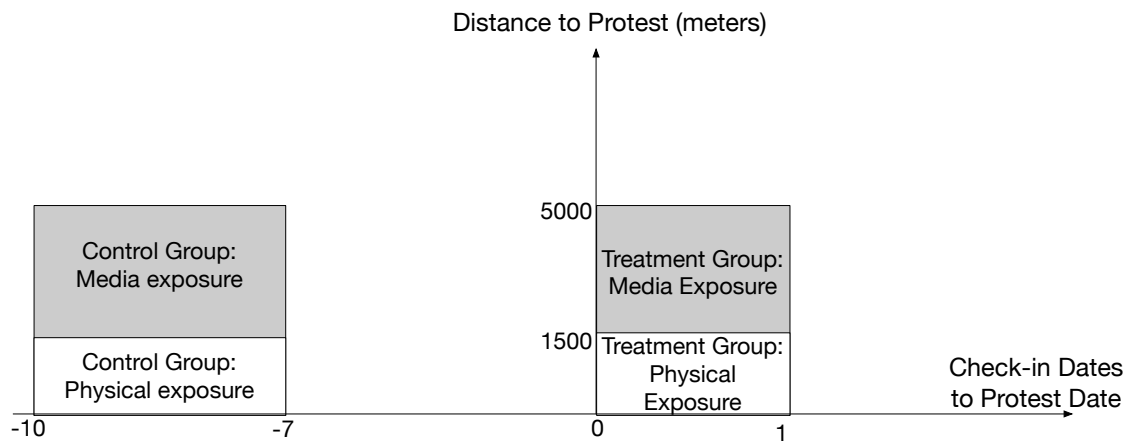
-10   -7   0   1
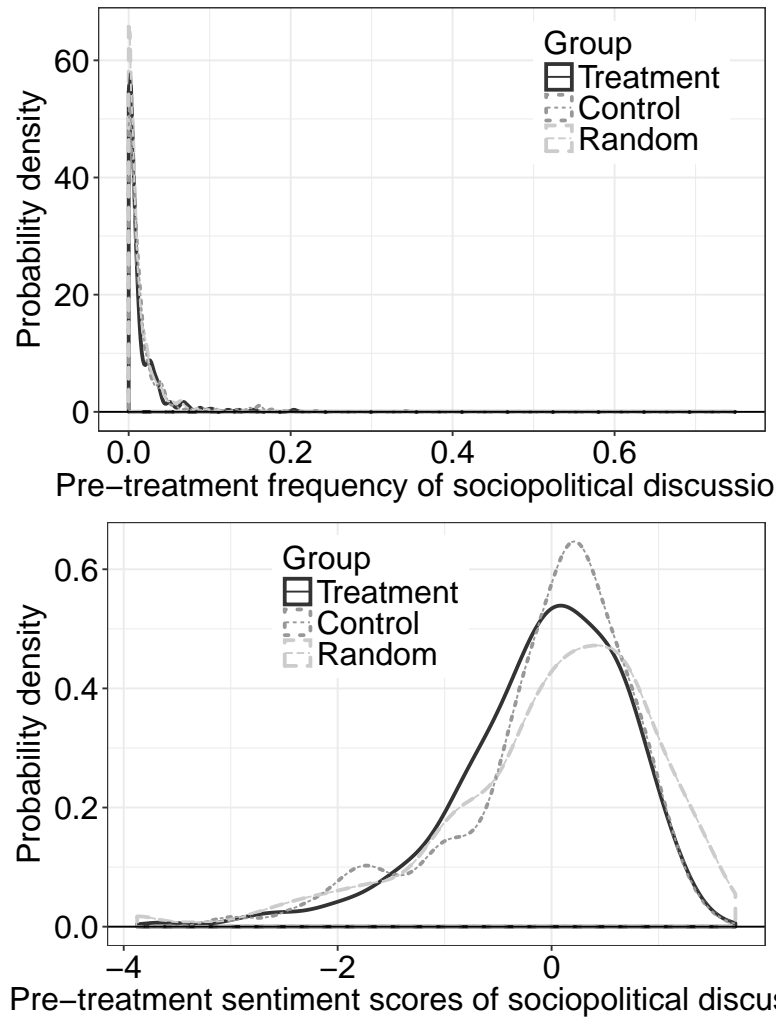
Figure 1: Illustration of the experiment design for a single protest.

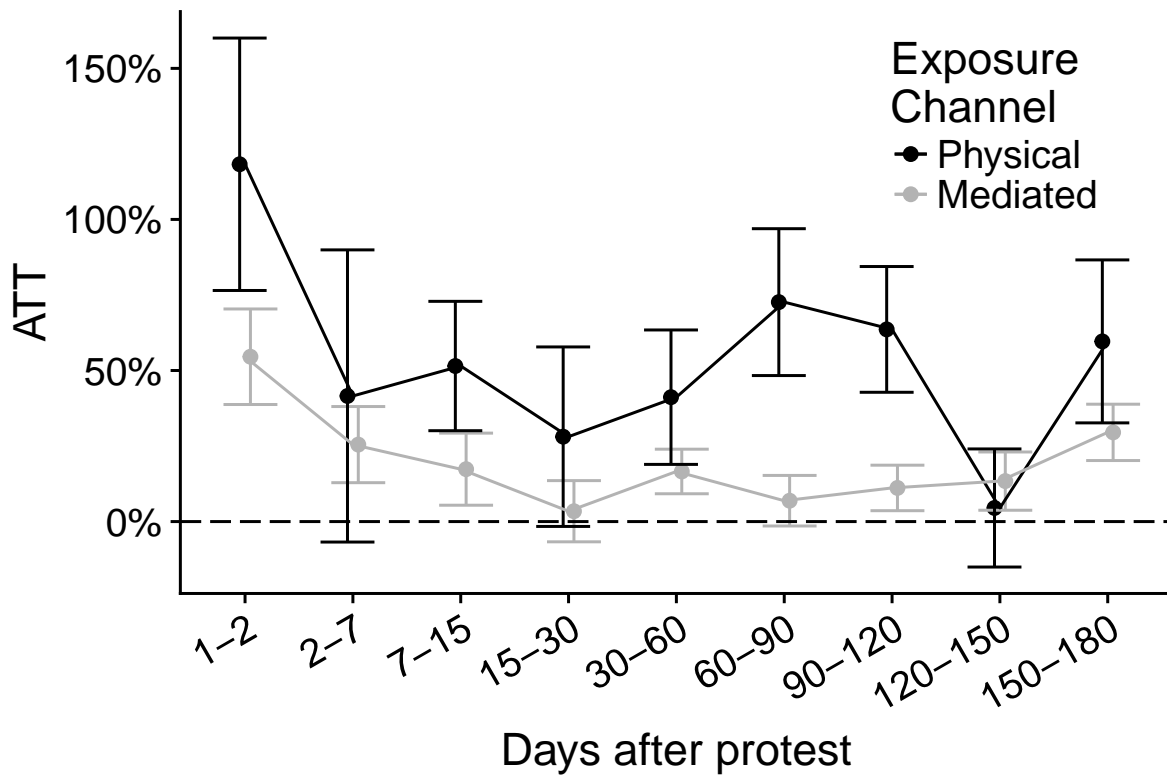Figure 2: Distribution of pre-treatment frequency and sentiment of discussion of social/political issues.

Figure 3: *ATT* on the frequency of political discussions, by exposure channel and time. The 95% confidence intervals are calculated from 10000 bootstrap draws.
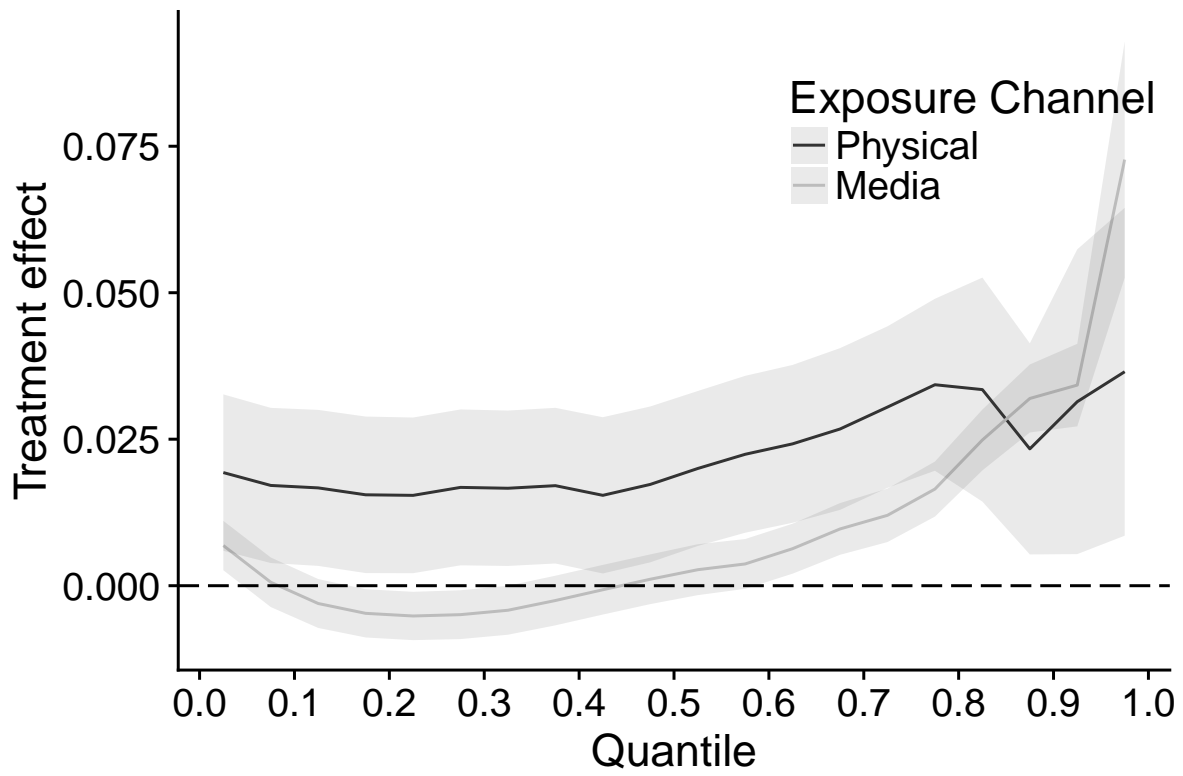
Figure 4: *QTT* on the frequency of political discussions. The 95% confidence intervals are calculated from 10000 bootstrap draws.
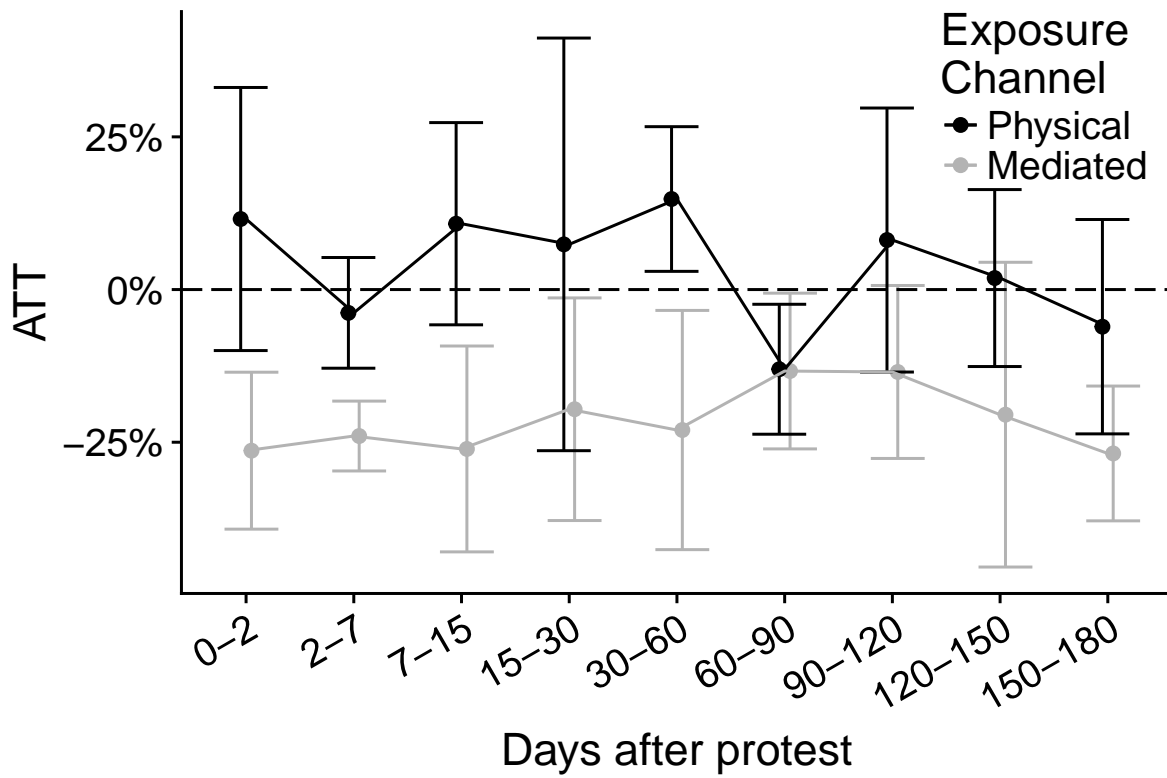
Figure 5: *ATT* for attitudes toward political democracy, by distance and time. This plot use the word2vec-based estimator. The 95% confidence intervals are calculated from 10000 bootstrap draws
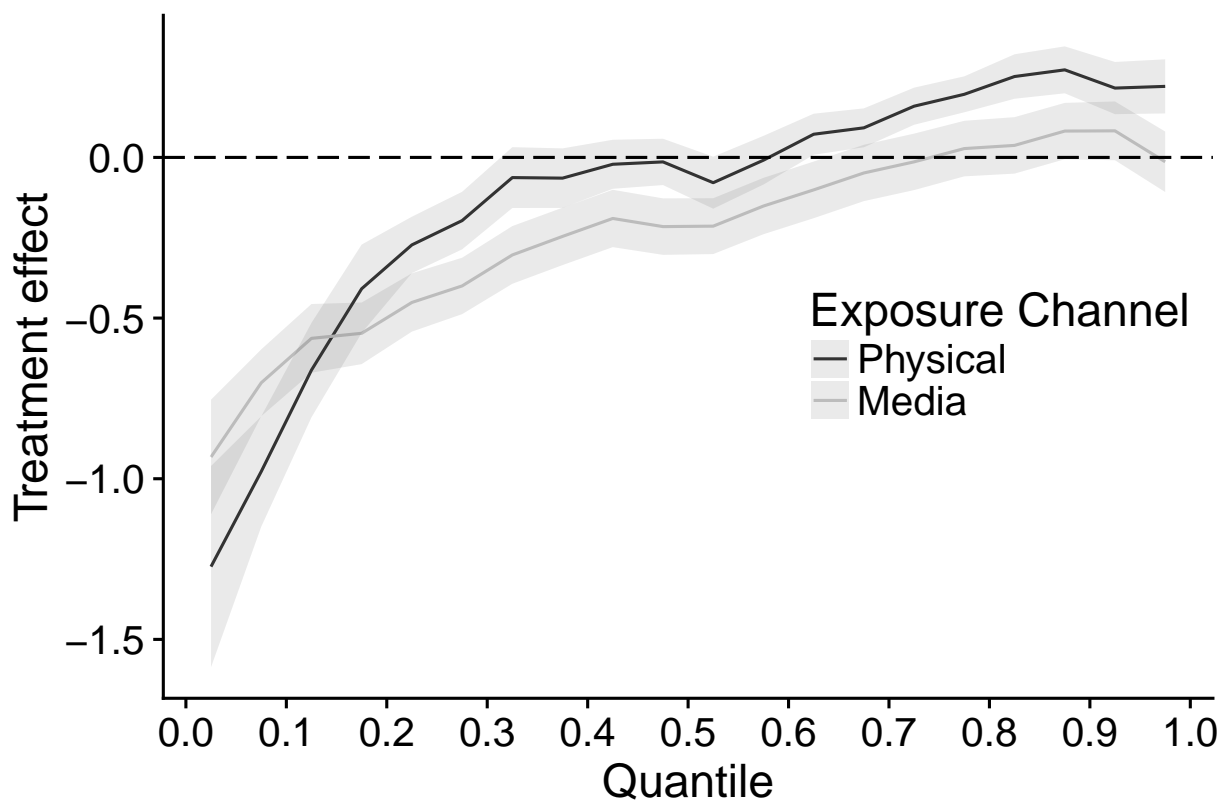
Figure 6: *QTT* on support for democracy. This plot use the word2vec-based estimator. The 95% confidence intervals are calculated from 10000 bootstrap draws.

List of Tables

Table 1: Descriptive statistics of the study population.

	Treatment		Control		Random
	Physical	Mediated	Physical	Mediated	
Sizes					
Total number of users	67	316	151	624	12000
Total number of posts [a]	70146	334675	186476	673309	10181040
Mean number of posts per user [a]	1046.9	1059.1	1234.9	1079.0	848.4
Pre-treatment Outcomes					
Frequency of political discussions [b]	1.31%	1.68%	1.26%	1.75%	1.43%
Pre-treatment covariates [c]					
Male%	38.0%	42.1%	34.7%	39.4%	50.5%
College Education%	44.7%	42.1%	42.5%	41.4%	34%
Age ≤ 25 %	89.47%	88.3%	90.21%	93.9%	70.8%

[a] The posts were between [-180, 180] date intervals of the protests.

[b] Within six months before the protests.

[c] Based on users' self-reporting in their user profiles on Weibo. Users who do not report their gender, education level and age are omitted.