Physical Repression and Online Dissent: Evidence from Saudi Arabia*

Jennifer Pan† Alexandra A. Siegel‡

October 29, 2018

Abstract

Governments worldwide are using physical intimidation to constrain online expression. While the study of online censorship has received widespread attention, the relationship between physical repression and online dissent is not well understood. We begin to fill this knowledge gap by examining the impact of arrests of Saudi elites on online dissent. We analyze over 300 million tweets and Google search data from 2010 to 2017 using automated text analysis and crowd-sourced human evaluation of content. We find that although physical repression decreased online activity by arrested elites and constrained their criticisms of the government, it did not reign in dissent overall. For the millions of Saudis who actively followed the arrested elites online, observing repression mobilized their interest in the elites and increased their criticism of the regime and its policies. For similar elite actors who were not arrested, observing repression had no impact on their online activity.

Keywords: repression; social media; dissent; censorship; Saudi Arabia

*Our thanks to Will Hobbs, Molly Roberts, and participants at the 2018 APSA pre-conference on politics and computational social science for their helpful comments and suggestions; to the Stanford Center on Global Poverty and Development and the National Science Foundation (Award # 1647450) for research support.

†Assistant Professor, Department of Communication, Building 120, Room 110 450 Serra Mall, Stanford University, Stanford CA 94305-2050; jenpan.com, (650) 725-7326.

‡Postdoctoral Fellow, Immigration Policy Lab, Encina Hall, 616 Serra Street, Stanford University, Stanford, CA 94305-2050

1 Introduction

Governments worldwide are using physical intimidation and repression, in addition to censorship, to constrain online expression. In 2017, more than thirty countries—ranging from authoritarian regimes such as China, Russia, and Iran to democracies such as India, Mexico, and Lebanon—used physical reprisals to rein in online speech. Online journalists, bloggers, and individuals who were critical of the government or its authority were the most frequent targets. Physical reprisals most often included arrests and physical punishment, but dissidents in eight countries were killed for writing about sensitive subjects online (Freedom House 2017).

In this paper, we examine the consequences of one form of physical repression—the arrests of Saudi Arabian elites—on online dissent. Saudi Arabia has one of the highest rates of Twitter penetration in the world.[1] The platform is very popular among demographically diverse individuals, and is widely used to discuss politics (Noman, Faris and Kelly 2015). Online dissent is remarkably commonplace in the Saudi Kingdom, despite the fact that it is a highly repressive theocracy where political rights and civil liberties are severely curtailed.[2] As a result, the Saudi Twittersphere represents an ideal space to investigate the effect of repression on online dissent. It is also an increasingly important context to examine from a policy perspective. Since the time we conducted our analysis, indiscriminate arrests of online activists and more severe forms of repression such as the murder of Saudi journalist Jamal Khashoggi, have heightened instability at home and sparked diplomatic crises abroad (Rauhala 2018).

We examine the consequences of physical repression by analyzing over 300 million tweets between 2010 and 2017, as well as Google search data for the same time period. Our unique data and methods allow us to 1) disaggregate the effects of repression on different actors, 2) to measure both the *volume* of online activity and changes in the *substance* of discussion, 3) to assess changes in public expression (tweets) and private interest (Google searches), and 4) to examine both the short and longer term consequences of re-

[1] As we describe in Section 3, an estimated 8 million people or 41% of the Saudi population uses Twitter (Al-Arabiya 2015).

[2] See Section 3 for a discussion of repression in Saudi Arabia.

pression. In this way, our empirical approach offers a novel opportunity to examine the effects of repression on diverse actors with high resolution.

We find that Saudi elites who personally experienced arrests were *demobilized*. Following their arrests, they decreased their level of Twitter activity. Moreover, large-scale human coding of the content of their tweets demonstrates that they reined in their criticisms of the regime, government policies, and Saudi society, as well as their discussions of collective action. These demobilizing outcomes can be seen immediately after their release from prison and persist up to a year afterward.

Moving beyond the behavior of those who directly experienced repression, we find that arrests *mobilized* online discussion, interest, and anti-regime sentiment among the millions of people residing in Saudi Arabia who actively engage with the arrested elites by rewteeting, mentioning, or replying to their tweets. Immediately following the arrests, these everyday Saudis interacted more with the arrested elites by tweeting about these elites and retweeting their content. The rates at which arrested actors were retweeted remained elevated both one month and one year following the arrests. General interest in the arrested elites, as measured by Saudi Google search behavior, also increased in the month after the arrests but returned to pre-arrest levels shortly thereafter. Because we observe the same heightened interest in both Twitter and Google search data, it does not appear that the public engaged in preference falsification following the arrests (Kuran 1997). When we turn to the content of tweets from ordinary Saudis who followed arrested elites, we find that everyday Saudis' criticisms of the regime and government policies and their discussion of collective action were not constrained, and perhaps even increased in both the immediate- and longer-term aftermath of the arrests. Only when we focus on the subset of users' retweets and mentions of arrested elites, which accounts for less than 2% of the users' tweets about politics, do we see content that is less critical after the arrests. This is unsurprising as retweets by construction include the text of the tweets produced by the elites, which are less critical after their release from prison.

We push a step further and examine the behavior of other elites, who were not arrested but who tweeted in similar ways to those who were arrested. These non-arrested elites

likely face higher risk of repression than ordinary Saudis, and may consequently be more likely to rein in their online dissent. Identifying elites who are similar is no trivial task, and we find them by conducting text matching on the tweets of over one million Saudi Twitter accounts that each have more than 10,000 followers with the tweets of arrested elites. We find that these similar Saudi elites who were not arrested did not change the volume or the content of their tweets following the arrests.

If the Saudi regime only intended to use physical repression to punish and rein in arrested elites, then its strategy was successful. However, the regime did not arrest everyone who expressed dissent online in the period under study, or even all of the well-known elites who dissented. This suggests that the Saudi regime selected a few highly visible opponents to repress in order to generate self-censorship among others not directly targeted. If this was the regime's goal, then its strategy was not successful. The public did not rein in their criticisms of the regime, and if anything increased their expression of online dissent. Elites who were not arrested also continued to disseminate critical content online. The Saudi regime might also have intended to use repression to decrease the influence of arrested online key opinion leaders. If this was its intent, then the strategy was similarly unsuccessful. Although the arrested elites constrained their behavior, they were, if anything, more influential online after their arrests as the average number of retweets for every message they tweeted was higher after their releases from prison than before their arrests.

These results begin to fill a gap in our understanding of how information is controlled in the age of digital technologies and social media. The use of censorship to constrain online speech has received widespread attention from scholars. However, even though there is increasing awareness that governments are using traditional repression to constrain online activities, there is little empirical evidence about the effects of these repressive strategies. This study is a first step in advancing our understanding of how physical repression influences online expression, and it is our hope that others will continue this effort by studying other forms of physical repression (beyond targeted arrests), studying repression in other geographic contexts and types of political regimes, and examining other outcomes

related to online activity.

Finally, while our study is focused on the relationship between repression and online dissent, it has implications more generally for the study of the "dissent-repression nexus" (Lichbach 1987). Scholars of repression and dissent have advocated for the disaggregation of the study of repression beyond country units (Davenport 2005; Davenport and Loyle 2012). Our study illustrates the analytical leverage gained by further disaggregating the effects of repression by type (separating physical repression from censorship), by actor (those who were subjected to repression versus those who observed repression), by behavior (public versus private dissent), and by time (short and long term effects). Because the effects of repression differ along many of these dimensions, without disaggregation, we would come to different conclusions about the relationship between physical repression and online dissent.

The paper proceeds in four sections. Section 2 discusses when we might expect governments to use physical repression to constrain online dissent, how such repression might influence online expression, and under what conditions this strategy is likely to succeed or fail. Section 3 provides background on repression in Saudi Arabia, including information on the arrests of online elites. Section 4 describes our unique sources of data and empirical strategy. Section 5 presents our results, showing the effect of physical repression on arrested elites, similar non-arrested elites, as well as the public, and Section 6 concludes.

2 Physical Repression and Online Dissent

Governments interested in suppressing dissent may adopt a number of different strategies. They could exert control over the technology platforms where opposition takes place, for example blocking access to entire platforms or deleting offending content (Bamman, O'Connor and Smith 2012; Hassanpour 2014; Hobbs and Roberts 2018; King, Pan and Roberts 2013, 2014; MacKinnon 2009; Roberts 2018; Zhu et al. 2013). They may spread disinformation and propaganda to manipulate public perceptions of reality, such as creating fake accounts to increase the appearance of public support for the government (King, Pan and Roberts 2017; Munger et al. 2018). Regimes may also use traditional offline

repression to induce self-censorship online. Governments often adopt these strategies in tandem, but here we discuss why governments might use physical repression to control online expression.

Governments may repress online dissent because physical repression is a part of their institutional culture (Gurr 1988) or general strategy for control (Blaydes 2018). Governments may also perceive social media and online dissent as a more serious threat after the Arab Spring, prompting them to respond more harshly (Hess 2013; Krieg 2016).[3] Governments may believe that physical repression can be applied in a targeted manner to induce broader self-censorship (Stern and Hassid 2012).

Governments may prioritize physical repression when they lack the resources to apply censorship effectively. Online censorship has been most successful in countries like China where Internet content providers censor content quickly and thoroughly in accordance with government demands (Chen and Yang 2018; King, Pan and Roberts 2013, 2014; Roberts 2018). However, for most regimes, the market for social media is dominated by U.S. firms that acquiesce to censorship demands slowly or impartially, such that censorship is limited to Internet blackouts, website blocking, and other visible strategies (Pan 2017).[4] Existing research suggests that censorship constrains dissent when it is invisible and undetectable. Otherwise, censorship may backfire, bringing attention to the very issues that governments sought to suppress.[5]

Finally, governments may use physical repression to constrain online opposition when propaganda and disinformation activities interfere with their ability to gather reliable information. When a government floods an online platform with pro-regime content, it can become more difficult to gauge the extent of public support or the likelihood of elite

[3]Despite debate over the exact effect of social media on mobilization during the Arab Spring (Howard and Hussain 2011; Steinert-Threlkeld 2017), widespread discussion of the role of social media in the Arab Spring has made social media more salient for governments (Morozov 2012).

[4]This was the case prior to the adoption of the General Data Protection Regulation (GDPR). Whether this will change after GDPR is not yet known.

[5]This phenomenon is known as the "Streisand effect" and is named after Barbara Streisand whose desire to censor online photos of her coastal mansion led to widespread online circulation of these photos (Jansen and Martin 2015; Morozov 2012). In examining why the Streisand effect exists, Jansen and Martin (2015) argue that it is in part because the action is visible, and its visibility draws attention—the very thing censorship is intended to avoid. Sometimes the visibility of censorship generates anger and increased discontent (Hassanpour 2014; Nabi 2014). Other times, it leads individuals to seek alternative sources of information that undermine the intent of the censorship (Hobbs and Roberts 2018).

defection.[6] This is especially problematic for authoritarian regimes that do not have electoral competition or free media to provide more accurate sources of information on public sentiment and elite behavior (Egorov, Guriev and Sonin 2009; Malesky and Schuler 2011). In these political contexts, online channels are one of the few reliable, or relatively more reliable, ways autocrats can gather accurate information (Gunitsky 2015; Pearce and Kendzior 2012; Qin, Strömberg and Wu 2015; Qin, Stromberg and Wu 2017).

Low-Intensity and High-Intensity Repression: Physical repression can be separated into low-intensity—imprisonment and violence targeting a relatively small number of individuals—and high-intensity—indiscriminate mass imprisonment and violence (Guriev and Treisman 2015; Stern and Hassid 2012; Way and Levitsky 2006). Physical repression to constrain online dissent is more likely to be low-intensity when a government is integrated into and dependent on international markets and institutions. High intensity repression can threaten economic productivity and growth. Individuals immobilized by fear are not very productive workers; they lack attributes such as innovation and risk-taking, which global markets increasingly demand. High intensity repression is also more likely to generate international censure and sanctions, which can have political and economic consequences. Finally, formal models show that when a government can use low-intensity repression to stay in power, the use of high intensity coercion signals to the public that the government is incompetent and therefore increases the vulnerability of the regime (Guriev and Treisman 2015).

How Does Low-Intensity Physical Repression Constrain Online Dissent? Low-intensity repression can work through direct deterrence (Oberschall 1973; Jenkins and Perrow 1977; Tilly 1978). Repression deters dissent because it represents a negative outcome that makes dissent costly (Hardin 1982; Olson 1965). When individuals are arrested for dissent, they rein in their behavior for fear of future punishment. Deterrence works the same way online and off—people who are arrested for expressing opposition online may

[6]For example, in China, where local governments produce online propaganda, they routinely manipulate information in ways that make it more difficult for central authorities to determine their level of competence and malfeasance (Pan 2016; Pan and Chen 2018).

stop doing so after being subjected to physical repression because they do not want to suffer again.

Low-intensity repression can also work through indirect deterrence (Walter 1969; Durkheim 1984) because repressive actions can inform the public about what is and is not deemed acceptable by the regime. Highly visible low-intensity repression demonstrates how opposition can lead to negative outcomes, which makes dissent more costly in the population as a whole. Indirect deterrence is successful when those who observe repression constrain their own actions. Applied to online dissent, indirect deterrence works when those who observe the repression of online key opinion leaders conclude that online dissent is unacceptable to the government, and believe that they themselves may also be subject to punishment for dissent. This then leads to self-censorship as observers of repression seek to avoid a similar fate.

Finally, low-intensity repression can limit online dissent through downstream effects. Here, repression changes mass behavior by altering the actions of those directly targeted. There are two types of downstream effects. The first operates through the power of leaders over their followers. For example, if a leader of an opposition movement is co-opted by a government, the participants in that movement may also be swayed to support the regime. Online, if a government arrests influential elites and the arrests change the elites from critics to supporters, this attitude may trickle down to their followers. The second type of downstream effect cuts off the power of leaders. For example, by arresting the leader of an opposition movement, a government can harm the organizational capacity of the movement, causing it to flounder (Friedrich and Brzezinski 1965; Bahry and Silver 1987). If arresting online opinion leaders silences them, their followers may unfollow them or stop engaging with them because they are no longer active.[7] Both types of downstream effects would suppress online dissent but through different mechanisms—by leveraging the influence of the online elites or by cutting off their influence.

[7]Note that here, followers are not disengaging because of fear of punishment but simply because the voice is no longer active and producing interesting content.

When Does Low-Intensity Physical Repression Succeed or Fail? For low-intensity physical repression to succeed as a direct deterrent, the costs of dissent must outweigh the benefits. We would only expect arrests to constrain behavior if imprisonment were painful enough that any benefit derived from expressing dissent would be overcome by the fear of being imprisoned again. To succeed as an indirect deterrent, those who observe the repression must want to avoid repression and also believe that they could be targeted. When large groups of people are engaged in dissent, the risk that any one of them will be repressed decreases. Repression may have limited effects as an indirect deterrent for online behavior. Social media allows millions of people to express themselves online, making individuals less likely to believe they will be punished, and thus less likely to rein in dissent after observing repression. People who are similar to those who have been directly targeted, however, may be more likely to change their behavior. For example, if a human rights lawyer with a large online following is arrested, then other human rights lawyers with many followers may be more likely to constrain their behavior. However, additional factors can influence these levels of perceived risk, such as a dissident's ability to defect and leave the country, as well as political connections, and sources of political support.

Repression is also more likely to succeed as an indirect deterrent when it increases uncertainty about who might be targeted next (Link 2002; Stern and Hassid 2012). When ambiguity of repression increases, observers of repression are unsure of whether their actions could also lead to punishment, which generates greater self-imposed constraints. On the flip side, if the reasons for physical repression are clear to observers then repression is less likely to be effective.

Physical repression is more likely to have downstream effects on online dissent if there are few or no leadership alternatives. Taking the first type of downstream effect—if a social media influencer with many followers is repressed and reins in his or her online dissent—the leader's followers will be more likely to constrain their behavior if there are no alternative voices expressing dissent. With regard to the second type of downstream effect, the repression of influential dissenters may limit overall criticism online. When

an online opinion leader is repressed, the viewpoints endorsed that person may fade over time as other topics gain the public's attention.

Finally, physical repression may have differing effects on online mobilization over time. Offline, repression demobilizes in the short term because it makes targeted dissidents fearful, but generates more diffuse anger and makes it easier to recruit and mobilize new dissenters in the longer-term (Rasler 1996). We may also see these differing time effects of repression online, but possibly for different reasons. On one hand, we may only see short term effects of repression on online expression because social media is bursty and trending topics change quickly.[8] On the other hand, we might observe long-term effects because of the social networks and connections embedded in some online spaces. For example, if physical repression has durable effects on those who are directly repressed, and those actors are online elites with many followers, the effects may persist across the social network.

Taken together, our understanding of how low-intensity repression might impact online dissent suggests a need to disaggregate the effect of physical repression in three key ways. First, we need to disaggregate the effect of repression between those who are directly targeted and those who observe it. Second, because physical repression is aimed at constraining behavior, its effects may be private. As a result, we need to disaggregate between public dissent (online tweets), and private behavior (Google searches). Finally, we also need to disaggreate by time—between changes in the short-term aftermath of physical repression and longer-term effects.

3 Repression in Saudi Arabia

There is very little space for dissent in Saudi Arabia. Almost all political rights and civil liberties are curtailed through laws, coercion, and surveillance (Freedom House 2018; Gibney et al. 2016; Human Rights Watch 2018*a*).[9] Political parties are banned, political

[8]For findings on the "burstyness" of social media data, see, for example, Cordeiro and Gama (2016); Cheng et al. (2016), and Zhao et al. (2012).

[9]Over the past several years, Saudi Arabia has received scores of three and four on the Political Terror Scale (PTS), which measures countries' annual level of state political violence and terror. The data used in compiling this index comes from three differential sources: the yearly country reports of Amnesty International, the U.S. State Department Country Reports on Human Rights Practices, and Human Rights Watch's

dissent is criminalized, and organized opposition only exists outside of the country. Activists who challenge the regime—whether by highlighting the monarchy's human rights record or demanding constitutional reform—are routinely arrested, and allegations of torture and abuse by police and prison officials are common (Alabaster 2018; Calamur 2018; ESHR 2017). Protests are rare and are violently repressed when they occur (Ménoret 2016).

In contrast to these severe constraints on offline dissent and real-world mobilization, political dissent has found a foothold in the Saudi online sphere. With high levels of literacy and internet penetration, social media adoption is high, and online platforms have provided an alternative space for political expression and civil society organizing (Freedom House 2018; Worth 2012). The Saudi Twittersphere has become a particularly popular venue for political discussion. An estimated 8 million people or 41% of the Saudi population use Twitter (Al-Arabiya 2015), and although most Saudi Twitter users are relatively young, because 70% of the Saudi population is under the age of 30, the Saudi Twittersphere constitutes a large and diverse subset of the population (Glum 2015). Tweets containing political and social commentary, political dissent and criticism of the monarchy, and complaints about corruption and the quality of public services proliferate. Indeed, politics is one of the most popular topics of conversation, just behind religion and soccer (Noman, Faris and Kelly 2015). While many Saudis tweet using pseudonyms—fearing government surveillance of their online activities and the repercussions of being identified—many well-known clerics, activists, and other elites have easily identifiable Twitter accounts with large followings (Ibahrine 2016; Siegel 2015).

The Saudi regime has struggled to regulate social media. The Minister of Information admitted in February 2013 that monitoring Twitter was difficult due to the large volume of users (al Rasheed 2013). Indeed, censors in Saudi Arabia likely do not have the capability in their filtering infrastructure to block access to specific Twitter accounts without block-

World Reports. A score of three on the PTS scale suggests that: "There is extensive political imprisonment, or a recent history of such imprisonment. Execution or other political murders and brutality may be common. Unlimited detention, with or without a trial, for political views is accepted." A score of four means that "civil and political rights violations have expanded to large numbers of the population. Murders, disappearances, and torture are a common part of life. In spite of its generality, on this level terror affects those who interest themselves in politics or ideas" (Gibney et al. 2016).

ing the platform entirely, which is politically unfeasible (Noman, Faris and Kelly 2015). Government officials have requested user information from Twitter to monitor dissent. For example, between January and June 2015, there were 93 requests to Twitter for account information, and Twitter reports turning over information for 69% of these requests (Report 2015). However, these information control efforts are incomplete and too slow to be effective, given how quickly information can spread online (Pan 2017).

The Saudi state has turned to the same strategies—including legal action, coercion, surveillance—that it uses offline to punish online dissent. Saudi Arabia carries out some of the most severe physical repression of Internet users in the world, ranking sixth behind China, Iran, Syria, Egypt, and Bahrain (Freedom House 2017). Official statements from the Saudi regime suggest that it will punish individuals who post tweets that offend the rulers, dissent against the monarchy, express attitudes of class superiority, ignite regional prejudices, offend clerics, promote intellectual deviation, promote extremism, or destabilize security (Makkah Newspaper 2015). In other words, as in other authoritarian regimes, what is punishable is broad and ambiguous. Saudi authorities have used a 2007 anti-cybercrime law to crack down on lawyers and activists who peacefully criticize the government on Twitter (Human Rights Watch 2014). The government has also criminalized a broad range of online activities by labeling them as acts of terrorism. These include questioning the Kingdom's religious foundation, "unsettling the social and national fabric...or any actions that touch the unity and stability of the Kingdom under any reason and in any form" (Human Rights Watch 2014).

Over the past decade, dozens of elite actors in the Saudi Twittersphere—including popular clerics, well-known human rights activists, women's rights activists, and Shia rights activists—have been arrested, often in response to their online activity. For example, on July 20, 2013, two well-known Saudi clerics with millions of online followers—Mohamed al-Arefe and Mohsen al-Awaji—were arrested for denouncing the Egyptian military coup ousting Muslim Brotherhood president Mohammed Morsi, which was in direct opposition to the stance of the Saudi regime that recognized the legitimacy of the coup. While no official charges were made public, activists and human rights organiza-

tions attributed these arrests to comments, written communiques, and Youtube videos circulated on Twitter and Facebook opposing the military coup in Egypt and supporting the Muslim Brotherhood. Both clerics were released after two days after promising to desist from further "intervention in the affairs of other countries" (Admoun 2013). In October of 2014, three lawyers—Abdulrahman al-Subaihi, Bandar al-Nogithan (a graduate of Harvard Law School), and Abdulrahman al-Rumaih—were sentenced to between five and eight years in prison for criticizing the justice system on Twitter. Official charges stated that the lawyers were convicted of disobeying the ruler and slandering the judicial system online. Their convictions were overturned in April 2015, and they were released shortly afterwards (Amnesty International 2016). In December 2014 women's rights activists Loujain al-Hathloul and Maysa al-Amoudi were arrested and detained for 73 days after al-Hathloul tried to drive into Saudi Arabia from the United Arab Emirates in defiance of the Saudi women's driving ban. Al-Amoudi, a UAE-based Saudi journalist, arrived at the border to support al-Hathloul and was arrested as well. Officially, al-Hathloul and al-Amoudi were charged under vague provisions of an anti-cybercrime law, but according to human rights organizations monitoring the case, the real motive behind the arrests was not defiance of the driving ban but voicing their opinions online. At the time of her arrest, al-Hathloul had 232,000 followers on Twitter, and her tweets detailing the 24 hours she spent waiting to cross into Saudi Arabia after border officers stopped her had gone viral. Al-Amoudi had 136,000 followers at the time of her arrest and was also well known for hosting a program on YouTube calling for an end to the Saudi driving ban (Alferaehy 2016).

These examples show that between 2010 and 2017, a wide variety of Saudi elite actors were arrested for relatively short periods of time because of their online activity. As such, Saudi Arabia offers an ideal context to test the effect of one form of physical repression on online dissent. The arrests of well-known elites, who are are active on Twitter and have large followings, allow us to assess the effect of repression on their online behavior, as well as that of their followers and other similar elites who were not arrested.

4 Data and Empirical Strategy

Here, we describe our unique, large-scale data, as well as the computational and statistical approaches we take to examine the effects of arrests on different actors over time.[10]

4.1 Data

We gather four datasets to assess how the online behavior of mass and elite actors changed in the aftermath of arrests: 1) a dataset of all tweets produced by arrested elites, 2) a dataset of tweets by ordinary Saudis who engaged with (retweet, mention, or reply to) arrested elites, 3) Google search data for the names of arrested elites, and 4) a dataset of all tweets produced by elites who are similar to those arrested in terms of the content of their tweets but who were not arrested in the period under study. The first dataset allows us to examine the online behavior of those who directly experienced physical repression. The second dataset allows us to analyze the content of tweets produced by everyday Saudis who engaged with the arrested elites. The third dataset enables us to assess private online interest in these actors, to determine whether results we observe in the second dataset might be driven by preference falsification. The last dataset allows us to examine the online behavior of actors who are similar to the arrested elites but did not directly experience physical repression in the period under study.

1. Tweets of Arrested Elites: In order to analyze the consequences of physical repression on the behavior of elites and ordinary people over time, we began by identifying Saudi elites who had been arrested and were active on Twitter between January 1, 2010 and January 1, 2017.[11] We compiled a list of 36 individuals whose arrests were widely reported in either the Saudi or international press that were active on Twitter. These elites, which include those we described in Section 3, range from prominent Sunni clerics to human rights activists, women's rights activists, lawyers, and Shia clerics and rights ac-

[10] Our data is entirely observational, and it is not our intent to make causal claims with our data or empirical strategy. We use terms such as "effects" for ease for exposition, not to signify causal inference.

[11] We chose January 2010 as a start date because Twitter became increasingly popular across the Arab World in the early days of the Arab Spring protests. We began our historical data collection in January 2017, which marks the end of our data collection period.

tivists. Table A1 in the Appendix lists all arrested elites, including a brief description of their background, the official justification for their arrests, as well as the unofficial reasons for arrest provided by human rights organizations. As Table A1 demonstrates, many of these individuals were explicitly arrested for the content of their tweets or their online activity. Several of the elites in our dataset were arrested on multiple occasions. These arrests tended to be separated by at least a year, and were often in response to different activities. For example, the Sahwa cleric Mohammed al-Arefe was first arrested and briefly detained in response to his comments about the Muslim Brotherhood in Egypt as described in Section 3, and then, over a year later, was arrested for his critical comments about the Saudi Hajj pilgrimage train. In our study, we limited our analysis to elites' first arrests in order to keep the analysis consistent across all arrested elites.[12] A table of the dates of these first arrests is provided in Table A2 in the Appendix.

After identifying these 36 elites, we then collected all of their tweets produced between January 2010 and January 2017 using Twitter's Historical PowerTrack API. This API provides access to the entire historical archive of public Twitter data—dating back to the first tweet—using a rule-based-filtering system to deliver complete coverage of historical Twitter data. This gave us a dataset of 408,511 tweets produced by our 36 elites from 2010 to 2017.

2. Tweets of Ordinary Saudis: We are also interested in how repression influences those who observed it but were not subject to it. The individuals most likely to have observed these arrests were individuals who had actively engaged with the arrested elites on Twitter. We do not include everyone who follows the arrested elites because that would include individuals who were not attentive—for example, people who have Twitter accounts but who do not use Twitter, or people who followed these elites because it was recommended by Twitter's algorithm but were not actually interested in these individuals. We again used the Historical PowerTrack API to download all public tweets engaging with the arrested elites using the @ sign (for example, @LoujainHathloul). We then filtered

[12] Our analysis is limited to those elites who were released during our time our data collection, 25 of the 36 elites.

this dataset to include only mentions made by individuals who were either geolocated in Saudi Arabia or contained location metadata in the location or timezone fields of their profiles indicating that they were located in Saudi Arabia. We collected 32,504,397 tweets produced by 8,506,400 users likely to be located in Saudi Arabia who retweeted, mentioned, or replied to our arrested elites between 2010 and 2017. In addition, we selected a random sample of about 30,000 of these actively engaged users, stratified by arrested elite, and used Twitter's API to scrape up to 3200 of each of their most recent tweets for a total of 47,886,355 tweets.[13] The set of 47,886,355 tweets allows us to examine the influence of arrests on how individuals who engaged with arrested elites express themselves in general discussions of politics.[14] The set of 32,504,397 retweets, mentions, or replies made by ordinary Saudis in response to arrested elites allows us to examine more narrowly the downstream effects of the arrests.

3. Google Search Data on Arrested Elites: As a means of measuring mass private, rather than public, interest in the arrested elites, we downloaded daily Saudi Google Search data for the Arabic names[15] of each arrested elite in the month preceding and following each arrest. This enabled us to see how often Saudi Google users searched for these individuals. We also downloaded weekly Google Search data for the year preceding and following each arrest to obtain data over a longer time horizon.[16] Because individuals conducting Google searches are generally alone, and there is no obvious record of their activity, they are more likely to express socially and politically taboo thoughts in their searches than they might in more public forums (Conti and Sobiesk 2007; Stephens-Davidowitz 2014, 2017). The global popularity of searches for pornography and embarrassing medical conditions is a clear example of this phenomenon. This data therefore

[13] For most of these non-elite users, this encompassed all of the tweets they have ever made.

[14] In order to use these tweets to assess the sentiment of political content, we first filtered these tweets to include only those that contain the most common political keywords in a random sample of tweets sent by the arrested and match elites coded on Crowdflower as relevant to the Saudi regime, politics, or society. This left us with a total of 16,427,785 potentially politically relevant tweets. The full list of these keywords and their translations is provided in Figure A2 in Appendix C.

[15] We manually checked to ensure that these search terms were in fact drawing results related to these elites by examining the "related queries" provided in the Google Search data. We excluded the names of elites from our analysis that had too low of a search volume to restrict the analysis to Saudi Arabia.

[16] It is possible to download daily Google search data for up to a 90 day period and weekly Google search data for up to a five year period.

provides a kind of mass-level "truth serum"—revealing sensitive attitudes and behaviors that may not be apparent in Twitter data. In this way, Google search data allows us to develop a real-time behavioral measure of how much attention everyday Saudis were privately paying to arrested elites, beyond the direct public engagement we measure using our dataset of Twitter mentions. Furthermore, by comparing the results of the Twitter mentions data to the Google search data, we may capture any preference falsification (Kuran 1997) that could have occurred in the mass public. For example, if we see evidence of deterrence in our Twitter mentions data and mobilization in our Google search data, this might indicate that everyday Saudi Twitter users were self-censoring or falsifying their preferences to avoid punishment.

4. Tweets of Similar, Non-Arrested Elites: A theoretically important population we also want to examine are Saudi elites who are similar to those arrested but who were not themselves subjected to repression during our period of study. To identify these elites not directly subjected to repression, we first used the Historical PowerTrack API to download all tweets sent by Twitter users who had over 10,000 followers located in Saudi Arabia, based on their geo-location and location metadata, between 2013 and 2014.[17] This resulted in 235,215,314 tweets sent by 1,048,568 unique accounts. We then measured the average cosine similarity between tweets sent by these accounts and our elite accounts to find "matches" for each elite.[18] Each arrested elite account was matched to the account of non-arrested elite that contained tweets most closely matching the arrested elite's tweets.[19] We found 13 unique matches because several elites had the same top match.[20]

Our method resulted in matches that made sense substantively—Sunni clerics were matched with Sunni clerics, Shia clerics were matched with Shia clerics or Shia-rights activists, women's right activists were matched with women's rights activists, human rights

[17] We choose 2013 and 2014 because this is approximately the middle of our data collection period for arrested elites.

[18] We did not use a specific threshold for cosine similarity but simply chose matches that had the highest rates. The average cosine similarity value was .22, and ranged from .09 to .57.

[19] As might be expected, the closest match for many of our arrested elites were other arrested elites, and the closest non-arrested match was not necessarily the closest match over all.

[20] For example, three of our arrested women's rights activists were matched to the women's rights activist Hala al-Dosari, who was not arrested. Several arrested human rights activists were matched to the human rights activist Waleed al-Sulais, who also was not arrested.

activists were matched with other human rights activists activists, etc. [21] We then used the Historical PowerTrack API to download all of their public tweets from 2010 to 2017, resulting in a dataset of 365,337 tweets.

Our goal in identifying matches is not to find individuals who are identical to the arrested elites. We are not conducting matching for purposes of causal inference. We are finding matches in order to identify elites who might face greater threat of repression than ordinary Saudis, and who may be most likely to rein in their online dissent after seeing their peers arrested.

4.2 Empirical Strategy

We analyze changes in the overall volume of online expression, the overall volume of private online behavior, and the content of what is expressed online between the pre and post arrest periods for arrested elites, their active followers, and similar non-arrested elites.

Interrupted Time Series Analysis: To examine the effect of arrests on the *volume* of online activity—both tweets and Google searches—we use interrupted time series analysis (ITSA) to measure online activity prior to arrest compared to after release for arrested elites, and to measure online activity prior to arrest compared to after the arrest for non-arrested elites and the public. We describe the model in detail in Appendix B.

Crowdsourced Evaluation of the Content of Tweets: Moving beyond changes in the volume of activity, we also evaluate how the content of tweets produced by elites and everyday Saudis changed in the aftermath of repression. In particular, we are interested in the effect of arrests on dissent, which includes four categories of content: 1) criticisms of the regime, 2) criticisms of government policies, 3) criticisms of Saudi society, and 4) discussions of collective action.

The first category focuses on tweets that express dissatisfaction with or criticize the Saudi monarchy including specific royal family members, members of the religious establishment such as state-sanctioned clerics, or religious doctrine. It also includes tweets

[21] see Table A3 in the Appendix for information and account metadata for all arrested elites and matches.

calling for democracy or other changes to the form of government. This category focuses on content that challenges the legitimacy of the religious monarchy, and as such likely represents the most intolerable form of online expression for the Saudi regime. The second category includes tweets that express dissatisfaction with or are critical of Saudi bureaucracy including the judiciary, government ministries, or the religious police. It also include tweets criticizing or expressing dissatisfaction with policies and policy outcomes such as the state of the economy, corruption, foreign policy, and infrastructure. This category is perhaps less problematic for the regime as it challenges policies but not the underlying legitimacy of the regime. The third category identifies tweets criticizing Saudi society for being too liberal or too conservative, as well as tweets criticizing the role of women in society. Because these tweets focus on Saudi society in general, they may be more likely to be tolerated. The final category are tweets discussing protest or organized crowd formation on the ground. While rare, these tweets represent a particularly problematic form of dissent for the monarchy in the post-Arab Spring period because they facilitate and spread awareness of offline mobilization.

To classify tweets into these categories, we crowdsourced large-scale human coding of tweets via Crowdflower, a platform similar to Mechanical Turk but with more native Arabic speakers. We used Crowdflower to code about 10,000 tweets produced by arrested and non-arrested elites and about 20,000 tweets produced by the engaged followers of arrested elites for a total of approximately 30,000 coded tweets. 5,000 of the elite tweets were selected through stratified random sampling of all tweets produced by the arrested and non-arrested elites over the month preceding arrest and the month following release, balanced by actor type (Sunni clerics, women's rights activists, liberal activists, lawyers, anti-corruption activists, and Shia rights activists). The other 5,000 tweets were sampled from all elite tweets produced between six months and one year following arrest, again balanced by actor type.[22] We similarly sampled tweets that mentioned or retweeted the arrested elites (another 10,000 tweets), and finally sampled tweets containing political keywords that were produced by the elites' engaged followers (another 10,000 tweets).

[22]We did not collect data from a year pre-arrest because substantively we wanted to compare content produced in the lead-up to the arrest (the period during which the regime decided to constrain the elites' behavior) with content produced in the immediate aftermath of repression and in the longer term

Three native Arabic speakers assessed each tweet on the Crowdflower platform.[23] Across all samples, intercoder agreement was very high, with 95% agreement among coders on average.[24] The majority of tweets about the Saudi regime, policies, and society expressed negative sentiment (72%, 75%, and 60% respectively) and very few tweets called for collective action (less than 1% of all coded tweets).[25]

5 Differential Effects of Physical Repression on Online Dissent

Here, we show how the volume and content of online activity changed in the aftermath of repression.

5.1 Demobilization of Arrested Elites

Elites who were subjected to physical repression were deterred from dissent. Figure 1 presents the pre-arrest and post-release trends in the volume of tweets produced by arrested elites plotted as local regression lines with loess smoothing and 95% confidence intervals. The left plot of Figure 1 shows the daily volume of tweets produced by the arrested elites in the month before arrest and month after release, and the right plot shows the daily volume of tweets in the year before arrest and year after release.[26]

As Figure 1 demonstrates, arrested elites tweeted significantly less in the post-release period relative to the pre-arrest period. This is true both in the short term (a month before arrest and a month after release—left panel of Figure 1), and this is also true in the longer term (comparing a year before arrest and a year after release—right panel of Figure 1). The results of our interrupted time series regressions (Table A4 and Table A5) demonstrate that these effects are negative and statistically significant both in the month and year time frames. In order to ensure that these effects are not just driven by the change of behavior of one or two elites, we also replicate this analyisis including elite fixed effects.

[23] The coding scheme used by the Crowdflower workers is presented in detail in subsection C.1 in the Appendix.

[24] A table of average intercoder agreement by coding category can be found in Table A13 in the Appendix.

[25] Histograms of these proportions can be found in Figure A1 in the Appendix.

[26] Regression tables showing the results of our interrupted time series analysis can be found in Table A4 and A5 in the Appendix.

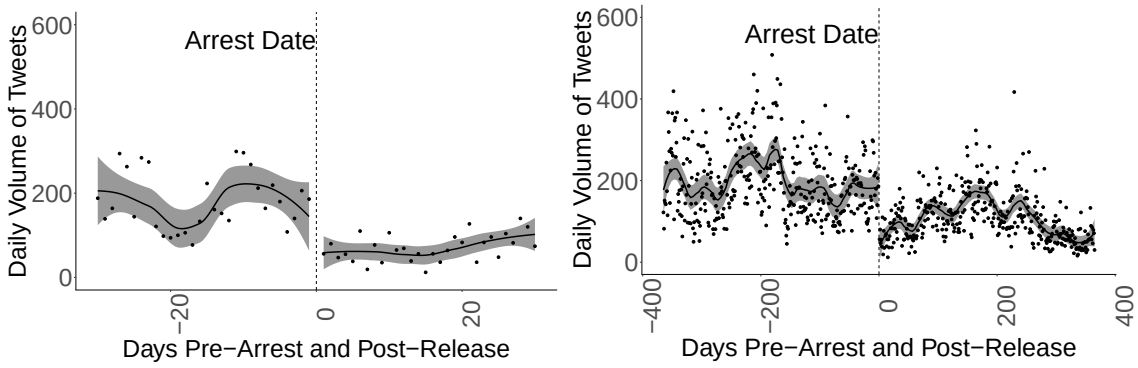


Figure 1: Pre-arrest and post-release trends plotted as local regression lines with loess smoothing and 95% confidence intervals based on the daily volume of tweets produced by the arrested elites in the month before arrest and month after release (left panel), and in the year before arrest and year after release (right panel); regression tables of these results found in Table A4 and A5 of the Appendix.

These results, reported in Table A6 again show demobilization in the month and year time frames.

We also observe a deterrent effect of physical repression on arrested elites in the content of their tweets. The left panel of Figure 2 shows barplots of the average sentiment of each tweet type pre-arrest (black bar), one month post-release (light gray bar), and one year post-release (dark gray bar). The right panel of Figure 2 shows the results of t-tests evaluating the change in average tweet sentiment of tweets produced by arrested elites one month before the arrests and one month (and one year) following the releases, with 95% confidence intervals.[27]

As the left panel of Figure 2 demonstrates, before the arrests, these elites expressed very negative sentiment toward the regime, its policies, and society in general. After their releases from prison, they expressed significantly more positive (less negative) sentiment toward the Saudi regime and Saudi policies or bureaucracy. This result is particularly strong immediately after their releases, but persists up to a year afterward. Whereas in the month before their arrests, the average sentiment of their tweets about the Saudi regime was quite negative (-.7 on a scale ranging from -1 to 1), in the month after their release, the average sentiment had risen to +.15. In the year following the releases, average sentiment

[27]Each tweet was coded by three coders on Crowdflower as expressing either a positive, negative, or neutral attitude toward the Saudi regime, policies, or society. Tweets coded as irrelevant were excluded from the analysis.

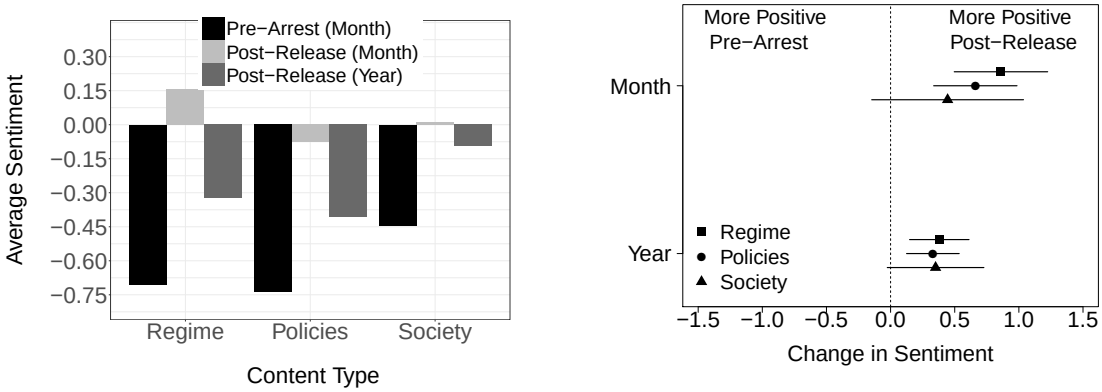


Figure 2: Barplots of the average sentiment of tweets of arrested elites in the month pre-arrest (black bar), the month post-release (light gray bar), and six months to one year post-release (dark gray bar) by content category (left panel); estimates of the change in average sentiment of tweets of arrested elites before arrests and after release based on t-tests with 95% confidence intervals (right panel).

was again negative, but far less negative than it had been in the pre-arrest period (-.3). A similar pattern is evident when examining their tweets about Saudi policies. Their tweets about Saudi society were also more positive (less negative) on average, though this result is not statistically significant. This suggests that arrested elites not only tweeted less following their releases, but that the content of their tweets became less negative toward the regime, policies, and society.

Figure 3 shows changes in arrested elites' online discussion of collective action. Ar-

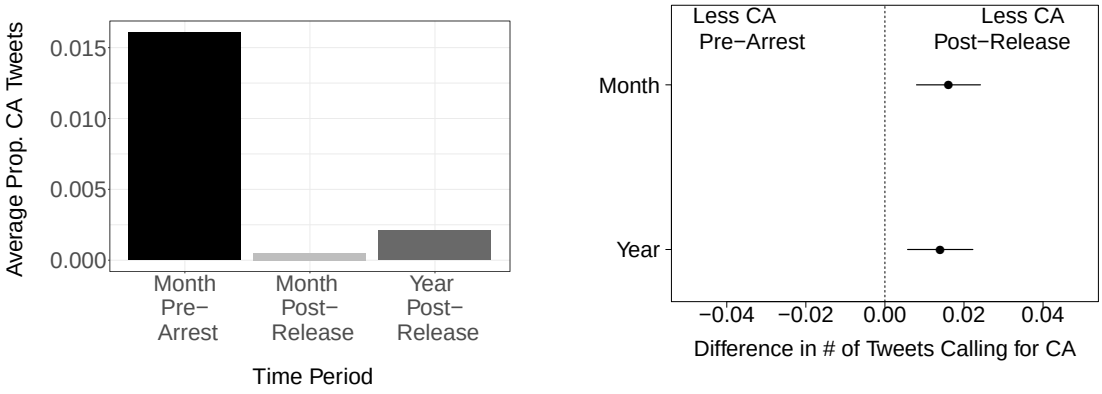


Figure 3: Barplots of the average proportion of arrested elite tweets calling for collective in the month pre-arrest (black bar), the month post-release (light gray bar), and six months to one year post-release (dark gray bar); estimates of the change in proportion of tweets calling for collective action before arrest and after release based on t-tests with 95% confidence intervals (right panel).

rested elites produced fewer tweets calling for collective action post-release, relative to the pre-arrest period, and again, this difference persists one year post-release. Although tweets calling for collective action are quite rare in the Saudi Twittersphere (around 1.5% of arrested elite tweets prior to arrest), they were nonetheless more likely to be produced by arrested elites in the pre-arrest period, and they disappeared almost entirely post-release.

These results suggest that low-intensity physical repression was a direct deterrent on the behavior of arrested elites engaged in dissent. Even though many of the arrests were short-term, arrested elites reined in their criticisms of the regime, its policies, and Saudi society, and already rare online posts about offline mobilization essentially disappeared.

5.2 Mobilization of the Public

The arrests mobilized short-term public engagement with the arrested elites on Twitter. This is seen in Figure 4, which plots pre- and post-arrest trends in the volume of mentions of arrested elites as local regression lines with loess smoothing and 95% confidence intervals. In the left panel of Figure 4, which shows the daily volume of tweets mentioning arrested elites in the month before arrests and month after, there is a large spike in Twitter engagement after the arrest dates.[28] This pattern holds in the right panel of Figure 4, which shows a peak in mentions immediately after the arrest when taking into account the daily volume of tweets mentioning arrested elites in the year before arrests and in the year after.

Our analysis of Google search data yields similar spikes of interest in the arrested elites following their arrest (Figure 5). We do not find evidence of preference falsification as private interest in arrested elites follows a similar pattern to public engagement, increasing significantly in the immediate aftermath of the arrest, but ultimately returning to pre-arrest levels. These findings are presented in Figure 5, which again plots the pre-arrest and post-arrest trends in our data as local regression lines with loess smoothing and 95% confidence

[28]There is also an uptick in daily mentions approximately ten days before the arrest. Many of the elites were arrested for their online activities. The uptick may denote the tweet(s) that the regime deemed to be problematic.

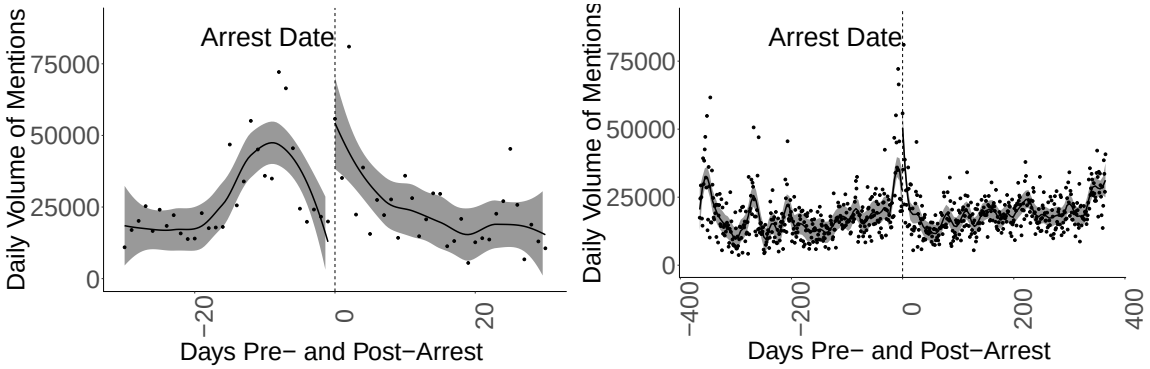


Figure 4: Daily volume of mentions of arrested elites in the month pre and post arrest (left panel) and in the year pre and post arrest (right panel) plotted as local regression lines with loess smoothing and 95% confidence intervals. Regression tables showing the results of our interrupted time series analysis can be found in Table A9 and Table A10 in the Appendix.

intervals, and show a large spike in search volume immediately following in the arrests.[29] These results provide further evidence of the brief mobilizing effect, which dissipates over time.

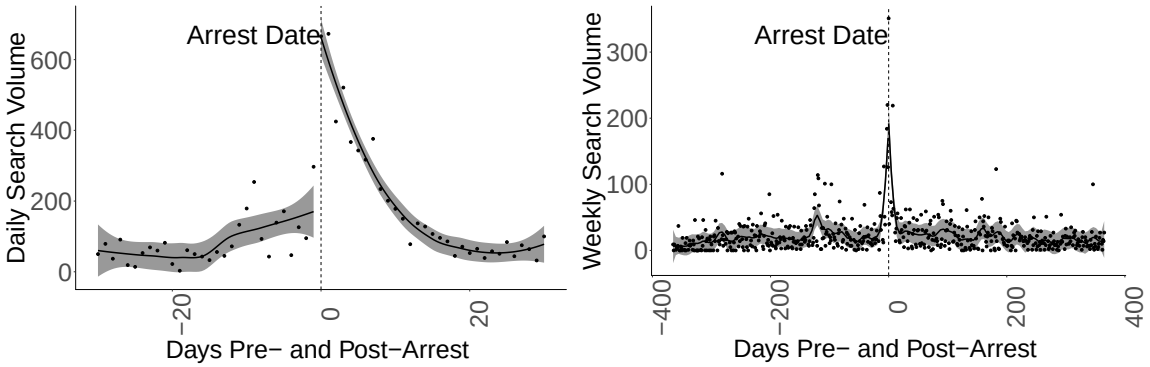


Figure 5: Daily relative volume of Google searches of arrested elites in the month pre and post arrest (left panel) and weekly relative volume of Google searches in the year pre and post arrest (right panel) plotted as local regression lines with loess smoothing and 95% confidence intervals. Google search data is a relative measure of the popularity of a given search term on Google. Each data point is the total number of searches for a given term divided by the total number of searches from that same geogrpahic region and time window. The resulting numbers are then scaled on a range of 0 to 100 based on a topic's proportion to all searches on all topics. Regression tables showing the results of our interrupted time series analysis can be found in Table A11 and Table A12 in the Appendix.

[29]Google search data is a relative measure of the popularity of a given search term on Google. Each data point for is the total number of searches for a given arrested elite divided by the total number of searches from that same geogrpahic region and time window. The resulting numbers are then scaled on a range of 0 to 100 based on a topic's proportion to all searches on all topics.

Although these spikes in public and private interest quickly dissipate, the level of engagement per tweet made by the arrested elites remains higher after their release than prior to their arrest in the longer term. As Figure 6 demonstrates, on average tweets sent by arrested elites garnered more retweets per tweet when comparing the month before and after the arrests (p=.084), and when comparing retweet per tweet made by an arrested elite in the year before and after the arrests (p=.007). The average number of retweets per tweet in the year pre-arrest period was 31 and the average number of retweets per tweet in the year post-release period was 110. This indicates that everyday Twitter users engaged more with elites in the aftermath of repression, suggesting that physical repression did not decrease the online influence of arrested elites over their followers.

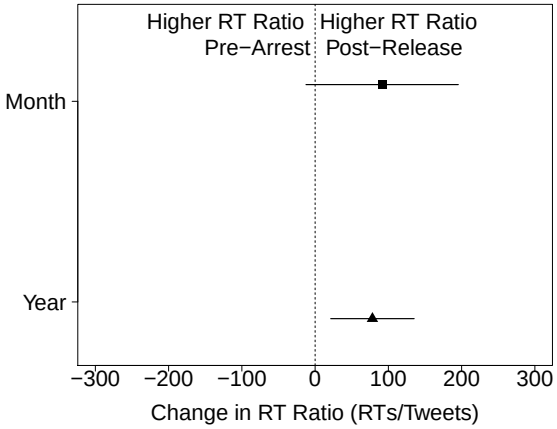


Figure 6: Change in average number of retweets per tweet made by arrested elites before and after their arrest based on t-test with 95% confidence intervals; compares the month before and after arrests and the year before and after arrests.

Arrests of elites did not reduce negative public attitudes toward the regime, and perhaps increased them. These results are shown in Figure 7. The left panel is a barplot of the average sentiment of political tweets produced by Saudi Twitter users who engaged with arrested elites. The average sentiment is always negative, but in the month (light gray bar) and year (dark gray bar) after the arrests, online sentiment is more negative toward the regime, policies, and society than before the arrests. The right panel presents t-tests of the difference in sentiment of tweets in the month and year before and after arrests with 95% confidence intervals, showing increasingly negative sentiment in all issue areas and in both time periods of comparison.

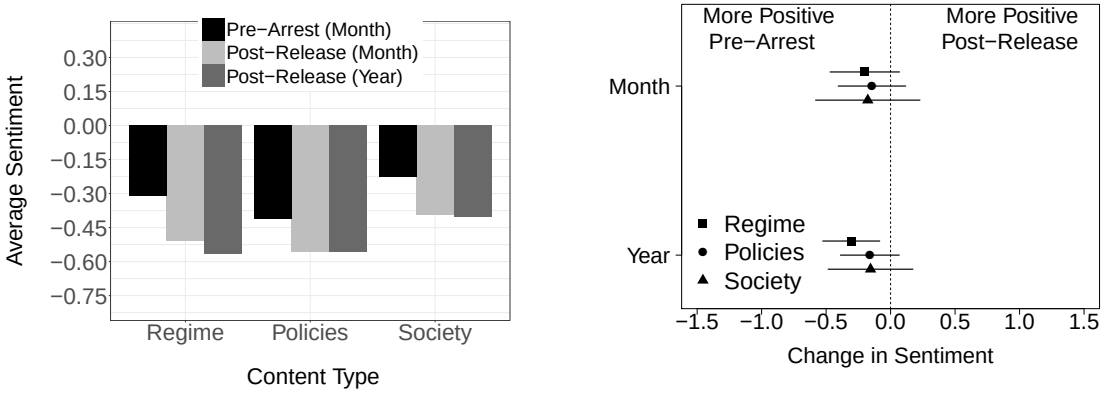


Figure 7: Barplots of the average sentiment of tweets made by those who engaged with arrested elites in the month pre-arrest (black bar), the month post-release (light gray bar), and six months to one year post-release (dark gray bar); estimates of the change in average sentiment of the public's tweets before arrest and after release based on t-tests with 95% confidence intervals (right panel).

This pattern of mobilization holds with regard to calls collective action. As Figure 8 demonstrates, there is an increase in tweets calling for collective action among the public in the post-release period, though these results were not statistically significant and these discussions of collective action are quite rare.

Our analysis of the content of posts made by Saudi Twitter users who engaged with the arrested elites shows that even though the arrested elites reined in their criticisms of the Saudi regime, this does not trickle down to the broader Saudi Twittersphere. We see no evidence of demobilization and perhaps some evidence of mobilization in the tweets of the public. The only exception is when we analyze the subset of tweets directly mentioning or

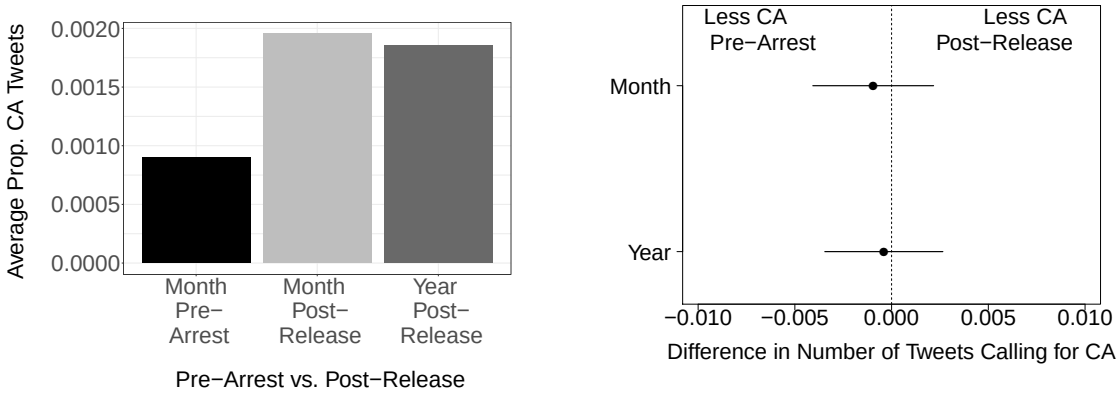


Figure 8: Barplots of the average proportion of tweets made by those who engaged with arrested elites calling for collective in the month pre-arrest (black bar), the month post-release (light gray bar), and six months to one year post-release (dark gray bar); estimates of the change in proportion of tweets calling for collective action before arrest and after release based on t-tests with 95% confidence intervals (right panel).

retweeting arrested elites, which represent less than 2% of their political tweets. Figure 9 shows the sentiment of tweets that retweet, mention, or reply to the arrested elites, and we see the public's sentiment reflecting changes in the less negative (more positive) sentiment of the arrested elites. In the left panel, average sentiment in these retweets, mentions, and replies are always negative, but become less negative in the month (light gray) and year (dark gray) after the arrests than in the month (black) before. The right panel of t-tests of differences in the average sentiment of the public's direct engagements with arrested elites shows that sentiment became more positive, and the result is statistically significant for sentiment toward the regime. We see no change in their tweets calling for collective action.[30]

These results suggest that low-intensity repression did not act as an indirect deterrence, and if anything, repression backfired. Immediately following the arrests, everyday Saudis tweeted more about the arrested elites, and general interest in the arrested elites, as measured by Saudi Google search behavior, also increased in the short term. Importantly, everyday Saudis' criticisms of the regime and government policies and their discussion of collective action were not constrained, and perhaps even increased in both the immediate and longer term aftermath of the arrests. This absence of a chilling effect among

[30]These results are displayed in Appendix C, Figure A3.

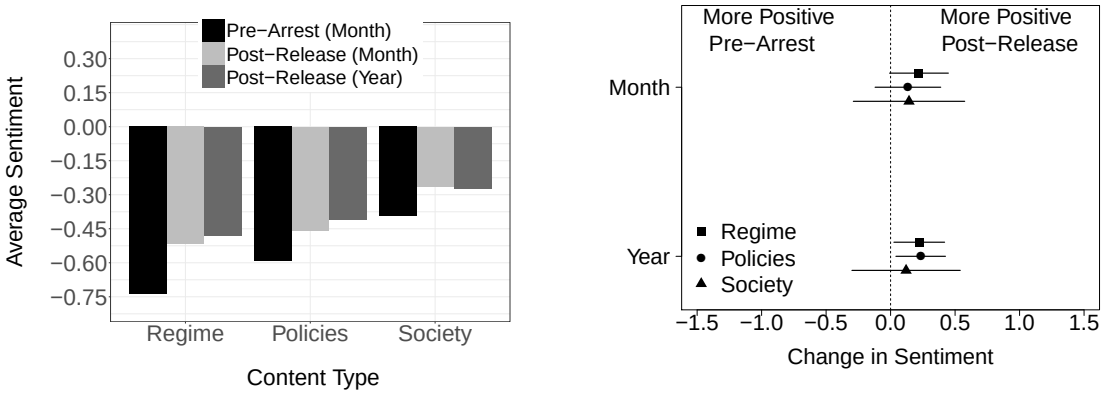


Figure 9: Barplots of the average sentiment of retweets, mentions, and replies to arrested elites in the month pre-arrest (black bar), the month post-release (light gray bar), and six months to one year post-release (dark gray bar); estimates of the change in average sentiment of the public's direct engagement with arrested elites before arrest and after release based on t-tests with 95% confidence intervals (right panel).

the masses suggests physical repression was ineffective in demonstrating the bounds of socially acceptable behavior.

These results also indicate that the constraint of elite behavior had very limited downstream effects in the broader Saudi Twittersphere. The repression of elites did not make them less influential or decrease public engagement with them, as the rates at which arrested actors were retweeted remained elevated both one month and one year following their arrests. Only when we explicitly examine retweets and mentions—content showing direct public engagement with the elites—do we see content becoming less critical over time. This is largely due to the fact that retweets by design include the text of the tweets produced by the elites, and because elite behavior is constrained, this content is constrained as well. However, among our sample of users, retweets and mentions of our arrested elites make up a tiny fraction of their tweets about politics and society.

While pinpointing exactly why these short-term arrests do not act as indirect deterrents and have limited downstream effects is beyond the scope of this paper, it could be related to the large volume of online dissenters that reduces the risk to any individual, to these arrests not changing the level of ambiguity around the possibility of repression, or to the networked nature of social media in facilitating the emergence of other leaders of dissent.

5.3 Unchanged Behavior Among Similar Elites

Individuals most at risk of repression are those who are most similar to the arrested elites—those who have engaged in similar dissent in the past and those who also have large online followings. However, even when we focus on this population, we do not find evidence of demobilization. Figure 10 shows that unlike the arrested elites, similar elites do not decrease their volume of tweets. There is little change in the daily volume of tweets

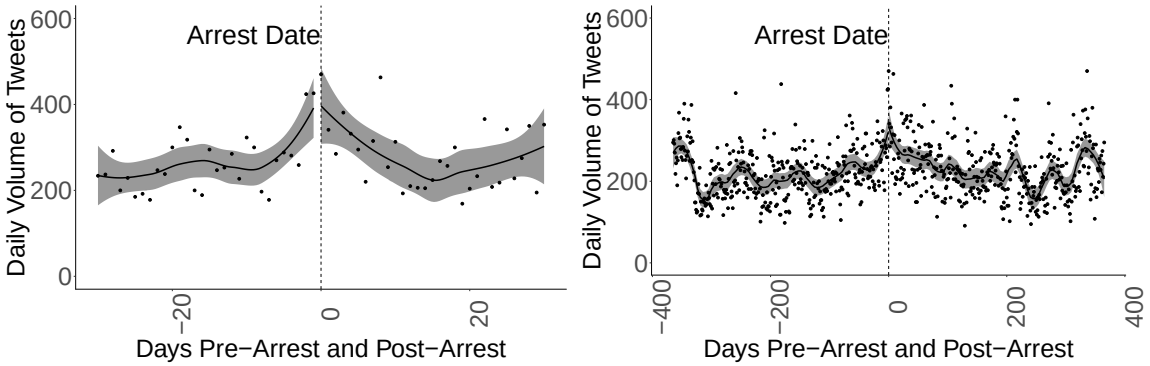


Figure 10: Pre-arrest and post-release trends plotted as local regression lines with loess smoothing and 95% confidence intervals based on the daily volume of tweets produced by similar elites who were not arrested in the month before arrest and month after release (left panel), and in the year before arrest and year after release (right panel); regression tables of these results found in Table A7 and A8 of the Appendix.

produced by the non-arrested elites in the month before and month after the arrests, or in the year before and year after arrests. These results suggests that the chilling effect of arrests on those who are directly targeted does not extend to similar non-arrested elites, who remain equally active on Twitter in the post-arrest period. Similarly, non-arrested elites do not change the content of their tweets. Figure 11 shows that non-arrested elites continue to express negative sentiment toward the regime, and Figure 12 shows that discussions of offline collective action remain unchanged.[31]

Qualitative evidence also suggests that the behavior of similar non-arrested elites was not constrained by arrests. In this period, activists and clerics frequently denounced the arrests of their friends and colleagues and did not appear deterred by their arrests. For example, non-arrested women's rights activist Hala al-Dosari spoke out against the arrests

[31] We do not code more tweets one year out for the non-arrested elites because the chilling effect on the arrested elites diminishes over time, and since we do not observe any change one month post-arrest, we are unlikely to observe any further away from the arrests.

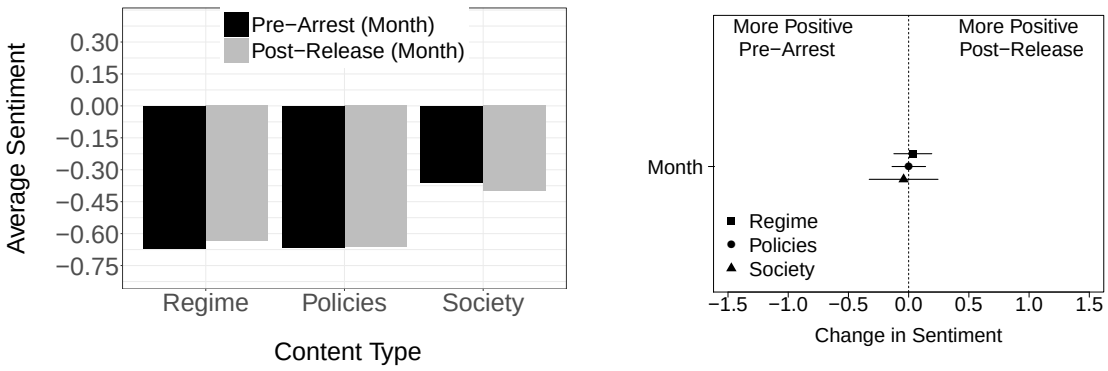


Figure 11: Barplots of the average sentiment of tweets of similar, non-arrested elites in the month pre-arrest (black bar) and the month post-release (light gray bar); estimates of the change in average sentiment of tweets of similar, non-arrested elites before arrests and after release based on t-tests with 95% confidence intervals (right panel).

Figure 12: Barplots of the average proportion of similar, non-arrested elite tweets calling for collective in the month pre-arrest (black bar) and the month post-release (light gray bar); estimates of the change in proportion of tweets calling for collective action before arrest and after release based on t-tests with 95% confidence intervals (right panel).

of Loujain al-Hathloul and Mayasa al-Amoudi in 2014 (BBC News 2014). Similarly, non-arrested clerics denounced the arrests of Mohamad al-Arefe and Mohsen al-Awaji in 2013 (Admoun 2013) and non-arrested human rights activists have protested the arrests of members of the Saudi Civil and Political Rights Association (The Daily Star 2011).

One reason we may not observe demobilization of non-arrested elites is that arrests did not change the already high level of ambiguity surrounding repression in the Saudi Kingdom. Because well-known elites are at risk of arbitrary arrest at all times, the arrest of other elites may not constrain their behavior any more than the daily reality of living under such repressive conditions. This is especially true since arrested elites were well-known figures, and the reason for their arrests was generally known. Another reason

for the lack of demobilization of non-arrested elites may related to their ability to leave Saudi Arabia and escape repression in this period. For example, non-arrested women's rights activist Hala al-Dosari currently has an academic position at Harvard University, and human rights defender Waleed Sulais was recently forced into exile.

6 Conclusions

By dissaggregating the effect of arrests on online expression by actor, behavior, and time, our results provide new insights into the relationship between physical repression and online dissent, the puzzle of repression and dissent more broadly, and how information is controlled in the digital age. Analyzing over 300 million tweets and Google search data between 2010 and 2017, this paper offers new real-time measures of the direct and indirect effects of repression on online dissent. Furthermore, by allowing us to capture both the volume and content of mass and elite messages on the same platform, Twitter data provides novel perspectives on how diverse actors behave in the aftermath of physical repression.

Our results suggest that physical repression was largely unsuccessful in constraining online expression, given that the goal of the Saudi regime was likely not just to rein in the behavior of particular elite actors but rather to reduce dissent overall. In particular, we find that while repression demobilized arrested elite actors in both the short and long term—causing them to tweet less and to produce less dissenting content—the indirect effect of repression on the behavior of other actors was more varied. The active followers of arrested elites were mobilized in the short and long term both regarding their level of interest and engagement with arrested elites. Moreover, the content of the tweets produced by these actively engaged followers was not constrained, and became more critical of the Saudi regime in the post-arrest period. Furthermore, similar elites who were not arrested exhibited no change in behavior with regard to either the volume or content of their tweets, suggesting that the chilling effects of repression did not extend to other elite actors who had also voiced dissent prior to these arrests.

Why would the Saudi regime utilize an ineffective strategy? Governments may go

through a learning process in how to control new technologies, and perhaps these targeted arrests were one phase in that learning. Government may also default to a particular style of repression depending on who is in power. What we know is that Saudi Arabia's use of physical repression has shifted since our period of analysis. Since 2017, the Saudi Kingdom has moved away from targeted arrests to more indiscriminate forms of physical repression. These include larger-scale arrests, such as the late 2017 "purge" of about 500 business people, princes, government ministers, and activists, more death sentences such as as that of popular cleric Salman al-Oudah, and even murder of opponents living abroad such as the recent murder of influential journalist Jamal Kashoggi (Rauhala 2018; Freedom House 2018; Human Rights Watch 2018*b*). Our work suggests that—as of 2017—despite the threat of repression, many elite and non-elite actors continued to take advantage of Twitter as one of the few avenues of political expression available in the Saudi Kingdom. Future research should examine the extent to which this pattern persists under current conditions of higher intensity repression.

While our study is focused on dissent in the Saudi Twittersphere, given the increasing use of physical repression to combat online opposition in authoritarian and democratic regimes worldwide, these findings may have important implications for the study of repression and online dissent in other contexts. We hope the analytical leverage gained by disaggreating the effect of repression on online dissent by type, actor, behavior, and time can be used in future studies examining other regions, regime types, other forms of physical repression, and other forms of online dissent.

References

Admoun, Y. 2013. "Saudi Authorities Irate Over Communique Condemning Ouster Of Former Egyptian President Mursi." *Middle East Media Research Institute* .

URL: *https://www.memri.org/reports/saudi-authorities-irate-over-communique-condemning-ouster-former-egyptian-president-mursi*

Al-Arabiya. 2015. "41 percent of Saudis have Twitter accounts: study." *Al-Arabiya* .

URL: *http://english.alarabiya.net/en/media/digital/2015/03/11/41-percent-of-Saudis-have-Twitter-accounts-study-.html*

al Rasheed, Nayaf. 2013. "Twitter Out of Our Control." *Al-Watan* .

URL: *http://www.alwatan.com.sa/Nation/NewsDetail.aspx?ArticleID=133168CategoryID=3*

Alabaster, Olivia. 2018. "Saudi Arabia using anti-terror laws to detain and torture political dissidents, UN says." *The Independent* .

Alferaehy, Rawan. 2016. My Right to Drive: Women's Rights in Saudi Arabia Explored through the Lens of Religion PhD thesis The George Washington University.

Amnesty International. 2016. "Jail Sentances of Three Lawyers Overturned." *Amnesty International* .

URL: *https://www.amnesty.org/download/Documents/MDE2333402016ENGLISH.pdf*

Bahry, Donna and Brian D Silver. 1987. "Intimidation and the Symbolic Uses of Terror in the USSR." *American Political Science Review* 81(4):1064–1098.

Bamman, David, Brendan O'Connor and Noah Smith. 2012. "Censorship and deletion practices in Chinese social media." *First Monday* 17(3-5).

BBC News. 2014. "Saudi terrorism court 'to try women drivers'." *BBC News* .

URL: *https://www.bbc.com/news/world-middle-east-30602155*

Blaydes, Lisa. 2018. *State of Repression: Iraq Under Saddam Hussein*. Princeton University Press.

Calamur, Krishnadev. 2018. "Saudi Arabia Rejects Human-Rights Criticism, Then Crucifies Someone." *The Atlantic* .

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

Cheng, Justin, Lada A Adamic, Jon M Kleinberg and Jure Leskovec. 2016. Do cascades recur? In *Proceedings of the 25th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee pp. 671–681.

Conti, Gregory and Edward Sobiesk. 2007. An honest man has nothing to fear: user perceptions on web-based information disclosure. In *Proceedings of the 3rd symposium on Usable privacy and security*. ACM pp. 112–121.

Cordeiro, Mário and João Gama. 2016. Online social networks event detection: a survey. In *Solving Large Scale Learning Tasks. Challenges and Algorithms*. Springer pp. 1–41.

Davenport, Christian. 2005. "Repression and mobilization: Insights from political science and sociology." *Repression and mobilization* pp. vii–xli.

Davenport, Christian and Cyanne Loyle. 2012. "The states must be crazy: Dissent and the puzzle of repressive persistence." *International Journal of Conflict and Violence (IJCV)* 6(1):75–95.

Durkheim, Emile. 1984. *The Division of Labor in Society*. Free Press. [Translated by W.D. Halls.].

Egorov, Georgy, Sergei Guriev and Konstantin Sonin. 2009. "Why Resource-poor Dictators Allow Freer Media: A Theory and Evidence from Panel Data." *American Political Science Review* 103(4):645–668.

ESHR. 2017. "Death Penalty 2017 Annual Report." https://bit.ly/2MApAbW.

Freedom House. 2017. "Freedom on the Net." *Freedom House* .
URL: *https://freedomhouse.org/report/table-country-scores-fotn-2017*

Freedom House. 2018. "Saudi Arabia: Thousands Held Arbitrarily Dramatic Increase in Detention Without Trial." *Freedom House* .
URL: *https://freedomhouse.org/report/freedom-world/2017/saudi-arabia*

Friedrich, Carl J and Zbigniew K Brzezinski. 1965. "Totalitarian dictatorship." *Cambridge, MA: Harvard UP* .

Gibney, Mark, Linda Cornett, Reed Wood, Peter Haschke and Daniel Arnon. 2016. "The Political Terror Scale 1976–2015. Political Terror Scale website.".

Glum, Julia. 2015. "Saudi Arabia's Youth Unemployment Problem Among King

Salman's Many New Challenges After Abdullah's Death." *International Business Times* .

Gunitsky, Seva. 2015. "Corrupting the cyber-commons: Social media as a tool of autocratic stability." *Perspectives on Politics* 13(01):42–54.

Guriev, Sergei and Daniel Treisman. 2015. How modern dictators survive: An informational theory of the new authoritarianism. Technical report National Bureau of Economic Research.

Gurr, Ted Robert. 1988. "War, revolution, and the growth of the coercive state." *Comparative Political Studies* 21(1):45–65.

Hardin, Russell. 1982. *Collective action*. Johns Hopkins University Press.

Hassanpour, Navid. 2014. "Media Disruption and Revolutionary Unrest: Evidence From Mubarak's Quasi-Experiment." *Political Communication* 31(1):1–24.

Hess, Steve. 2013. "From the Arab Spring to the Chinese Winter: The institutional sources of authoritarian vulnerability and resilience in Egypt, Tunisia, and China." *International Political Science Review* 34(3):254–272.

Hobbs, William R and Margaret E Roberts. 2018. "How sudden censorship can increase access to information." *American Political Science Review* pp. 1–16.

Howard, Philip N and Muzammil M Hussain. 2011. "The role of digital media." *Journal of democracy* 22(3):35–48.

Human Rights Watch. 2014. "Saudi Arabia: Assault on Online Expression." *Human Rights Watch* .

URL: *http://www.hrw.org/news/2014/11/22/saudi-arabia-assault-online-expression*

Human Rights Watch. 2018*a*. "Human Rights Watch World Report 2017: Saudi Arabia." *Human Rights Watch* .

URL: *https://www.hrw.org/world-report/2017/country-chapters/saudi-arabia*

Human Rights Watch. 2018*b*. "Saudi Arabia: Thousands Held Arbitrarily Dramatic Increase in Detention Without Trial." *Human Rights Watch* .

URL: *https://www.hrw.org/news/2018/05/06/saudi-arabia-thousands-held-arbitrarily*

Ibahrine, Mohammed. 2016. "The dynamics of the Saudi Twitterverse." *Political Islam*

and Global Media: The boundaries of religious identity p. 203.

Jansen, Sue Curry and Brian Martin. 2015. "The Streisand effect and censorship backfire." *International Journal of Communication* 9:656–671.

Jenkins, J Craig and Charles Perrow. 1977. "Insurgency of the powerless: Farm worker movements (1946-1972)." *American sociological review* pp. 249–268.

King, Gary, Jennifer Pan and Margaret E Roberts. 2013. "How Censorship in China Allows Government Criticism but Silences Collective Expression." *American Political Science Review* 107(2):1–18.

King, Gary, Jennifer Pan and Margaret E Roberts. 2014. "Reverse-engineering censorship in China: Randomized experimentation and participant observation." *Science* 345(6199):1–10.

King, Gary, Jennifer Pan and Margaret E Roberts. 2017. "How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, not Engaged Argument." *American Political Science Review* 111(3):484–501.

Krieg, Andreas. 2016. Gulf security policy after the Arab Spring: considering changing security dynamics. In *The Small Gulf States*. Routledge pp. 57–73.

Kuran, Timur. 1997. *Private truths, public lies: The social consequences of preference falsification*. Harvard University Press.

Lichbach, Mark Irving. 1987. "Deterrence or escalation? The puzzle of aggregate studies of repression and dissent." *Journal of Conflict Resolution* 31(2):266–297.

Link, Perry. 2002. "The anaconda in the chandelier: Chinese censorship today." *New York Rev. of Books, April* 11.

Lopez Bernal, J, S Cummins and A Gasparrini. 2016. "Interrupted time series regression for the evaluation of public health interventions: a tutorial [published online ahead of print June 9, 2016]." *International Journal of Epidemiology* .

MacKinnon, Rebecca. 2009. "China's Censorship 2.0: How companies censor bloggers." *First Monday* 14(2).

Makkah Newspaper. 2015. "[10 proposals to protect Saudi Arabia from Twitter]." *Makkah Newspaper* .

URL: *http://www.makkahnewspaper.com/makkahNews/loacal/122727.html.VQsfK2SUfXo*

Malesky, Edmund and Paul Schuler. 2011. "The Single-Party Dictator's Dilemma: Information in Elections without Opposition." *Legislative Studies Quarterly* 36(4):491–530.

Ménoret, Pascal. 2016. "Repression and Protest in Saudi Arabia." *Middle East Brief n* 101.

Morozov, Evgeny. 2012. *The Net Delusion: the Dark Side of Internet Freedom*. Public Affairs.

Munger, Kevin, Richard Bonneau, Jonathan Nagler and Joshua A Tucker. 2018. "Elites Tweet to Get Feet Off the Streets: Measuring Regime Social Media Strategies During Protest." *Political Science Research and Methods* pp. 1–20.

Nabi, Zubair. 2014. "Censorship is futile." *arXiv preprint arXiv:1411.0225* .

Noman, Helmi, Robert Faris and John Kelly. 2015. "Openness and Restraint: Structure, Discourse, and Contention in Saudi Twitter.".

Oberschall, Anthony. 1973. *Social conflict and social movements*. Prentice-Hall Englewood Cliffs, NJ.

Olson, Mancur. 1965. *The Logic of Collective Action: Public Goods and the Theory of Groups*. Harvard University Press.

Pan, Jennifer. 2016. "How Chinese Officials Use the Internet to Construct their Public Image." *Political Science Research and Methods* Forthcoming.

Pan, Jennifer. 2017. "How market dynamics of domestic and foreign social media firms shape strategies of internet censorship." *Problems of Post-Communism* 64(3-4):167–188.

Pan, Jennifer and Kaiping Chen. 2018. "Concealing Corruption: How Chinese Officials Distort Upward Reporting of Online Grievances." *American Political Science Review* Forthcoming.

Pearce, Katy E and Sarah Kendzior. 2012. "Networked authoritarianism and social media in Azerbaijan." *Journal of Communication* 62(2):283–298.

Qin, Bei, David Strömberg and Yanhui Wu. 2015. "The Political Economy of Social Media in China." Working Paper.

Qin, Bei, David Stromberg and Yanhui Wu. 2017. "Why Does China Allow Freer Social Media? Protests versus Surveillance and Propaganda." *CEPR working paper* DP11778.

Rasler, Karen. 1996. "Concessions, repression, and political protest in the Iranian revolution." *American Sociological Review* pp. 132–152.

Rauhala, Emily. 2018. "Saudi Arabia's spat with Canada was a lesson. Trump ignored it." *The Washington Post* .

Report, Twitter Transparancy. 2015. "Saudi Arabia: Information Requests." *Twitter Transparancy Report* .

URL: *https://transparency.twitter.com/country/sa*

Roberts, Margaret E. 2018. *Censored: Distraction and Diversion Inside Chinas Great Firewall*. Princeton University Press.

Siegel, Alexandra. 2015. *Sectarian Twitter Wars: Sunni-Shia Conflict and Cooperation in the Digital Age*. Vol. 20 Carnegie Endowment for International Peace.

Steinert-Threlkeld, Zachary C. 2017. "Spontaneous collective action: peripheral mobilization during the Arab spring." *American Political Science Review* 111(2):379–403.

Stephens-Davidowitz, Seth. 2014. "The cost of racial animus on a black candidate: Evidence using Google search data." *Journal of Public Economics* 118:26–40.

Stephens-Davidowitz, Seth. 2017. *Everybody lies: big data, new data, and what the internet can tell Us about who we really are*. HarperCollins New York.

Stern, Rachel E and Jonathan Hassid. 2012. "Amplifying silence: uncertainty and control parables in contemporary China." *Comparative Political Studies* 45(10):1230–1254.

The Daily Star. 2011. "Saudi group raps authorities over recent arrests." *The Daily Star* .

URL: *http://www.dailystar.com.lb/ArticlePrint.aspx?id=135177mode=print*

Tilly, Charles. 1978. "Collective violence in European perspective.".

Walter, Eugene Victor. 1969. *Terror and resistance: a study of political violence, with case studies of some primitive African communities*. Vol. 1 Oxford University Press New York.

Way, Lucan A and Steven Levitsky. 2006. "The dynamics of autocratic coercion after the Cold War." *Communist and Post-Communist Studies* 39(3):387–410.

Worth, Robert. 2012. "Twitter gives Saudi Arabia a revolution of its own." *The New York Times* 20.

Zhao, Wayne Xin, Baihan Shu, Jing Jiang, Yang Song, Hongfei Yan and Xiaoming Li. 2012. Identifying event-related bursts via social media activities. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*. Association for Computational Linguistics pp. 1466–1477.

Zhu, Tao, David Phipps, Adam Pridgen, Jedidiah R. Crandall and Dan S. Wallach. 2013. The Velocity of Censorship: High-fidelity Detection of Microblog Post Deletions. In *Proceedings of the 22Nd USENIX Conference on Security*. Berkeley, CA, USA: USENIX Association pp. 227–240.

Appendix

A Descriptive Data

Table A1: Arrested Elite Actors

	Name	Type	Official Arrest Reason	Unofficial Arrest Reason
1	Bandar al-Nogaithan	Lawyer	Disobeying ruler / Slandering judiciary	Tweets critical of judiciary
2	Abdulrahman al-Subaihi	Lawyer	Disobeying ruler / Slandering judiciary	Tweets critical of judiciary
3	Abdulrahman Al-Rumaih	Lawyer	Disobeying ruler / Slandering judiciary	Tweets critical of judiciary
4	Raif Badawi	Liberal Activist (Religion)	Apostosy /Insulting Islam / Violating Anti-Cyber Crime Law	Comments on his website debating political and rel
5	Omar al-Saeed	Liberal Activist (Human Rights)	Harming public order / Setting up unliscensed organization	Calling for Democracy/Criticizing Saudi HR Recor
6	Abdullah al Hamid	Liberal Activist (Human Rights)	Sowing Discord and Chaos/Violating Public Safety	Calling for Prison reform / resignation of Interior M
7	Issa al-Nukheifi	Liberal Activist (Human Rights)	Disobedience to the ruler/ Violating cybercrimes law	Accused authorities of corruption / human rights vi
8	Abdulaziz al-Hussan	Liberal Activist (Human Rights)	Providing inaccurate information about the government	Representing arrested lawyers / tweeting about thei
9	Mohammed al-Bajady	Liberal Activist (Human Rights)	Establishing HR Org/ Distorting state's reputation / Impugning judicial independence / instigating relatives of detainees to protest / Possessing censored books	Organized protest against arbitrary detention
10	Abdulkarim Al-Khoder	Liberal Activist (Human Rights)	Disobeying the ruler / Inciting disorder / Harming the image of the state / Founding an unlicensed organization	Crackdown on Saudi Civil and Political Rights Ass
11	Fowzan al-Harbi	Liberal Activist (Human Rights)	Inciting disobedience to the ruler / Describing KSA as a 'police state'	Crackdown on Saudi Civil and Political Rights Ass
12	Khaled al-Johani	Liberal Activist (Human Rights)	Being present at a prohibited demonstration/ Distorting the kingdom's reputation/ Contact with known Saudi dissident abroad	Participated in 'Day of Rage' and spoke to internati
13	Mohammad Fahad al-Qahtani	Liberal Activist (Human Rights)	Sowing discord / Disturbing public order / Breaking allegiance with the ruler	Calling for Prison reform / resignation of Interior M
14	Suliman al-Rashoodi	Liberal Activist (Human Rights)	Breaking Allegiance with Ruler / Attempting to distort reputation of kingdom	Arrested for Publication 'The Legitimacy of Demo
15	Saleh al-Ashwan	Liberal Activist (Human Rights)	Breaking allegiance to and disobeying the ruler/ Questioning the integrity of officials/ Member of unliscensed organization	Crackdown on Saudi Civil and Political Rights Ass
16	Waleed Abul Khair	Liberal Activist (Human Rights)	Disobeying the ruler and seeking to remove his legitimacy/ Insulting the judiciary and questioning the integrity of judges / Setting up an unlicensed organization / Harming the reputation of the state	oners in Iraq Establishing human rights organization / criticizing
17	Zuhair Kutbi	Liberal Activist (Human Rights)	Sowing discord/ Inciting public opinion /Reducing the government's prestige	Calling for Constitutional Monarchy / Combattting
18	Alaa Brinji	Liberal Activist (Religion)	Insulting rulers / Inciting public opinion	Critical tweets about imprisonment of human right the driving ban
19	Hamza Kashagri	Liberal Activist (Religion)	Apostosy/ Crossing red lines / Denigrating religious beliefs in God and His Prophet	Popular calls for his death online following tweets l
20	Turki al-Hamad	Liberal Activist (Religion)	No public charges	Tweets criticizing Saudi interpretation of Islam
21	Hassan Farhan Al-Malki	Moderate Cleric	Supporting proximity among Islamic sects	Defending Shia rights surrounding Nimr al-Nimr's
22	Abdulaziz al-Tarifi	Sahwa Cleric	Calling for Constitutional Monarchy	Tweet criticizing monarchy for religious police refo
23	Mohammad al-Arefe	Sahwa Cleric	No public charges	Supporting Morsi / Muslim Brotherhood / Criticizi
24	Mohsen al-Awaji	Sahwa Cleric	No public charges	Supporting Morsi /Muslim Brotherhood/ Signing C (second arrest)
25	Ibrahim al-Sakran	Salafi Cleric	Damaging fabric of society / Inciting public opinon / Interefering in international affairs	Tweets criticizing foreign policy in Yemen / treatm
26	Adel Ali al-Labbad	Shia Activist	Disobedience to Ruler/Disturbing Public Order	Poems criticizing arrests / treatment of dissidents
27	Mohamed Baqir al-Nimr	Shia Activist	No public charges	Tweeting about Nimr al Nimr's trial
28	Ahmed al-Musheikhis	Shia Activist	No public charges	Protesting Detentions / Advocating Shia Rights / B tion
29	Nimr al-Nimr	Shia Cleric	Disturbing security /Seeking Foreign Meddling /Terrorism	Giving anti-regime speeches/ Defending political p following
30	Tawfiq al-Amer	Shia Cleric	Defaming ruling system /Ridiculing religious leaders/ Inciting sectarianism/ Calling for change/ Disobeying the ruler	Criticizing treatment of Shia / Calling for reforms
31	Sahar Al-Khashrami	Anti-University Corruption	Defamation / Violating Anti-Cyber Crime Law	Hashtag campaign condemning academic fraud, fo
32	Lujain al-Hathloul	Women's Rights Activist	Tried under vague provisions of anti-cybercrime law	Comments on social media calling for end to drivin
33	Manal al-Sharif	Women's Rights Activist	Disturbing public order / Inciting Public Opinion	Social media campaigns calling for protests / filmin
34	Mayasa al-Amoudi	Women's Rights Activist	Tried under vague provisions of anti-cybercrime law	Comments on social media calling for end to drivin
35	Samar Badawi	Women's Rights Activist	No public charges	Women's Driving Campaign / Managing jailed husl
36	Souad al-Shammari	Women's Rights Activist	Insulting Islam / Inciting rebellion	Women's Driving Campaign / Critcizing Guardians

Table A2: Elite Arrest Dates (First Arrest)

	Name	Type	First Arrest Date	First Release Date
1	Bandar al-Nogaithan	Lawyer	10/27/14	4/15/15
2	Abdulrahman al-Subaihi	Lawyer	10/27/14	5/15/15
3	Abdulrahman Al-Rumaih	Lawyer	10/27/14	4/15/15
4	Raif Badawi	Liberal Activist (Religion)	6/17/12	not released
5	Omar al-Saeed	Liberal Activist (Human Rights)	4/30/13	12/24/15
6	Abdullah al Hamid	Liberal Activist (Human Rights)	9/2/12	not released
7	Issa al-Nukheifi	Liberal Activist (Human Rights)	9/1/12	4/6/16
8	Abdulaziz al-Hussan	Liberal Activist (Human Rights)	3/11/13	3/12/13
9	Mohammed al-Bajady	Liberal Activist (Human Rights)	3/21/11	8/6/13
10	Abdulkarim Al-Khoder	Liberal Activist (Human Rights)	6/28/13	not released
11	Fowzan al-Harbi	Liberal Activist (Human Rights)	12/26/13	6/24/14
12	Khaled al-Johani	Liberal Activist (Human Rights)	3/1/11	8/6/12
13	Mohammad Fahad al-Qahtani	Liberal Activist (Human Rights)	3/9/13	not released
14	Suliman al-Rashoodi	Liberal Activist (Human Rights)	12/12/12	12/12/17
15	Saleh al-Ashwan	Liberal Activist (Human Rights)	7/7/12	not released
16	Waleed Abul Khair	Liberal Activist (Human Rights)	4/15/14	not released
17	Zuhair Kutbi	Liberal Activist (Human Rights)	7/15/15	not released
18	Alaa Brinji	Liberal Activist (Religion)	5/12/14	not released
19	Hamza Kashagri	Liberal Activist (Religion)	2/7/12	10/29/13
20	Turki al-Hamad	Liberal Activist (Religion)	12/24/12	6/5/13
21	Hassan Farhan Al-Malki	Moderate Cleric	10/14/14	12/24/14
22	Abdulaziz al-Tarifi	Sahwa Cleric	4/25/16	8/25/17
23	Mohammad al-Arefe	Sahwa Cleric	7/20/13	7/22/13
24	Mohsen al-Awaji	Sahwa Cleric	7/20/13	7/22/13
25	Ibrahim al-Sakran	Salafi Cleric	6/14/16	not released
26	Adel Ali al-Labbad	Shia Activist	10/10/12	not released
27	Mohamed Baqir al-Nimr	Shia Activist	10/15/14	11/1/14
28	Ahmed al-Musheikhis	Shia Activist	1/5/17	2/1/17
29	Nimr al-Nimr	Shia Cleric	7/8/12	executed 1/2/2016
30	Tawfiq al-Amer	Shia Cleric	2/27/11	3/6/11
31	Sahar Al-Khashrami	University Corruption	4/15/15	4/15/15
32	Lujain al-Hathloul	Women's Rights Activist	12/2/14	2/3/15
33	Manal al-Sharif	Women's Rights Activist	5/21/11	5/30/11
34	Mayasa al-Amoudi	Women's Rights Activist	12/2/14	2/3/15
35	Samar Badawi	Women's Rights Activist	1/1/16	1/13/16
36	Souad al-Shammari	Women's Rights Activist	10/28/14	1/28/15

Table A3: Arrested Elites and Non-Arrested "Match" Elites

	Name	Twitter Handle	Arrested	Match	Type	Followers Count	Tweet Count
1	Omar al-Saeed	181Umar	arrested	Abdullah al-Nasri	Liberal Activist (Human Rights)	2497	2947
2	Abdulaziz al-Tarifi	abdulaziztarefe	arrested	Suhail bin Mualla al-Mutairi	Sahwa Cleric	1029931	11023
3	Suhail bin Mualla al-Mutairi	aborazan2011	match		Sunni Cleric	126755	22824
4	Abdullah al Hamid	Abubelal_1951	arrested	Abdullah al-Nasri	Liberal Activist (Human Rights)	83609	10253
5	Adel Ali al-Labbad	adel_lobad	arrested	Saeed Abbas	Shia Activist	7140	196
6	Issa al-Nukheifi	aesa_al_nukhifi	arrested	Mujtahidd	Liberal Activist (Human Rights)	26549	39280
7	Abdulaziz al-Hussan	Ahussan	arrested	Abdullah al-Nasri	Liberal Activist (Human Rights)	41966	14987
8	Mohammed al-Bajady	albgadi	arrested	Waleed Sulais	Liberal Activist (Human Rights)	23482	666
9	Alaa Brinji	albrinji	arrested	Waleed Sulais	Liberal Activist (Religion)	1448	2633
10	Abdullah al-Nasri	alnasri1	match		Lawyer	18307	30199
11	Abbas Said	alsaeedabbas	match		Shia Cleric	11892	894
12	Abdulrahman al-Subaihi	Alsubaihiabdul	arrested	Abdullah al-Nasri	Lawyer (anti-regime)	38067	40159
13	Abdullah Rahman al-Sudais	assdais	match		Sunni Cleric	315890	21191
14	Abdulkarim Al-Khoder	drkhdar	arrested	Waleed Sulais	Liberal Activist (Human Rights)	38400	8006
15	Sadeq al-Jibran	DrSadeqMohamed	match		Lawyer	16913	5894
16	Fowzan al-Harbi	fowzanm	arrested	Waleed Sulais	Liberal Activist (Human Rights)	2611	1032
17	Hala al-Dosari	Hala_Aldosari	match		Women's Rights Activist	57010	31825
18	Hamza Kashagri	Hmzmz	arrested	Rashad Hassan	Liberal Activist (Religion)	18917	2714
19	Hassan Farhan Al-Malki	HsnFrhanALmalki	arrested	Abdullah Rahman al-Sudais	Moderate Cleric	309260	90053
20	Ibrahim al-Sakran	iosakran	arrested	Abdullah Rahman al-Sudais	Salafi Cleric	237899	3109
21	Khaled al-Johani	KhaledLary	arrested	Mujtahidd	Liberal Activist (Human Rights)	4401	1385
22	Abdulrahman Al-Rumaih	LawyerAMRumaih	arrested	Sadeq al-Jibran	Lawyer (anti-regime)	8970	16066
23	Lujain al-Hathloul	LoujainHathloul	arrested	Hala al-Dosari	Women's Rights Activist	307382	6573
24	Mohamad Ali Mahmoud	ma573573	match		Liberal Writer	51322	32662
25	Manal al-Sharif	manal_alsharif	arrested	Hala al-Dosari	Women's Rights Activist	275666	24139
26	Mayasa al-Amoudi	maysaaX	arrested	Hala al-Dosari	Women's Rights Activist	202147	3764
27	Mohamed Baqir al-Nimr	mbanalnemer	arrested	Saeed Abbas	Shia Activist	43483	8851
28	Mohammad Fahad al-Qahtani	MFQahtani	arrested	Waleed Sulais	Liberal Activist (Human Rights)	81054	10287
29	Mohammad al-Arefe	MohamadAlarefe	arrested	Abdullah Rahman al-Sudais	Sahwa Cleric	21325719	32208
30	Mohsen al-Awaji	MohsenAlAwajy	arrested	Yousef Ahmed Qasem	Sahwa Cleric	1624072	3602
31	Ahmed al-Musheikhis	mshikhs	arrested	Saeed Abbas	Shia Activist	2158	1414
32	mujtahid	mujtahidd	match		Liberal Regime Critic	2082608	15476
33	Fawaz al-Ruwaili	Muwafig	match		University Corruption Activist	66359	159704
34	Sahar Al-Khashrami	Profsahar	arrested	Fawaz al-Ruwaili	University Corruption	7580	15126
35	Raif Badawi	raif_badawi	arrested	Wadad Khaled	Liberal Activist	77643	20135
36	Suliman al-Rashoodi	s_alrushodi	arrested	Waleed Sulais	Liberal Activist (Human Rights)	35985	3971
37	Saleh al-Ashwan	saleh_alashwan	arrested	Taha al-Hajji	Liberal Activist (Human Rights)	2943	12098
38	Samar Badawi	samarbadawi15	arrested	Hala al-Dosari	Women's Rights Activist	5134	576
39	Bandar al-Nogaithan	SaudiLawyer	arrested	Sadeq al-Jibran	Lawyer	37174	47355
40	Nimr al-Nimr	ShaikhNemer	arrested	Saeed Abbas	Shia Cleric	15431	17614
41	Tawfiq al-Amer	sk_tawfeeq	arrested	Saeed Abbas	Shia Cleric	550	1226
42	Souad al-Shammari	SouadALshammary	arrested	Wadad Khaled	Women's Rights Activist	246932	49244
43	Taha al-Hajji	tahaalhajji	match		Liberal Activist	7663	16697
44	Turki al-Hamad	TurkiHAlhamad1	arrested	Mohamad Ali Mohamed	Liberal Activist (Religion)	283607	7704
45	Waleed Abul Khair	WaleedAbulkhair	arrested	Waleed Sulais	Liberal Activist (Human Rights)	89906	36405
46	Waleed Sulais	WaleedSulais	match		Liberal Activist	21297	16871
47	Rashad Hassan	watheh1	match		Professor	249871	23739
48	Wadad Khaled	wdadkhaled	match		Liberal Activist	56921	1734
49	Yousef Ahmed Qasem	Yqasem	match		Sahwa Cleric	117982	46024
50	Zuhair Kutbi	zuhairkutbi	arrested	Waleed Sulais	Liberal Activist (Human Rights)	9530	12342

B Interrupted Time Series Analysis

Using Interrupted Time Series Analysis, we model changes in the volume of online behavior as follows:

$$Y_t = \beta_0 + \beta_1(T) + \beta_2(X_t) + \beta_3(X_t T) \tag{1}$$

In Equation 1, Y_t is the number of tweets (or google searches) made at time t, T is the time (number of days) since the elite was arrested, X_t is a dummy variable representing the arrest (for arrested elites the pre-arrest period is coded as 0 and the post-release period is coded as 1[32]), and $X_t T$ is an interaction term. β_0 represents the baseline volume of tweets (or google searches) produced at $t = 0$, β_1 shows the change in the volume of tweets (or google searches) associated with a one unit time increase, representing the underlying daily pre-arrest trend. β_2 captures the immediate effect of the arrest on the volume of tweets (or google searches) produced, or an intercept change, and β_3 captures the slope change in the volume of tweets (or google searches) following the release, relative to the pre-arrest trend. In other words, ITSA is a segmented regression model. Segmented regression simply refers to a model with different intercept and slope coefficients for the pre and post-intervention time periods. It is used to measure the pre-arrest trend, the immediate change in the volume of tweets (or google searches) following the release, as well as the change in the slope of the daily volume of tweets (or google searcehs) in the post-release period. In order to address serial autocorrelation in our data, we use a first order autoregressive (AR1) model in our analysis instead of the standard OLS ITSA model Lopez Bernal, Cummins and Gasparrini (2016).

If repression backfires and is followed by increased online activity, then we should see a positive shift immediately after the release β_2 or a non-negative immediate effect β_2 followed by a positive slope change in the volume of tweets in the post-release period β_3. If repression acts as a deterrent, then we should see a a negative shift immediately after the release β_2 or a non-positive immediate effect β_2 followed by a negative slope change in the volume of tweets in the post-release period β_3.

[32] If elites were not released from prison in the period under study they are excluded from the analysis. The release dates, as well as those elites that were not released, are described in Table A2 in the Appendix.

B.1 Regression Tables

Table A4: Effect of Arrests on Daily Volume of Tweets (Arrested Elites)
One Month Pre-Arrest vs. One Month Post-Release

	Model 1
Baseline	181.312^{***}
	(31.288)
Pre-Arrest Trend	0.017
	(1.751)
Post-Release Level Change	-133.167^{**}
	(40.625)
Post-Release Slope Change	1.389
	(2.611)
AIC	627.284
BIC	639.436
Log Likelihood	-307.642
Num. obs.	60

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$

Table A5: Effect of Arrests on Daily Volume of Tweets (Arrested Elites)
One Year Pre-Arrest vs. One Year Post-Release

	Model 1
Baseline	179.692^{***}
	(11.745)
Pre-Arrest Trend	$-0.094^{\cdot}$
	(0.056)
Post-Release Level Change	-56.150^{***}
	(16.509)
Post-Release Slope Change	0.002
	(0.079)
AIC	8223.851
BIC	8251.376
Log Likelihood	-4105.925
Num. obs.	730

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$

Table A6: Effect of Arrests on Daily Volume of Tweets (Arrested Elites)
One Year Pre-Arrest vs. One Year Post-Release
Disaggregated Models with Elite Fixed Effects

	Month 1	Month (FE) 2	Year	Year (FE)
Baseline	10.587^{***}	6.121	11.255^{***}	2.080
	(2.609)	(4.192)	(1.147)	(1.452)
Pre-Arrest Trend	0.001	0.001	0.001	-0.003
	(0.126)	(0.129)	(0.005)	(0.004)
Post-Release Level Change	-7.030	-8.115^{**}	-3.600^{*}	-3.739^{***}
	(3.714)	(3.002)	(1.663)	(1.097)
Post-Release Slope Change	0.055	0.087	-0.003	-0.002
	(0.182)	(0.197)	(0.008)	(0.005)
Elite Fixed Effects	No	Yes	No	Yes
AIC	8198.052	8038.178	101721.324	100305.740
BIC	8227.414	8145.479	101765.597	100482.793
Log Likelihood	-4093.026	-3997.089	-50854.662	-50128.870
Num. obs.	990	990	11838	11838

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table A7: Effect of Arrests on Daily Volume of Tweets (Non-Arrested Elites)
One Month Pre-Arrest vs. One Month Post-Arrest

	Model 1
Baseline	310.396^{***}
	(33.916)
Pre-Arrest Trend	$3.190^{\cdot}$
	(1.902)
Post-Arrest Level Change	10.129
	(44.843)
Post-Arrest Slope Change	-5.764^{*}
	(2.694)
AIC	671.370
BIC	683.629
Log Likelihood	-329.685
Num. obs.	61

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$

Table A8: Effect of Arrests on Daily Volume of Tweets (Non-Arrested Elites)
One Year Pre-Arrest vs. One Year Post-Arrest

	Model 1
Baseline	226.105^{***}
	(10.980)
Pre-Arrest Trend	0.044
	(0.052)
Post-Arrest Level Change	22.942
	(15.359)
Post-Arrest Slope Change	-0.154^{*}
	(0.074)
AIC	8006.114
BIC	8033.647
Log Likelihood	-3997.057
Num. obs.	731

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$

Table A9: Effect of Arrests on Daily Volume of Mentions of Arrested Elites
One Month Pre-Arrest vs. One Month Post-Arrest

	Model 1
Baseline	36865.670^{***}
	(7590.993)
Pre-Arrest Trend	633.552
	(425.249)
Post-Arrest Level Change	4475.659
	(9838.871)
Post-Arrest Slope Change	-1687.593^{**}
	(610.016)
AIC	1276.251
BIC	1288.509
Log Likelihood	-632.126
Num. obs.	61

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$

Table A10: Effect of Arrests on Daily Volume of Mentions of Arrested Elites
One Year Pre-Arrest vs. One Year Post-Arrest

	Model 1
Baseline	18905.851^{***}
	(1412.542)
Pre-Arrest Trend	6.178
	(6.687)
Post-Arrest Level Change	-1850.453
	(1978.944)
Post-Arrest Slope Change	7.742
	(9.485)
AIC	15160.490
BIC	15188.024
Log Likelihood	-7574.245
Num. obs.	731

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$

Table A11: Effect of Arrests on Daily Search Volume of Arrested Elites
One Month Pre-Arrest vs. One Month Post-Arrest

	Model 1
Baseline	119.299
	(78.644)
Pre-Arrest Trend	2.902
	(4.313)
Post-Arrest Level Change	345.517^{***}
	(65.320)
Post-Arrest Slope Change	-19.479^{**}
	(7.057)
AIC	661.413
BIC	673.671
Log Likelihood	-324.706
Num. obs.	61

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$

Table A12: Effect of Arrests on Weekly Search Volume of Arrested Elites
One Year Pre-Arrest vs. One Year Post-Arrest

	Model 1
Baseline	35.546^{***}
	(4.395)
Pre-Arrest Trend	0.074^{***}
	(0.021)
Post-Arrest Level Change	4.635
	(6.173)
Post-Arrest Slope Change	-0.164^{***}
	(0.029)
AIC	5013.541
BIC	5039.156
Log Likelihood	-2500.771
Num. obs.	532

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$

C Content Analysis

C.1 Crowdflower Coding Scheme

Overview: In this job you will be presented with Arabic language tweets related to society and politics posted by Saudi Arabian Twitter users. You will answer several brief questions about the content of each tweet.

Steps:

- Read each tweet carefully.
- Answer a series of brief questions about the content of each tweet.

1. What attitude does this tweet express about the Saudi monarchy, ruling regime, leaders, religious establishment, or religious doctrine?

- Positive
- Negative
- Neutral
- Irrelevant
- Unclear

2. What attitude does this tweet express about Saudi policies or bureaucracy?

- Positive
- Negative
- Neutral
- Irrelevant
- Unclear

3. What attitude does this tweet express about Saudi society?

- Positive
- Negative
- Neutral
- Irrelevant
- Unclear

4. Is this tweet calling for collective action (social mobilization to achieve a particular goal)?

- Yes
- No
- Unclear

Question 1 Instructions:

- Positive tweets include tweets praising or expressing satisfaction with the Saudi monarchy, ruling regime, leaders, religious establishment, or religious doctrine such as tweets praising specific royal family members or clerics, tweets supporting the legitimacy of the Saudi regime or religious establishment, or tweets praising Saudi Wahabbi religious doctrine.
- Negative tweets include tweets expressing dissatisfaction with or critical of the Saudi monarchy, ruling regime, leaders, religious establishment, or religious doctrine such as tweets criticizing specific royal family members or clerics, tweets calling for democracy or other forms of government, or tweets criticizing Saudi Wahabbi religious doctrine.
- Neutral tweets neither express satisfaction nor dissatisfaction with the Saudi monarchy, ruling regime, leaders, religious establishment, or religious doctrine. These include news articles or factual statements about the regime or religious establishment.

- Irrelevant tweets do not mention the Saudi monarchy, ruling regime, leaders, religious establishment, or religious doctrine.

Question 2 Instructions:

- Positive tweets include tweets praising or expressing satisfaction with the Saudi bureaucracy including the judiciary, the ministry of education, or the religious police. They also include tweets praising or expressing satisfaction with policies and policy outcomes including the state of the economy, corruption, foreign policy, or infrastructure.

- Negative tweets include tweets expressing dissatisfaction with or critical of the Saudi bureaucracy including the judiciary, the ministry of education, or the religious police. They also include tweets criticizing or expressing dissatisfaction with policies and policy outcomes including the state of the economy, corruption, foreign policy, and infrastructure.

- Neutral tweets neither express satisfaction nor dissatisfaction with Saudi policies or bureaucracy. These include news articles or factual statements about policies or bureaucracy.

- Irrelevant tweets do not mention Saudi policies or bureaucracy.

Question 3 Instructions:

- Positive tweets include tweets expressing satisfaction with or praising Saudi society including the role of women, the piety or industriousness of the population, or youth culture.

- Negative tweets include tweets expressing dissatisfaction with or critical of Saudi society, including tweets criticizing Saudi society for being too liberal or conservative or tweets criticizing the role of women in society or youth culture.

- Neutral tweets neither express satisfaction nor dissatisfaction with Saudi society. These include news articles or factual statements about Saudi society.

- Irrelevant tweets do not mention Saudi society.

Question 4 Instructions:

- Tweets calling for collective action (social mobilization to achieve a specific goal) include tweets discussing protest or organized crowd formation.

C.2 Intercoder Agreement

Table A13: Average Intercoder Agreement

	n	mean	sd
policies	16764	0.93	0.15
regime	16764	0.93	0.15
society	16764	0.95	0.13
collective action	16764	0.99	0.05

This table shows average intercoder agreement by category among the three human coders that coded each tweet on Crowdflower.

Figure A1: Distribution of Tweet Content

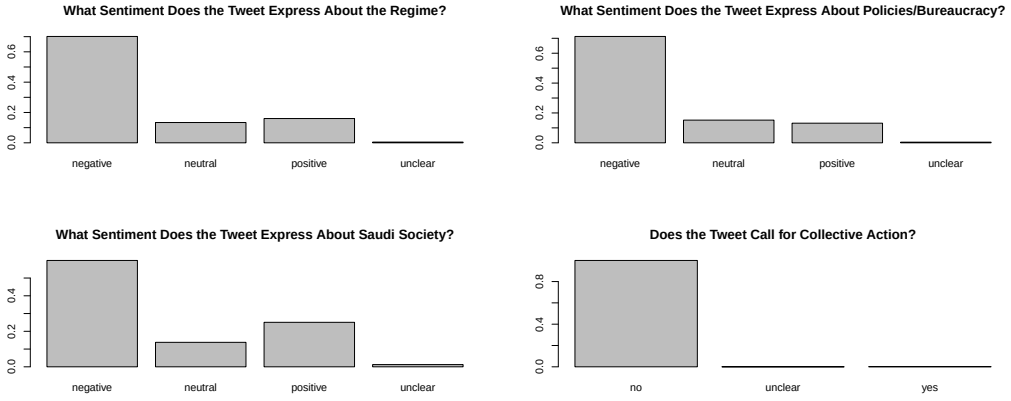
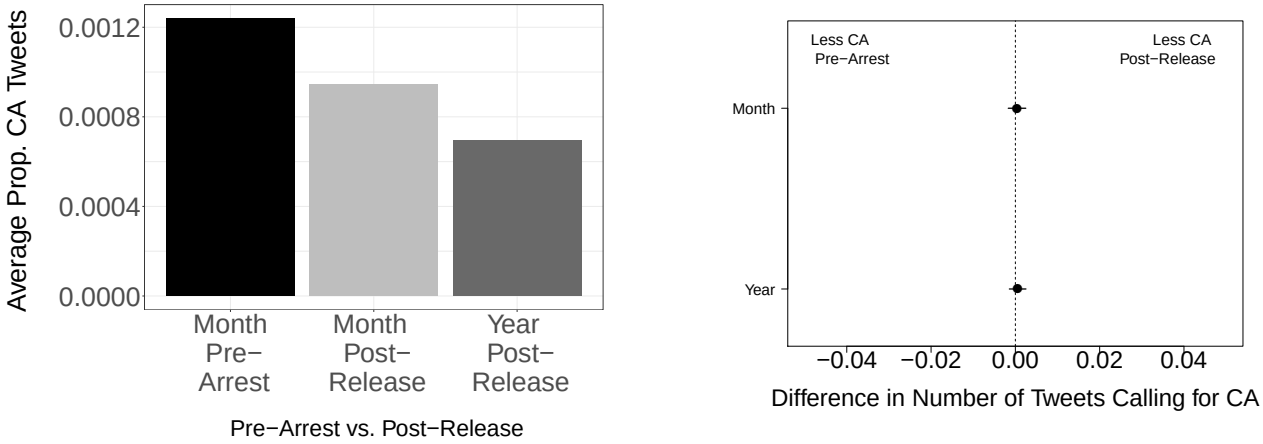


Figure A2: Top Political Keywords in Elite Tweets Relevant to Regime, Policies, or Society

keyword	translation	keyword	translation	keyword	translation
سلمان	King Salman	السجناء	prisoners	السيدات	women
الداخلية	Interior Ministry	منظمة	organization	تخاف	fear
الملك	King	المواطن	citizens	الشيعة	the Shia
نايف	Nayef (Interior Minister)	سياسي	political	سياسيا	politics
الناس	the people	خلف	behind/backwards	موافقة	agreement
السلطة	power	#داعش	#Daesh (ISIS)	النمر	Al-Nimr (Shia cleric)
النساء	women	المواطنين	citizens	الامير	prince
السياسية	politics	الأمير	prince	الشيعي	Shia
النظام	regime	الحاكم	rule	القطاع	sector
التعليم	education	القضاء	judge	حقوق	rights
وزارة	ministry	الحقوق	rights	الطائفية	sectarianism
المرأة	women	هوية	identity	الشرطة	police
الحكم	governance	القرار	decision	الحق	rights
وزير	minister	سياسة	politics	الجيش	army
ولي	crown	oct26driving	oct26driving	الحر	free
الدولة	the state	هدر	waste	إسرائيل	Israel
الشعب	the people	الخاص	private	السياسة	politics
الحكومة	government	القانون	law	الحوثيين	Houthi
women	women	موقع	position	الحرب	war
#مصر	egypt	حرب	war	واضح	clear
العلمية	academic research	مشروع	project	جامعة	university
العدل	justice	مصر	egypt	السرقات	theft
المجتمع	society	الجمعة	university	الجامعات	universities
سجن	prison	السياسي	political	قتل	killed
الفساد	corruption	اهل	people	اليمن	Yemen
قيادة	leadership	الظلم	injustice		
سوريا	Syria	وكيل	representative		
المصري	Egyptian	معالي	his excellency		
الوطن	homeland	#هلكوني	#they_stole_from_me		
ملف	issue	العمل	work		
#سرقوني	#they_stole_from_me	القطيف	Qatif (Shia region)		

D Additional Results

Figure A3: Change in % of Mentions of Arrested Elite Tweets Calling for Collective Action



This figure shows a barplot of the average proportion of tweets calling for collective action and the results of t-tests evaluating the change in the proportion of tweets calling for collective action in tweets directly mentioning or retweeting arrested elites one month before the arrests and one month (and one year) following the releases. Error bars on the left panel shows 95% confidence intervals. Each tweet was coded by three coders on Crowdflower as either containing discussions of collective action or not.