American Political Science Review (2020) 114, 1, 109–125
doi:10.1017/S0003055419000650

How Saudi Crackdowns Fail to Silence Online Dissent

JENNIFER PAN *Stanford University*
ALEXANDRA A. SIEGEL *Stanford University*

Saudi Arabia has imprisoned and tortured activists, religious leaders, and journalists for voicing dissent online. This reflects a growing worldwide trend in the use of physical repression to censor online speech. In this paper, we systematically examine the consequences of imprisoning well-known Saudis for online dissent by analyzing over 300 million tweets as well as detailed Google search data from 2010 to 2017 using automated text analysis and crowd-sourced human evaluation of content. We find that repression deterred imprisoned Saudis from continuing to dissent online. However, it did not suppress dissent overall. Twitter followers of the imprisoned Saudis engaged in more online dissent, including criticizing the ruling family and calling for regime change. Repression drew public attention to arrested Saudis and their causes, and other prominent figures in Saudi Arabia were not deterred by the repression of their peers and continued to dissent online.

INTRODUCTION

On July 20, 2013, Mohamed al-Arefe, a popular Saudi religious leader affiliated with the Sahwa movement, was imprisoned without formal charges. His imprisonment came after he circulated comments and videos to his millions of Twitter and Facebook followers supporting the Muslim Brotherhood, an organization the Saudi regime views as an existential threat (Lacroix 2014). Hundreds of thousands of Saudis identify with the ideas of the Sahwa movement, even though its formal organizations have been repressed and co-opted by the state (Lacroix 2011). Following al-Arefe's arrest, his online supporters were outraged, and the hashtag #FreeMohamadalArefe quickly went viral.

On November 30, 2014, Loujain al-Hathloul live-tweeted her attempt to drive into Saudi Arabia from the United Arab Emirates as part of the #Women2Drive movement. Beginning in 2011, the #Women2Drive social media campaign generated a tide of videos of women defying the Saudi ban on women driving, increasing the domestic and international visibility of protest against these Saudi policies. Al-Hathloul was stopped at the Saudi border, and as we know from her continued tweets, her passport was confiscated and she was kept in her car without water overnight. In the morning, al-Hathloul was told to drive into Saudi Arabia, where she was immediately taken into police custody (Mackey 2014). Al-Hathloul's tweets and the hashtag #FreeLoujain spread rapidly online.

Saudi Arabia is an absolute monarchy and theocracy with great oil wealth. It is also one of the least free and most repressive countries in the world.[1] As a Saudi activist wrote in 2014:

> Saudis live under repression, in fact we breathe repression with the air; it haunts us in our dreams. It is our hell before we encounter hell. Even our appearance, streets, and houses are designed by repression. Repression has shaped the media, religion, security services, universities, and institutions (Al-Rasheed 2016, 119).

In Saudi Arabia, traditional media is tightly controlled. Political dissent is criminalized. Political parties, trade unions, political demonstrations, and strikes are banned. All types of organized opposition are suppressed. Formal social movement organizations are largely absent or short-lived.[2] When a rare on-the-ground protest occurs, it is violently quashed (Ménoret 2016). In the past decade, many activists and reformers have turned to online platforms to sustain their social movements.

Jennifer Pan, Assistant Professor, Department of Communication, Stanford University, jp1@stanford.edu.

Alexandra A. Siegel, Postdoctoral Fellow, Immigration Policy Lab, Stanford University, aasiegel@stanford.edu.

Our thanks to Charles Crabtree, Killian Clarke, Christian Davenport, Martin Dimitrov, Jennifer Earl, Will Hobbs, Holger Kern, Beatriz Magaloni, Elizabeth Nugent, Molly Roberts, Arturas Rozenas, Anton Sobolev, Rory Truex, Lauren Young, Zachary Steinert-Threlkeld, and participants at the 2018 APSA pre-conference on politics and computational social science for their helpful comments and suggestions; to SMaPP Global for making our collaboration possible; to the Stanford King Center on Global Poverty and Development and the National Science Foundation (Award #1647450) for research support. We would also like to thank Twitter for providing us with access to historical data as well as Steve Eglash for facilitating this access. Replication files are available at the American Political Science Review Dataverse: https://doi.org/10.7910/DVN/9AMKHL.

Received: November 5, 2018; revised: May 27, 2019; accepted: September 18, 2019.

[1] Saudi Arabia ranks 201 out of 211 on Freedom House's Freedom in the World index (Freedom House 2018). On the Political Terror Scale (PTS), Saudi Arabia has received a score of four in recent years, which 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" (Gibney et al. 2016). In the most recent year of CIRI Human Rights data (2010), Saudi Arabia is described as a country where torture and disappearances occur frequently, and where many people are imprisoned because of their religious, political, or other beliefs (Cingranelli and Richards 2010). In addition to repression, the Saudi regime also has a long history of selectively co-opting opposition (Wehrey 2015).

[2] Examples of formal organizations include the Shia Organization for the Islamic Revolution in the Arabian Peninsula, which came into existence in 1979 and disappeared around 1993 when its leaders were either arrested or co-opted by the regime (Wehrey 2013); and the human rights-oriented Saudi Civil and Political Rights Association (SCPRA), which formed in 2009 and was disbanded in 2013 after the arrests of its leaders. Example alliances include calls for the establishment of a constitutional monarchy in the late 1990s and early 2000s by Sahwa clerics along with liberals and Shia activists.

They use social media to distribute information about their causes, to organize and facilitate offline protests, to disseminate online petitions, and to organize online campaigns such as the #Women2Drive movement.

The Saudi regime has reacted to this online mobilization by stepping up online censorship—for example, blocking websites promoting Shia rights and those affiliated with the Muslim Brotherhood (Ibahrine 2016; Noman, Faris, and Kelly 2015).[3] But increasingly, the regime has turned to physical repression, which we define following Cingranelli and Richards (2010) as violations of physical integrity rights. Dozens of high-profile individuals connected to diverse social movements in Saudi Arabia have been imprisoned, publicly flogged, and tortured for using social media to criticize the regime and to mobilize support for political and social reforms (Alabaster 2018; Calamur 2018; ESHR 2017; Human Rights Watch 2018a).

Repression of individuals for their online speech is not limited to Saudi Arabia. In 2017, more than thirty countries—including authoritarian regimes such as China, Russia, and Iran, and democracies such as India, Mexico, and Lebanon—used repression to rein in online expression. The most frequent targets were prominent figures with large online followings—who we refer to as "online opinion leaders"—including journalists and dissidents. Political imprisonment and torture were the most common forms of repression, but people in eight countries were executed in 2017 for speaking out about sensitive subjects online (Freedom House 2017).

Despite governments' increased use of repression to control online expression and mobilization, we know very little about what happens when physical repression is used to suppress online dissent. There is a substantial literature on offline repression and its effects on offline mobilization.[4] There is also a great deal of research on how online mobilization reinforces offline mobilization and generates new pathways for dissent.[5] However, research on efforts by governments to quash online mobilization has focused mainly on online censorship and online disinformation campaigns.[6]

We provide the first large-scale, systematic study of the effects of offline repression on online dissent. The form of offline repression we focus on is political imprisonment, and the type of online dissent we study are criticisms of the Saudi regime and calls for political and social reform.[7]

We build on the repression, online mobilization, and censorship literatures to test three mechanisms by which repression might deter dissent: direct deterrence, indirect deterrence, and downstream effects. *Direct deterrence* occurs when individuals who experience repression rein in their behavior for fear that they will be punished again in the future (Jenkins and Perrow 1977; Oberschall 1973; Tilly 1978). *Indirect deterrence* occurs when observers of repression—those who see it but are not directly targeted—change their behavior for fear that they too will be subjected to repression (Durkheim 1984; Walter 1969). Repression might also affect those who do not experience or even observe repression through the *downstream effects* resulting from changes in the behavior of those who are directly repressed. For example, if an online opinion leader becomes supportive rather than critical of the government following repression, this change in sentiment may trickle down and be repeated by their followers. Finally, instead of acting as a deterrent, repression could cause *backlash* and intensify dissent among those targeted or among supporters or bystanders who are mobilized by observing repression (Sullivan and Davenport 2017; Young 2017).

We identify well-known individuals imprisoned by the Saudi regime for online dissent between 2010 and 2017.[8] We collect and analyze over 300 million tweets, as well as daily and weekly Google search data for the same time period. First, these data allow us to disaggregate the effects of repression on different actors and online behaviors. We analyze tweets by imprisoned online opinion leaders, tweets by non-imprisoned online opinion leaders, as well as tweets by individuals who retweeted, mentioned, or replied to the imprisoned leaders on Twitter prior to their arrests, who we refer to as "engaged followers." We also analyze online search behavior of the general public. Second, these data enable us to measure changes in both the *volume* and *substance* of online activity. Third, our data allow us to assess changes in both public expression (tweets) and private interest (Google searches). Finally, our data enable us to examine the immediate consequences of repression as well as its effects up to one year later. We focus on content produced on Twitter because Saudi Arabia has one of the highest levels of Twitter penetration in the world,[9] Saudis frequently

[3] High levels of literacy and internet penetration in Saudi Arabia have propelled social media adoption. To date, the Saudi regime have not imposed wholesale blocks of Twitter or Facebook. In addition, while Saudi Arabia can ask Twitter to remove content, Twitter does not comply in all instances (Noman, Faris, and Kelly 2015; Pan 2017).

[4] For overviews of this literature, see Davenport (2007), Davenport and Inman (2012), and Tilly (2005).

[5] Ayres (1999), Fisher (1998), and Myers (1994) explore how online mobilization reinforces offline mobilization while Bennett and Segerberg (2012), Bimber, Flanagin, and Stohl (2005), Earl and Schussman (2002), Earl and Kimport (2008), Earl et al. (2010), and Tufekci and Wilson (2012) examine how online mobilization creates new pathways for dissent.

[6] Hassanpour (2014), King, Pan, and Roberts (2013, 2014), and Roberts (2018) explore how governments use censorship to deter online dissent, while King, Pan, and Roberts (2017), Munger et al. (2018), Stukal et al. (2017), and Woolley and Howard (2016) examine governments' use of disinformation.

[7] We focus on political imprisonment because it is often a precursor to torture, disappearances, and extrajudicial killings in Saudi Arabia. At times in this paper, we use the term "arrest." We use this term to denote political imprisonment, not arrests as conceptualized in research on policing and arrest in democratic contexts (Davenport, Soule, and Armstrong 2011; Earl 2011; Earl and Soule 2010).

[8] We focus on well-known individuals because we are interested in the broader, public effects of repression (e.g., evidence of indirect deterrence), and repression must be publicly known to have these broader, public effects.

[9] An estimated 41% of the Saudi population uses 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).

discuss politics on Twitter (Noman, Faris, and Kelly 2015), and Twitter's networked structure enables us to examine the behavior of diverse actors on the same platform.

We find that Saudi online opinion leaders who were imprisoned were deterred from dissent. They decreased their online activity, reined in their criticisms of the state, and halted calls for reform after they were released. The altered content of their tweets also had downstream effects on the content of retweets, replies to their tweets, and mentions. However, when we examine the overall Twitter activity of engaged followers of imprisoned leaders, we find that these public cases of political imprisonment generated backlash. Among followers, repression *increased* criticisms of the Saudi monarchy, its religious authority, its institutions, and its policies. Repression also *increased* calls for political and social change, including calls to change the regime from an absolute monarchy to a constitutional monarchy or a democracy. Among other online opinion leaders who were not imprisoned but who had also expressed dissent on social media, repression did not have a deterrent effect. Despite observing the arrests of their peers, these actors continued to tweet and publicly voice support for political and social reform.

By showing the varied effects of repression on online dissent, these results tie into the broader literature on the dissent–repression or conflict–repression nexus (Davenport 2005; Lichbach 1987; Moore 1998).[10] Our findings also reinforce research on the backlash thesis, showing how repression mobilizes dissent,[11] as well as the literature on how censorship can backfire.[12]

Given that we observe backlash in Saudi Arabia, one of the most repressive countries in the world, we expect these findings may extend to other authoritarian contexts. In particular, we think the patterns we observe may be most likely to occur in countries where social movement organizations are largely absent, where repression is public and overt, where the repression leveled at online opinion leaders does not alter observers' calculation about their own risk of repression, and where the state lacks fine-grained control over the online sphere.

[10] The dissent-repression nexus literature shows that although dissent consistently increases state repression, the impact of repression on dissent is highly variable (Goldstone and Tilly 2001). In addition to variation by an individual's level of commitment to a social movement (Davenport, Armstrong, and Zeitzoff 2019; Sullivan and Davenport 2017), effects have been found to vary depending on time frame (Rasler 1996), for violent versus nonviolent forms of dissent (Moore 1998), for different individuals (Opp and Gern 1993), for different organizational categories (Davenport 2015), and for different societal categories (Goldstein 1978).

[11] For examples, see Eckstein (1965), Feierabend, Feierabend, and Gurr (1972), Gurr and Duvall (1973), Kuran (1989), Francisco (1996), Khawaja (1993), Kurzman (1996), Lichbach (1998), and Olivier (1991).

[12] For examples, see Hassanpour (2014), Hobbs and Roberts (2018), Jansen and Martin (2015), Nabi (2014), and Roberts (2018).

DATA AND EMPIRICAL STRATEGY

Since the 1950s, three broad and sometimes overlapping social movements have been present in the Saudi Kingdom: (1) the Sahwa or "Awakening" Sunni Islamist movement with historical ties to the Muslim Brotherhood, (2) a Shia-rights movement calling for rights, protections, and at times secession for the Shia minority in the oil rich Eastern province, and (3) networks of human rights, women's rights, and anti-corruption activists calling for social and political reform.[13] Many of the imprisoned online opinion leaders that we study—and similar non-imprisoned opinion leaders—are broadly part of the same informal networks, connected by family members, friends, and colleagues.[14] All three movements have used social media for mobilization, and from 2010 to 2017, leaders of all three movements—Sahwa clerics, Shia clerics and activists, women's rights activists, anti-corruption activists, judicial reformers, and human rights activists—were imprisoned and sometimes subsequently tortured.

Twitter and Google search data from this period offers detailed digital footprints of the online activity of these arrested individuals, similar non-arrested individuals, the arrested individuals' followers, and the Saudi public, allowing us to test the effects of repression on these diverse actors. These rich and temporally granular data sources are described in detail below.

Data

We gathered five datasets to assess how repression affects online dissent through direct, indirect, and downstream effects:

Tweets by Imprisoned Opinion Leaders

In order to measure the direct effects of physical repression, we began by identifying Saudi opinion leaders who had been imprisoned and were active on Twitter between January 1, 2010 and January 31, 2017.[15] We

[13] For overviews see Ménoret (2016), Lacroix (2011), and Wehrey (2015). We could also consider the labor movement to be a fourth social movement in this context.

[14] For example, Loujain al-Hathloul and Mayasa al-Amoudi (two arrested women's rights activists) were friends who were active in the women's right to drive online movement, as was Hala al-Dosari, the non-arrested opinion leader that we compared them too in the study. Other activists have family ties—for example, women's rights activist Samar Badawi is the sister of prominent human rights activist Raif Badawi, another arrested opinion leader in our dataset. Raif Badawi's wife is also a Saudi human rights activist. We also see family ties in among Shia dissidents—Mohammed al-Nimr is the brother of prominent cleric Nimr al-Nimr, and both were arrested during the period under study.

[15] 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 conducted our historical data collection in January 2017, which marks the end of our data collection period.

attempted to identify all well-known Saudi individuals who were imprisoned in connection with their online activity and who still had active Twitter accounts at the time of our data collection in January 2017 by conducting automated and manual searches of news and human rights reports in Arabic and English. This yielded a list of 49 individuals whose political arrests were widely reported in the Saudi, other Arabic language, or international press. Thirty-six of them had active Twitter accounts at the time of data collection. Online Appendix A lists all imprisoned opinion leaders, including a brief description of their backgrounds, the official justification for their imprisonment, as well as the unofficial reasons for arrest provided by human rights organizations.[16] After identifying these 36 opinion leaders, 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.

Tweets Engaging with Imprisoned Opinion Leaders

To measure the downstream effects of imprisonment on the content of tweets—how changes in the behavior of imprisoned opinion leaders are spread through the Saudi Twittersphere—we used the Historical PowerTrack API to download all public tweets engaging with the imprisoned opinion leaders using the @ sign (for example, @LoujainHathloul). We then filtered this dataset to only include tweets by individuals who were either geolocated in Saudi Arabia or contained location metadata in the location or time zone fields of their profiles indicating that they were located in Saudi Arabia, resulting in a dataset of 32,504,397 tweets produced by 8,506,400 unique users.

We use this dataset to measure the downstream effects of arrests on engagement—retweets, mentions, or replies—with imprisoned opinion leaders.[17] We think of these engaged followers as being similar to *supporters* in the social movements literature (Davenport, Armstrong, and Zeitzoff 2019).[18] The majority of these users (58.2%) engaged both before and after the arrests of the online opinion leaders. Slightly over 10% (13.3%) only engaged with arrested opinion leaders prior to their arrests, and slightly less than a third (28.4%) only engaged after the arrests. Our primary analyses of tweet volume includes all engaged followers.[19]

Tweets of Engaged Followers

We measure indirect deterrence by selecting a random sample of about 30,000 of the users who retweeted, mentioned, or replied to the imprisoned opinion leaders at least once preceding and once following their arrests,[20] stratified by imprisoned opinion leader. We then used Twitter's API to scrape up to 3,200 of each of their most recent tweets for a total of 47,886,355 tweets.[21]

Tweets by Non-Imprisoned Opinion Leaders

Collecting tweets of Saudi opinion leaders who are similar to those imprisoned but were not themselves subjected to repression during our period of study enables us to test the indirect effects of arrest on individuals who might have been most likely to be deterred by seeing their peers imprisoned.

To identify these individuals, 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, which is approximately the middle of the data collection period for the imprisoned opinion leaders. This resulted in 235,215,314 tweets sent by 1,048,568 unique accounts. We then measured the average cosine similarity between tweets produced by these

[16] Several of the opinion leaders in our dataset were imprisoned on multiple occasions. These arrests tended to be separated by at least a year, and were often in response to different activities. For example, Mohamad al-Arefe was first imprisoned in response to his comments about the Muslim Brotherhood as described in the introduction, and then, over a year later, was imprisoned for his critical tweets about the Saudi Hajj pilgrimage train. In our study, we limited our analysis to opinion leaders' first arrests in order to keep the analysis consistent across all imprisoned opinion leaders. Our analysis of direct effects is limited to 25 opinion leaders who were released by the time of our data collection. A table of the dates of these first arrests is provided in Online Appendix A.

[17] We do not include everyone who follows the imprisoned opinion leaders because that would include individuals who were not attentive and may not have observed repression—for example, people who have Twitter accounts but who do not use Twitter, or people who followed these opinion leaders because it was recommended by Twitter's algorithm but were not actually interested in these individuals.

[18] Given the absence of social movement organizations in Saudi Arabia, there are no formal organizational leaders or members. We treat the most prominent voices of a social movement as its leaders, and individuals who engage with these leaders as supporters.

[19] Substantive results of our volume analysis remained unchanged if we subset to the 58.2% of users who engage both before and after arrests, or if we subset to the 71.5% (58.2% + 13.3%) of users who engaged before the arrests.

[20] When we sampled opinion leaders' engaged followers, we only sampled users who had engaged with an opinion leader at least once preceding and once following political imprisonment. This ensures that we are only looking at content produced by users that had engaged with the imprisoned opinion leaders prior to their imprisonment, rather than comparing tweets by users who started tweeting about the imprisoned opinion leaders after their imprisonment (potentially bolder or more politically engaged individuals) to tweets produced by individuals who engaged with the opinion leaders pre-arrest.

[21] For most of these users, 3,200 tweets encompassed all of the tweets they have ever tweeted. In order to use these tweets to assess the sentiment of political content (described below), we first filtered these tweets to include only those that contained the most common political keywords in a random sample of tweets sent by the imprisoned and match opinion leaders coded 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 A.11 in Online Appendix D. The proportion of political tweets in our data was 0.34 on average across the entire period and there was no significant difference from the pre-arrest to post-arrest periods.

accounts and our imprisoned opinion leader accounts to find matches for each opinion leader.[22] Our matching method matched Sunni clerics with Sunni clerics, Shia clerics with Shia clerics or Shia-rights activists, women's right activists with women's rights activists, etc., giving us confidence in the validity of the approach.[23] We then used the Historical PowerTrack API to download all of the public tweets of matched non-imprisoned opinion leaders from 2010 to 2017, resulting in a dataset of 365,337 tweets.

Google Search Data

To measure private interest in the imprisoned opinion leaders among the general public, we downloaded daily and weekly Saudi Google Search data for the Arabic names[24] of each imprisoned opinion leader in the month and year preceding and following their arrests. This enabled us to see how often Saudi Google users privately searched for these individuals.[25] This real-time behavioral measure of how much attention everyday Saudis—who the social movements literature refer to as bystanders (Davenport, Armstrong, and Zeitzoff 2019)—were privately paying to imprisoned opinion leaders allows us to assess indirect and downstream effects. 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). By contrasting this data to public tweets, we can therefore capture preference falsification (Kuran 1997).

Empirical Strategy

Analyses of Tweet and Search Volume

We first calculate the average change for all imprisoned opinion leaders combined in the *volume* of both tweets and Google searches from the pre-arrest to the post-arrest (or post-release) period. We then conduct placebo tests to generate a null distribution of changes in tweet and search volume by choosing a placebo intervention date at random, and repeating this procedure 10,000 times. The resulting null distribution of change in volume allows us to conduct a non-parametric test of our hypotheses. We determine whether or not the combined change in volume we observe using the actual dates of imprisonment falls outside the mass of the distribution of changes in volume generated by choosing placebo dates at random. Specifically, we compute a one-sided p-value representing the proportion of simulated differences in volume that are at least the size of the actual observed difference in volume. We also use interrupted time series analysis (ITSA) and event count models (negative binomial autoregressive models) to test the robustness of these results (See Online Appendix B.1 and B.2, respectively, for details).

Changes in the volume of tweets give us our first measure of direct, indirect, and downstream effects. If we find evidence of a direct deterrent effect of repression we should see a lower volume of tweets from the imprisoned opinion leaders post-release compared to pre-arrest. If we observe an indirect deterrent effect we should see less engagement with imprisoned opinion leaders from their followers and fewer tweets from similar non-imprisoned opinion leaders post-arrest compared to pre-arrest. If we observe direct or indirect backlash effects, we should see the opposite results.

Examining private behavior, if we see a decrease in Google Search volume, this might be evidence of a downstream effect whereby people lose interest in the imprisoned opinion leaders and their causes following imprisonment. If we see an increase in Google Search volume, this might be evidence of an indirect backlash effect, whereby the arrest draws greater attention to the imprisoned opinion leaders and their causes.

Crowdsourced Evaluation of Tweet Content

Moving beyond changes in the volume of activity, we also evaluate how the content of tweets produced by imprisoned opinion leaders, their engaged followers, and non-imprisoned opinion leaders changed in the aftermath of repression. In particular, we are interested in the effect of imprisonment on four categories of content: (1) criticism or praise of the regime, (2) criticism or praise of government policies, (3) criticism or praise of Saudi society, and (4) discussion of collective action.[26]

With regard to criticisms, 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 associated with the monarchy. It also includes tweets calling for democracy, constitutional monarchy, and other changes

[22] Our goal in identifying these matches was to study a theoretically interesting population who might face greater threat of repression than ordinary Saudis, not to conduct matching for causal inference. 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 0.22, and ranged from 0.09 to 0.57. As might be expected, the closest matches for many of our imprisoned opinion leaders were other imprisoned opinion leaders, so the closest non-imprisoned match was not necessarily the closest match overall. We found 14 unique matches because several opinion leaders had the same top match. For example, three of our imprisoned women's rights activists were matched to the women's rights activist Hala al-Dosari, who was not imprisoned. Several arrested human rights activists were matched to the human rights activist Waleed al-Sulais, who also was not imprisoned.

[23] See Online Appendix A for information and account metadata for all imprisoned opinion leaders and matches.

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

[25] We gather daily Google search data, which is available for up to a 90 day period, as well as weekly Google search data, which is available for up to a five year period.

[26] The coding categories we used for the content of tweets are designed to be sufficiently broad to capture a wide range of politically or socially relevant content. The precise wording of the coding instructions is provided in Online Appendix D. In order to capture a wide range of topics we did not constrain our coding to particular categories of interest such as women's driving, the Egyptian coup, government corruption, or disbursement of oil wealth, for example.

to the political regime, including rights for the marginalized Shia minority. 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 its policies but not its underlying legitimacy.

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 includes tweets discussing protest or organized crowd formation on the ground. While rare, these tweets represent a particularly threatening 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 crowd-sourced large-scale human coding of tweets via Figure Eight (formerly Crowdflower), a platform similar to Mechanical Turk but with more native Arabic speakers. This content analysis gives us a second measure of direct, indirect, and downstream deterrent or backlash effects. If there is a direct deterrent effect then we should see less critical and more supportive content in the tweets of imprisoned opinion leaders post-release. There might also be a downstream deterrent effect if tweets directly mentioning or retweeting imprisoned opinion leaders become more supportive and less critical as a consequence of the direct deterrent effect of repression on the arrested individuals. If there is an indirect deterrent effect we should see content that is less critical or more supportive of the regime in the tweets produced by engaged followers and tweets produced by similar non-imprisoned opinion leaders post-arrest. Backlash would produce the opposite results.

RESULTS

We first present evidence of direct deterrence of repressed public opinion leaders and its downstream effects, before moving to evidence of indirect backlash effects—increased dissent by engaged online followers of the imprisoned opinion leaders and increased attention from everyday Saudis. We then show a lack of indirect deterrence among non-imprisoned opinion leaders.

Direct Deterrent Effects on Arrested Opinion Leaders

Opinion leaders who were imprisoned were deterred from dissent following their releases. First, their volume of tweets decreased. This can be seen in Figure 1, which presents the pre-arrest and post-release volume of tweets produced by imprisoned opinion leaders with a loess smoothed trend line for the month before arrest and the month after release (panel a), and for the year before arrest and the year after release (panel b).

Our placebo tests show that imprisoned opinion leaders collectively tweeted significantly less in the post-release period relative to the pre-arrest period. The lower panels (c and d) of Figure 1 present the results of non-parametric placebo tests, which compare the actual difference in tweet volume associated with the arrest to the difference in volume generated by placebo intervention dates chosen at random at the month (panel c) and year (panel d) time frames. The doted vertical line shows the actual average daily difference in tweet volume between the pre-arrest and post-release month (panel c) and year (panel d). Imprisoned opinion leaders tweeted less when comparing the month before arrest to the month after release. Importantly, they also tweeted less when comparing the year before arrest to the year after release, which suggests that the deterrent effect was not a temporary phenomenon that only lasts for a few-days or weeks. These results are consistent with the results of our interrupted time series analysis and event count models reported in Online Appendix B.1 and B.2, which both show statistically significant decreases in tweet volume among imprisoned opinion leaders at the 0.05 level one month and one year after release.

These direct deterrent effects were not simply driven by the change of behavior of a specific subgroup of public opinion leaders. When we disaggregate these effects by the type of opinion leader (Sunni clerics, liberal reformers and activists, and Shia clerics and activists), by the length of the arrests, by whether or not the individual was explicitly arrested for their online dissent,[27] by the time period in which he or she was arrested, and by his or her number of followers, we see decreased volume of activity across all of these subgroups (See Online Appendix C for details).

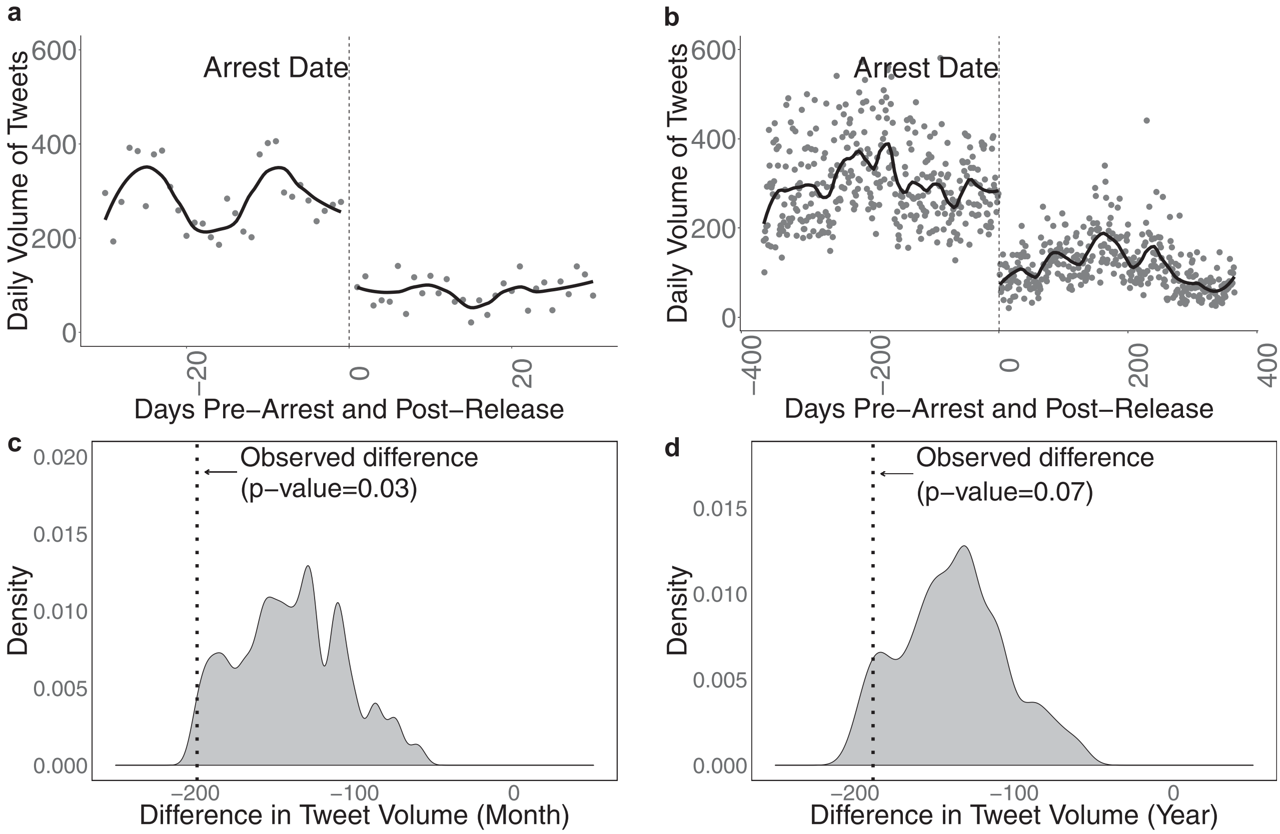
The content of what imprisoned public opinion leaders tweeted also changed. For example, prior to his arrest, one liberal activist in our dataset tweeted his support for Islamists (opposed by the Saudi regime) who gained power after the Arab Spring:

> They told me: Why support the Arab spring revolutions if the Islamists have benefited from them? I said: the Arab Spring revolutions have restored dignity and trust to the peoples.
>
> قالوا لي: لماذا تؤيد ثورات الربيع العربي وقد استفاد منها الإسلامويون؟
> قلت: ثورات الربيع العربي أعادت الكرامة والثقة إلى الشعوب

By contrast, after he was released, some of his tweets expressed support for the Saudi regime and its leaders:

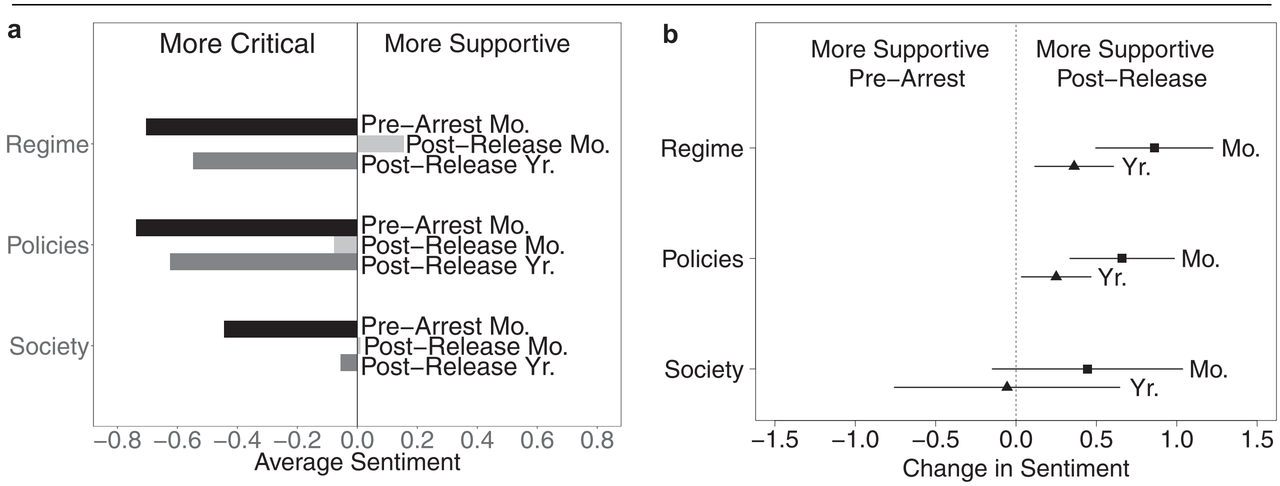
[27] While all of the opinion leaders in our study were speculated to have been arrested for their online activity, the official Saudi government rationale for the arrests did not always include online activity. For example, the official reason for the imprisonment of the three judicial reform activists was "disobeying rule/slandering judiciary," but according to media and observer reports, the "disobedience" and "slander" all took place on Twitter (see Online Appendix A for details of every arrested individual).

FIGURE 1. Volume of Tweets by Imprisoned Opinion Leaders



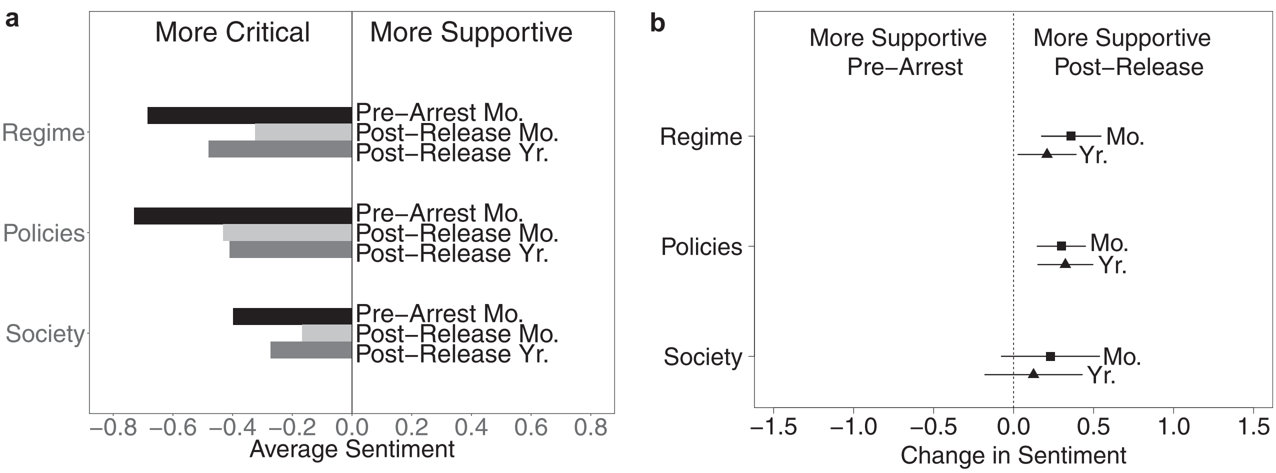
Note: Daily volume of tweets by imprisoned opinion leaders in month before arrest and after release (panel a), in year before arrest and after release (panel b) with loess smoothed line. Non-parametric placebo tests comparing observed difference in volume pre-arrest to post-release (dotted line) to a null distribution of placebo dates in month (panel c) and year (panel d) periods.

FIGURE 2. Sentiment of Tweets by Imprisoned Opinion Leaders



Note: Average sentiment of tweets by imprisoned opinion leaders in each time period (panel a). Change in average tweet sentiment from pre-arrest to post-release with 95% CI (panel b).

FIGURE 3. Sentiment of Tweets Mentioning, Retweeting, or Replying to Imprisoned Opinion Leaders



Note: Average sentiment of tweets mentioning, retweeting, or replying to imprisoned opinion leaders in each time period (panel a). Change in average tweet sentiment from pre-arrest to post-release with 95% CI (panel b).

> I followed the al-Arabiya interview with Prince Mohammed bin Salman. The man truly impressed me: fluent in his speech, transparent and practical, knows what he is talking about, his vision is clear, the future is his focus.
>
> تابعت مقابلة العربية مع الأمير محمد بن سلمان. حقيقة أبهرني الرجل: متمكن في حديثه، شفاف وعملي،يعرف ما يتحدث عنه،صاحب رؤية واضحة المستقبل هاجسه

More systematic analysis demonstrates that opinion leaders who were very critical of the regime, its policies, and Saudi society, who called for political and social reforms before their imprisonment, changed their tune after their release. As we described above, Arabic speakers coded the content of tweets as expressing either support (positive sentiment), criticism (negative sentiment), or neutral attitudes toward the Saudi regime, policies, or society.[28] The bar plots in Figure 2 (panel a) show the average sentiment of these three types of tweets one month pre-arrest (black bar), one month post-release (light gray bar), and one year post-release (dark gray bar).

In the month before imprisonment, the average sentiment of their tweets about the Saudi regime was quite negative (-0.7 on a scale ranging from -1 to 1). In the month after their release, the average sentiment became positive ($+0.15$), echoing what we saw in the example tweets. In aggregate, instead of criticizing the regime and calling for change, immediately after their releases from prison, the opinion leaders tweeted more supportive content about the regime.

In the year following their releases, average sentiment again became negative, but less negative than it had been in the pre-arrest period. A similar pattern is evident when examining their tweets about Saudi policies and Saudi society, both of which became less negative after release. These pre-arrest and post-release changes in the content of public opinion leaders' tweets about the regime and policies are statistically significant, as pictured in the coefficient plot displaying the results of t-tests with 95% confidence intervals in Figure 2 (panel b). Although tweets calling for collective action were very rare in the Saudi Twittersphere in this period (around 1.5% of arrested opinion leader tweets prior to arrest), they disappeared almost entirely post-release (See Online Appendix E for details). Together, these results suggest that arrests had a direct deterrent effect on the online dissent of imprisoned opinion leaders. Imprisoned opinion leaders reined in their criticisms of the regime, its policies, and Saudi society, and their already rare online posts about offline mobilization essentially ceased.

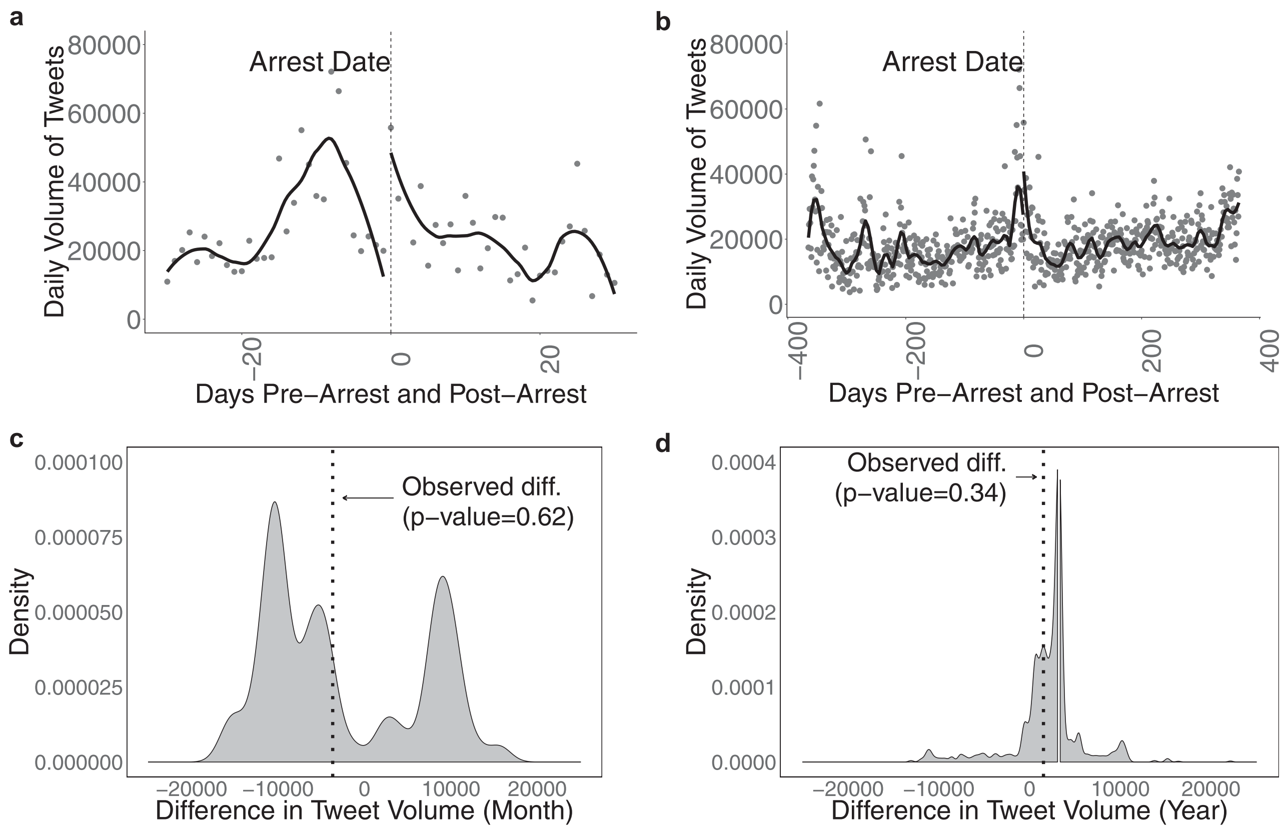
Downstream Deterrent Effects on Engaged Followers

As a consequence of the direct deterrent effect on arrested opinion leaders tweet content, we also see a downstream deterrent effect in which tweets directly mentioning, retweeting, or replying to them become less critical. Less than 2% of tweets produced by engaged followers of imprisoned opinion leaders were mentions, replies, or retweets of the imprisoned opinion leaders. However, largely because these tweets often contain the text of tweets produced by the imprisoned opinion leaders—which are less critical post-release—this 2% of tweets is also less critical of the regime.

Panel a of Figure 3 shows this—the sentiment of retweets, mentions, and replies are very critical in the month pre-arrest (black bar), and become less critical in the month (light gray bar) as well as year (dark gray bar) post-release. The results of t-tests with

[28] Tweets coded as irrelevant were excluded from the analysis.

FIGURE 4. Volume of Tweets by Engaged Followers of Imprisoned Opinion Leaders



Note: Daily volume of mentions, retweets, and replies of imprisoned opinion leaders in pre and post-arrest month (panel a) and year (panel b) with a loess smoothed line. Non-parametric placebo tests comparing observed difference in volume pre- and post-arrest (dotted line) to a null distribution of placebo dates in the month (panel c) and year (panel d) periods.

95% confidence intervals displayed in panel b of Figure 3 show that retweets, replies, and mentions became more supportive of the regime, policies, and Saudi society, and the results are statistically significant for tweets about the regime and policies.[29]

Indirect Backlash Effects on Engaged Followers

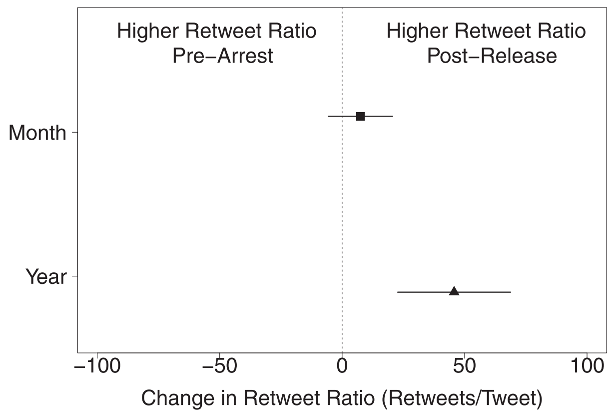
Despite evidence of downstream deterrent effects of repression in the content of the small subset of tweets directly engaging with the imprisoned opinion leaders, when we look at overall levels of engagement with arrested opinion leaders we see no evidence of deterrence. Moreover, when we examine changes in the content of engaged followers' tweets overall—rather than just those tweets that retweet, reply, or mention imprisoned opinion leaders—we see evidence of an indirect backlash effect.

First, examining the volume of retweets, mentions, and replies by engaged followers we find that the arrests did not deter them from interacting with the arrested opinion leaders on Twitter. This can be seen panels a and b of Figure 4, which plots pre- and post-arrest trends in the volume of tweets by engaged followers of imprisoned opinion leaders as local regression lines with loess smoothing for the month before and after arrests (panel a), and for the year before and after arrests (panel b).[30] Panels c and d of Figure 4 show the results of placebo tests, which demonstrate that our observed difference falls right in the middle of the null distribution of volume differences generated by using placebo intervention dates. This suggests that the arrests did not have a deterrent effects on engagement with imprisoned opinion leaders either a month or a year following the arrest. When we analyze these data in other ways—using interrupted time series regressions and event count

[29] We also see a decrease in the number of tweets calling for collective action, though again the overall volume of these tweets is very small (see Online Appendix E).

[30] Figure 4 also shows an uptick in daily mentions approximately ten days before the arrest. Many of the opinion leaders were arrested for their online activities. The uptick may capture some of the tweet(s) that the regime deemed to be problematic and online attention around these tweets.

FIGURE 5. RTs of Imprisoned Opinion Leaders



Note: Change in the average number of retweets per tweet of tweets by imprisoned opinion leaders from the pre-arrest to post-release periods with 95% CI.

models—we also do not find any statistically significant declines in volume (See Online Appendix B.1 and B.2).

We also disaggregate users by when they began engaging with the imprisoned opinion leaders. The subset of tweets produced by engaged followers who only engaged with the imprisoned opinion leader pre-arrest contains a spike in volume immediately post-arrest. Similarly, the subset of tweets produced by those who engaged both pre- and post-arrest also shows a spike in volume immediately post-arrest. This suggests that individuals who actively followed imprisoned opinion leaders were not deterred by their imprisonment. We also find that the spike in tweets by engaged followers who followed the leader prior to arrest is slightly smaller than what we observe in Figure 4. This indicates that the arrest also led to activity from newly engaged followers who had not engaged previously (See Online Appendix B.3).

When we disaggregate this analysis by different characteristics of imprisoned opinion leaders, we do not see statistically significant differences in volume between any subgroup of followers—engaged followers of different types of opinion leaders, engaged followers of opinion leaders who have more or fewer followers, engaged followers of opinion leader arrested in different time periods or for different lengths of time, or engaged followers of opinion leaders who were explicitly arrested for the content of their tweets (See Online Appendix C for details).

Additionally, engaged followers of the imprisoned online opinion leaders retweeted the tweets of imprisoned leaders at higher rates post-release. As Figure 5 demonstrates, on average, tweets produced by imprisoned opinion leaders garnered more retweets per tweet when comparing the month before the arrests and after the releases, and in the year before the arrests and after the releases, though these results are only statistically significant in the year period (p-value $= 0.001$). The average number of retweets per tweet in the pre-arrest year was 13 and the average number of retweets per tweet in the year post-release was 60. This suggests that physical repression did not scare away ordinary Twitter users from engaging with imprisoned opinion leaders or their causes.

Not only did repression not decrease engagement with imprisoned opinion leaders, we also observe a backlash in the content of tweets produced by engaged followers of the imprisoned opinion leaders. Engaged followers stepped up their dissent by criticizing the regime and calling for reform. Although the liberal activist quoted above tweeted in support of Saudi leaders following his release from prison, his online followers continued to criticize the regime. For example, one follower tweeted in the post-arrest period:

> A society that denies and condemns and calls for the killing of all who do not agree with its ideology, religion and beliefs, is a society that is intellectually sick and the people suffer.
>
> المجتمع الذي يُكفر ويُزندق ويدعوا لقتل كل من لا يوافق فكره ودينه
> ومعتقداته، هو مجتمع مريض فكريا وتعاني الشعوب

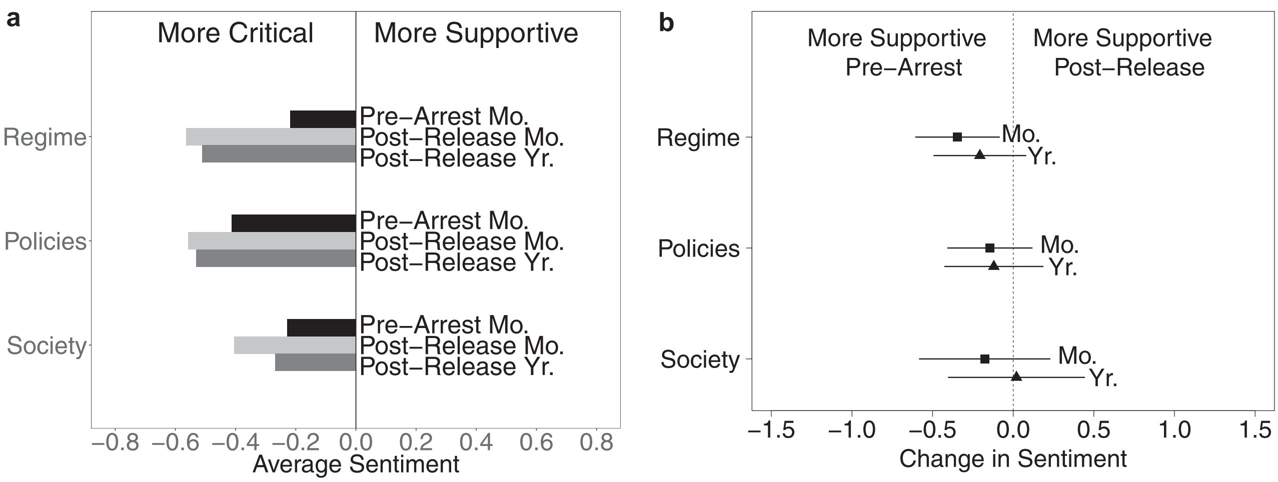
When we systematically examine tweets by engaged followers of imprisoned opinion leaders, we observe the same pattern of increased criticism and dissent. Panel a of Figure 6 displays a bar plot of the average sentiment of political tweets produced by Saudi Twitter users who engaged with imprisoned opinion leaders. The average sentiment is always negative, but in the month (light gray bar) and year (dark gray bar) after the releases, online sentiment was more critical toward the regime, policies, and society than before the arrests. Panel b of Figure 6 presents results of t-tests of differences in online sentiment before the arrests and after the releases and shows that online followers were more critical of the regime, its policies, and Saudi society, though these results are not consistently statistically significant.

Indirect Backlash Effects on the Saudi Public

Private interest in the imprisoned opinion leaders by the Saudi public also spiked. Panels a and b of Figure 7 show a very large increase in the popularity of Saudi Google searches for imprisoned opinion leaders in the immediate aftermath of the arrests. The top of Figure 7 plots the pre-arrest and post-arrest trends in our data with loess smoothed trend lines, based on daily search volume for the month prior to and post-arrest (panel a) and based on weekly search volume for the year prior to and post-arrest (panel b).[31] Searches increased

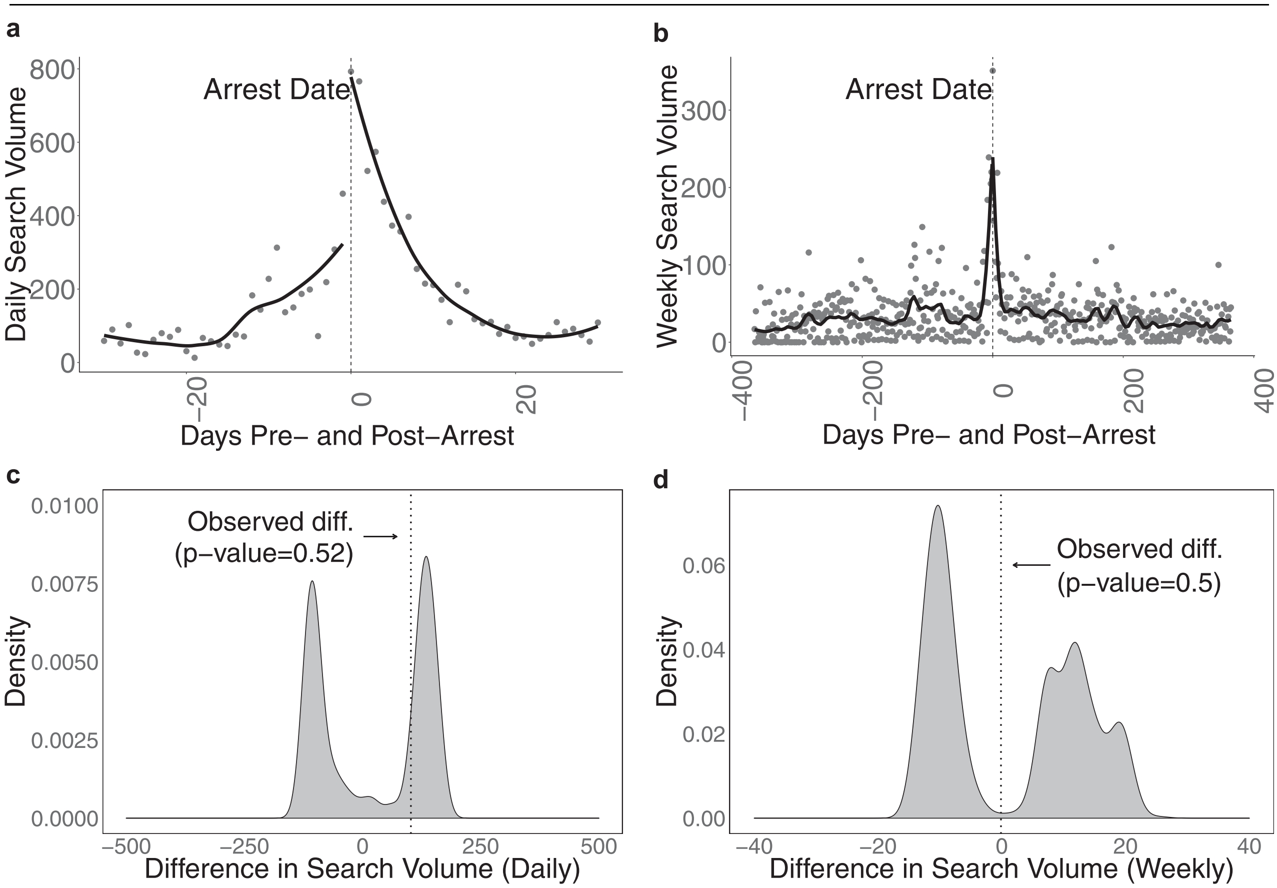
[31] Google search data is a relative measure of the popularity of a given search term on Google. For each imprisoned opinion leader, the relative popularity is calculated as the total number of searches for that person's Arabic name divided by the total number of searches from that same geographic region and time window. The resulting numbers for each imprisoned opinion leader are then scaled on a range of 0 to 100 based on a topic's proportion to all searches on all topics. The plots show the daily (or weekly) sums of these relative scores for all imprisoned opinion leaders.

FIGURE 6. Sentiment of Tweets by Engaged Followers of Imprisoned Opinion Leaders



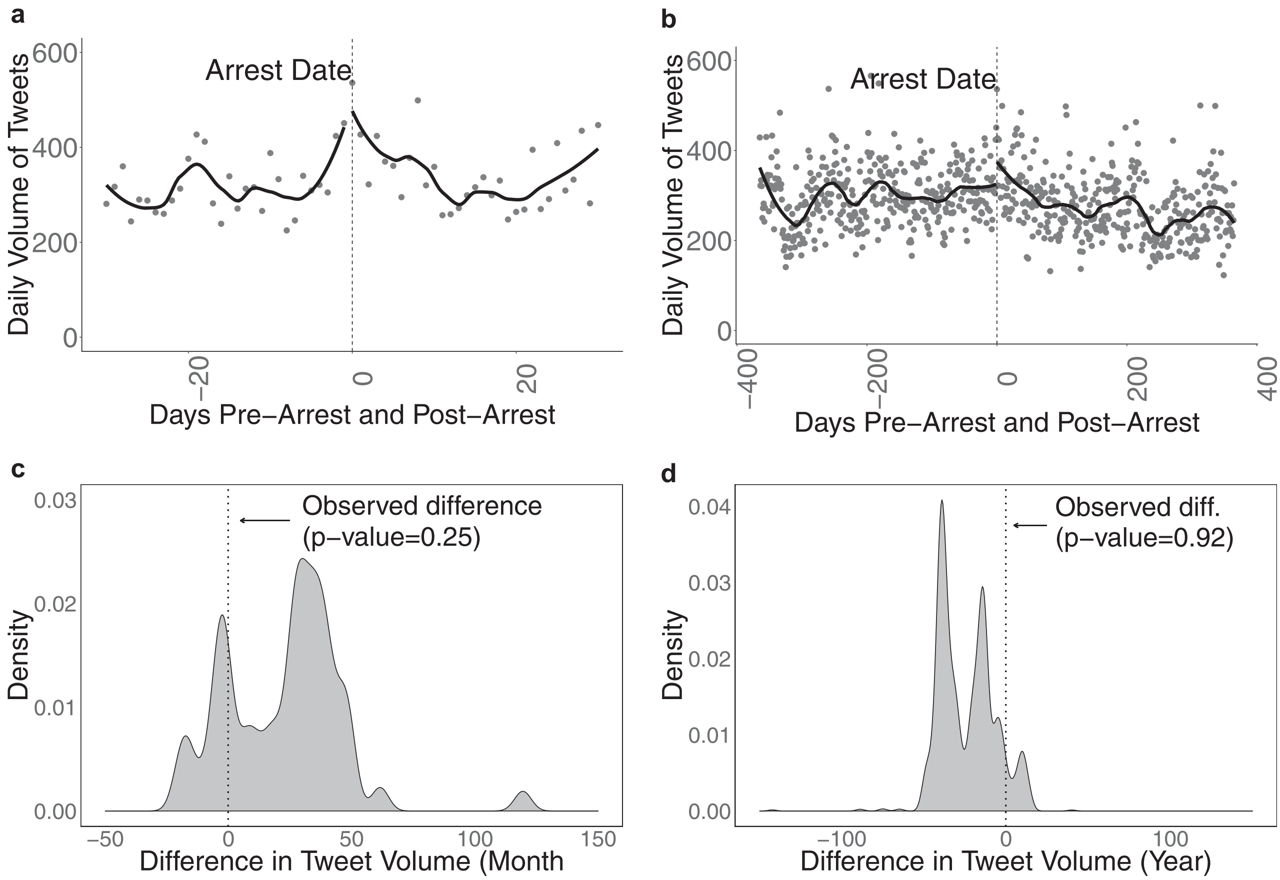
Note: Average sentiment of tweets by engaged followers of imprisoned opinion leaders in each time period (panel a). Change in average tweet sentiment from pre-arrest to post-release with 95% CI (panel b).

FIGURE 7. Relative Volume of Saudi Google Searches for Imprisoned Opinion Leaders



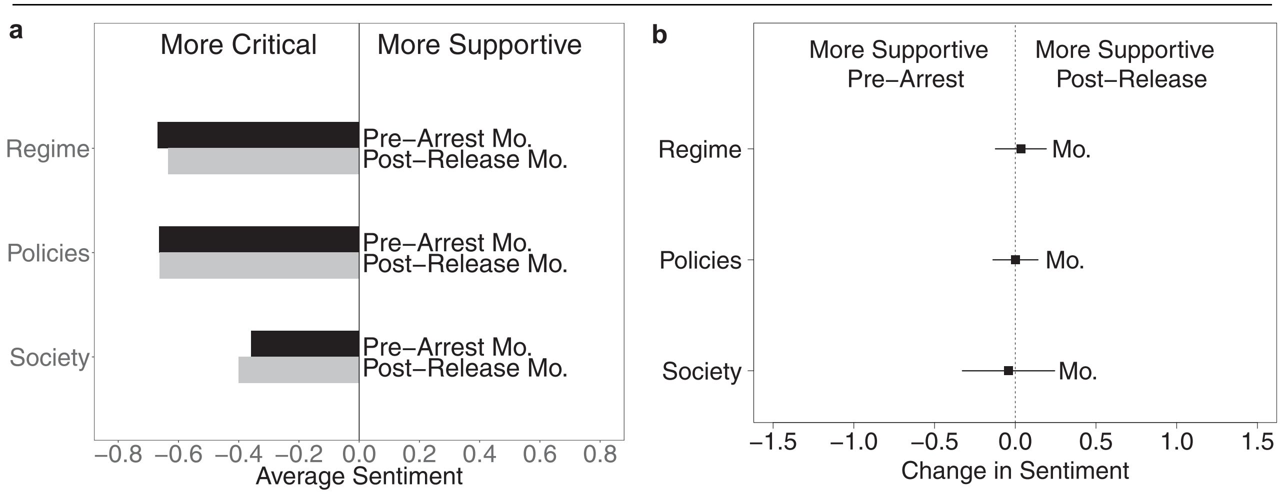
Note: Daily relative volume of Google searches for the Arabic names of imprisoned opinion leaders in the month before and after arrest (panel a) and weekly relative volume in the year before and after arrest (panel b) with loess smoothed lines. Non-parametric placebo tests comparing the observed change in relative volume pre-arrest to post-arrest (dotted line) to a null distribution of changes generated by choosing placebo dates at random the in month (panel c) and year (panel d) time periods.

FIGURE 8. Volume of Tweets by Non-Imprisoned Opinion Leaders



Note: Daily volume of tweets by non-imprisoned opinion leaders in the month before and after arrest (panel a) and in the year before and after arrest (panel b) with loess smoothed lines. Non-parametric placebo tests comparing the observed change in volume pre-arrest to post-arrest (dotted line) to a null distribution of changes generated by choosing placebo dates at random the in month (panel c) and year (panel d) time periods.

FIGURE 9. Sentiment of Tweets by Non-Imprisoned Opinion Leaders



Note: Average sentiment of tweets by non-imprisoned opinion leaders in each time period (panel a). Change in average tweet sentiment from pre-arrest to post-release with 95% CI (panel b).

significantly in the immediate aftermath of arrest and returned to pre-arrest levels soon after. Furthermore, the similar results we observe across the tweets engaging with imprisoned opinion leaders and the Google search data suggest that there is no evidence of preference falsification (Kuran 1997) or self-censorship that we might have observed if individuals had been afraid to publicly engage on Twitter but nonetheless continued to search on Google in private. While the uptick in Google searches quickly returns to normal and therefore is not captured by our placebo analysis, our interrupted time series analysis shows a statistically significant level change in the immediate aftermath of the arrest.[32]

No Indirect Deterrent Effect on Similar Opinion Leaders

Turning to individuals who were perhaps most at risk of repression—those who had engaged in similar dissent and also had large online followings—we find no evidence that they were deterred. Figure 8 shows that unlike the imprisoned opinion leaders, similar opinion leaders did not decrease their volume of tweets after the arrests of their peers. There was little change in the daily volume of tweets produced by the non-imprisoned opinion leaders in the month before and month after the arrests, or in the year before and year after arrests. When we disaggregate this analysis by different characteristics of imprisoned opinion leaders, we do not see statistically significant declines in volume among any subgroup of matched individuals. If anything, in some subgroups (non-imprisoned opinion leaders matched to imprisoned leaders with long arrest periods and imprisoned leaders with fewer followers), we see an increase in the volume of non-imprisoned opinion leader tweets. The results of interrupted time series analysis and event count models are similar and are displayed in Online Appendix B along with the subgroup analysis reported in Online Appendix C.

Similarly, non-imprisoned opinion leaders did not change the content of their tweets. For example, the non-arrested liberal activist matched to the activist quoted above tweeted equally negative content in the pre-arrest period, and did not reign in his dissent after the first activist's arrest. Following the arrest, the non-imprisoned activist called for freedom, referencing that Arabs have fought for freedom since pre-Islamic times:

> Freedom is an innate state inherent in the primitive existence of man. And therefore Arabs believe in it and have fought for it since the Sa'lek (pre-Islamic) rebellion of the Arabs.
>
> الحرية حالة فطرية متأصلة في الوجود البدائي للإنسان . ولذلك آمن بها العرب، وناضل في سبيلها منذ تمرد صعاليك العرب

Our systematic content analysis reveals the same lack of deterrence. Figure 9 shows that non-imprisoned opinion leaders continued to express negative sentiment toward the regime, and there were no statistically significant changes in the content of their tweets about the Saudi regime, policies, or society. Discussion of offline collective action by these other opinion leaders also remained unchanged.[33]

Our results also align with what we know anecdotally from this period—that activists and clerics frequently denounced the arrests of their friends and colleagues and did not appear deterred. For example, non-imprisoned women's rights activist Hala al-Dosari spoke out against the arrests of Loujain al-Hathloul and Mayasa al-Amoudi in 2014 (BBC News 2014). Non-imprisoned clerics denounced the arrests of Mohamad al-Arefe and Mohsen al-Awaji in 2013 and non-imprisoned human rights activists protested the arrests of members of the Saudi Civil and Political Rights Association in 2011 (The Daily Star 2011).

Discussion

Why did repression deter imprisoned opinion leaders but create backlash among their followers? And why did it fail to deter similar non-arrested individuals from dissent? Interpreting our results through the lens of the repression, online mobilization, and censorship literatures provides key insights into their generalizability as well as important scope conditions of our study.

Our finding of a direct deterrent effect contrasts with research in democratic contexts showing that repression can generate backlash among its targets, especially leaders of social movement organizations (Davenport 2015; Sullivan and Davenport 2017). This may be a consequence of the high level of repression in Saudi Arabia, which is virtually unconstrained by social norms, cultural practices, or law, as well as the absence of formal social mobilization organizations to offer a modicum of protection for dissenters. The future risk of repression is high (likely much higher than in any democratic context), not only for individuals who have been repressed but also for their family members and associates. This direct deterrence may be more likely to occur in contexts like Saudi Arabia where social movement organizations are largely absent. In regimes where social movement organizations exist and where offline mobilization is more prevalent, we may see backlash instead of deterrence among repressed individuals in line with Sullivan and Davenport (2017).

The lack of indirect deterrence and the backlash we observe among followers of arrested opinion leaders aligns with the large repression literature on backlash[34]

[32] Once again, these results are similar when disaggregated by subgroup, as we report in detail in Online Appendix C.

[33] We do not code more tweets one year out for the non-imprisoned opinion leaders because the chilling effect on the imprisoned opinion leaders diminishes over time, and since we do not observe any change one month post-arrest for non-imprisoned leaders, we are unlikely to observe any further away from the arrests of their peers. Results showing no change in the volume of tweets discussing collective action are displayed in Online Appendix E.

[34] For example: Eckstein (1965), Feierabend, Feierabend, and Gurr (1972), Gurr and Duvall (1973), Kuran (1989), Francisco (1996), Khawaja (1993), Kurzman (1996), and Lichbach (1998).

and echoes research on how censorship can backfire. There are two components to the backlash we observe. One is increased engagement and higher levels of criticism by individuals who had already been actively engaging with imprisoned opinion leaders online prior to their arrests. The second is new engaged followers who began interacting with the imprisoned opinion leaders or privately searching for them after their arrests. Backlash in both instances was likely related to the fact that online dissent and mobilization are extremely low cost, allowing large numbers of people to participate. As a result, the risk of repression for any one supporter is very low. This means that publicly visible repression may generate interest without incentivizing individuals to self-censor or to stop supporting a movement online (Earl and Kimport 2011).

Given that we observe backlash among engaged followers in this extremely repressive context—and that we find no evidence of preference falsification—we expect these results to generalize to other contexts where social media is used to sustain social movements online. We also expect these backlash results to extend to regimes that are less repressive and those that have freer media. We have no reason to believe these results are unique to absolute monarchies, theocracies, or resource-rich dictatorships.

There are, however, important scope conditions with regard to these indirect backlash effects. First, the repression we analyze in the Saudi context is publicly visible. Just as censorship is more likely to generate backlash when it is obvious than when it is hidden, when repression of online dissent is more covert, we might not observe backlash. We should therefore expect to see indirect backlash when those targeted by repression are prominent figures. In other contexts, where there have been high-profile cases of repression targeting ordinary social media users,[35] backlash might be less likely if online supporters learn new information about their own level of risk by observing the arrest. More generally, we would expect to see similar results in contexts where repression does not alter observers' calculation about their own risk of repression. In regimes where repression of online activity is rarer, or makes an example of ordinary citizens, repression might be more likely to change peoples' perceived level of risk and constrain their behavior.

Indirect backlash effects of repression may also be limited to regimes that lack fine-grained control over the online sphere. The backlash we observe is likely accelerated by the fact that the Saudi regime cannot quickly, reliably, and selectively censor tweets. Unlike in China, for example, where Chinese social media platforms carry out the censorship requests of the Chinese government to prevent certain online movements from going viral (King, Pan, and Roberts 2014), Twitter does not necessarily comply with Saudi government censorship requests. Indeed, the Minister of Information admitted in February 2013 that monitoring Twitter was difficult due to the large volume of users (Al-Rasheed 2013). The increased criticisms and calls for reform we observe suggest that in the time period of our study, online disinformation campaigns by the Saudi government (including large-scale fabrication of pro-government content) also did not stifle critical voices.[36] If a regime exercised a high degree of control over social media, it could censor dissent after the arrests. However most authoritarian regimes lack high degrees of control, with the exception of China and perhaps Russia.

The lack of indirect deterrence we observe among online opinion leaders who were not imprisoned is perhaps our most surprising result. First, it differs from the results of Sullivan and Davenport (2017), which show that repression demobilized movement members who did not experience repression. We believe this difference is due to the fact that the dissent we describe in this paper is taking place online instead of offline, and occurring within social movements but not in social movement organizations. In Sullivan and Davenport (2017), repression was applied to individuals who participated in an offline protest. Members of this organization who were not present at the protest were not repressed and later withdrew from the organization perhaps because they were perceived by protest participants as lacking commitment and socially stigmatized. Our context differs dramatically. Because of the large and diffuse nature of online mobilization, the Saudi regime could not and did not arrest everyone who participated in online dissent. In contrast to Sullivan and Davenport (2017), other leaders who were not repressed did not lack commitment—they also took risks and spoke out.

What else might explain why these non-arrested opinion leaders were not deterred by observing repression despite the seemingly high levels of risk they faced relative to ordinary Saudi social media users? Perhaps these non-imprisoned opinion leaders differed in some way from those who were imprisoned, and did not actually face as high a level of risk. However, our results suggest that non-imprisoned opinion leaders tweeted at a similar rate in general and produced a similar proportion of tweets expressing negative sentiment toward the regime, policies, and society in the pre-arrest period. It therefore seems unlikely that non-imprisoned opinion leaders believed they were immune from state repression because their online activity differed from that of their arrested peers. But there may be other characteristics of non-imprisoned opinion leaders we do not observe—such as the nature of their offline dissent, political connections, or their ability to leave Saudi Arabia—that alter their perceived level of risk and enable them to continue dissenting.[37]

35 For example, in 2013, when China began cracking down against online dissent, not only were prominent figures arrested but ordinary citizens as well, including the highly publicized story of the arrest of a middle school student from Western China, see https://nyti.ms/2wcIoHI (Accessed May 22, 2019).

36 For news reports of Saudi Arabia's troll army, see https://nyti.ms/2RXKNzP, https://wapo.st/2w8YClj (Accessed May 21, 2019).

37 For example, non-imprisoned women's rights activist Hala al-Dosari currently has an academic position at Harvard University, and human rights defender Waleed Sulais left Saudi Arabia for exile.

A final reason we do not observe an indirect deterrent effect on non-imprisoned opinion leaders may be that—like ordinary social media users—observing arrests did not alter their calculus about the risk they faced. Because well-known opinion leaders in Saudi Arabia are at risk of arbitrary imprisonment at all times, the political imprisonment of other opinion leaders may not provide any new information to constrain their behavior any more than the daily reality of living under such repressive conditions. This may stand in contrast to other contexts where the repression of a small number of actors might be sufficient to alert others to acceptable norms and deter dissent (Link 2002; Stern and Hassid 2012).

CONCLUSION

Analyzing over 300 million tweets and Google search data between 2010 and 2017, this paper offers new temporally granular measures of the direct, indirect, and downstream effects of repression on online dissent. Furthermore, by allowing us to capture both the volume and content of mass and opinion leader messages on the same platform, Twitter data provides novel perspectives on how diverse actors behave in the aftermath of physical repression. Together, our results suggest that while physical repression had a direct deterrent effect on the individuals who were imprisoned, it had an indirect backlash effect on their engaged followers and the public, and did not deter similar opinion leaders who were not imprisoned.

Given these results, why would the Saudi regime—or other regimes around the world—use physical repression in response to online dissent? Governments may go through a learning process in how to suppress online mobilization, and perhaps these targeted arrests were one phase in that process. Governments may also default to a particular style of repression depending on their institutional history or who is in power. Indeed 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. Examples include larger-scale political imprisonment, such as the late 2017 purge of about 500 business people, princes, government ministers, and activists; an increase in death sentences such as that of popular cleric Salman al-Oudah; and even the targeting of opponents living abroad such as the recent murder of influential journalist Jamal Kashoggi (Rauhala 2018; Freedom House 2018; Human Rights Watch 2018b). Our work suggests that—as of 2017—despite the threat of repression, many opinion leaders and everyday Saudis 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 more recent conditions of intensified repression in Saudi Arabia.

In general, more research is needed to capture how physical repression is being used by authoritarian and democratic regimes in response to online opposition worldwide. We hope the analytical leverage gained by disaggregating the effect of repression on online dissent by type, actor, behavior, and time will be used in future studies examining other regions, regime types, forms of physical repression, and forms of online dissent.

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit https://doi.org/10.1017/S0003055419000650.

Replication materials can be found on Dataverse at: https://doi.org/10.7910/DVN/9AMKHL.

REFERENCES

Al-Arabiya. 2015. *41 Percent of Saudis Have Twitter Accounts: Study*: Al-Arabiya. http://english.alarabiya.net/en/media/digital/2015/03/11/41-percent-of-Saudis-have-Twitter-accounts-study-.html.

Al-Rasheed, Madawi. 2016. *Muted Modernists: The Struggle over Divine Politics in Saudi Arabia*. Oxford, UK: Oxford University Press.

Al-Rasheed, Nayaf. 2013. *Twitter Out of Our Control (Arabic)*: Al-Watan. https://bit.ly/2wjYTBZ.

Alabaster, Olivia. 2018. "Saudi Arabia Using Anti-terror Laws to Detain and Torture Political Dissidents, UN Says." *The Independent*. https://ind.pn/2Htch9O.

Ayres, Jeffrey M. 1999. "From the Streets to the Internet: The Cyber-Diffusion of Contention." *The Annals of the American Academy of Political and Social Science* 566 (1): 132–43.

BBC News. 2014. "Saudi Terrorism Court 'To Try Women Drivers'." *BBC News*. https://www.bbc.com/news/world-middle-east-30602155.

Bennett, W. Lance, and Alexandra Segerberg. 2012. "The Logic of Connective Action: Digital media and the Personalization of Contentious Politics." *Information, Communication & Society* 15 (5): 739–68.

Bimber, Bruce, Andrew J. Flanagin, and Cynthia Stohl. 2005. "Reconceptualizing Collective Action in the Contemporary media Environment." *Communication Theory* 15 (4): 365–88.

Calamur, Krishnadev. 2018. "Saudi Arabia Rejects Human-Rights Criticism, Then Crucifies Someone." *The Atlantic*. https://bit.ly/2LbVyOx.

Cingranelli, David L., and David L. Richards. 2010. "The Cingranelli and Richads (CIRI) Human Rights Data Project." *Human Rights Quarterly* 32: 401.

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*, ed. Lorrie Faith Cranor. Pittsburgh, PA: ACM, 112–21.

Davenport, Christian. 2005. "Repression and Mobilization: Insights from Political Science and Sociology." In *Repression and Mobilization*, eds. Christian Davenport, Hank Johnston, and Carol Mueller. Minneapolis, MN: University of Minnesota Press, vii–xli.

Davenport, Christian. 2007. "State Repression and Political Order." *Annual Review of Political Science* 10: 1–23.

Davenport, Christian. 2015. *How social movements die*. Cambridge, UK: Cambridge University Press.

Davenport, Christian, Dave Armstrong, and Thomas Zeitzoff. 2019. "Acceptable or Out of Bounds? Scaling Perceptions of Challenger and Government Tactics as Well as Understanding Contentious Interaction." In *Workshop on Human Rights and Repression: Latin America in Comparative Perspective*. Working Paper.

Davenport, Christian, and Molly Inman. 2012. "The State of State Repression Research since the 1990s." *Terrorism and Political Violence* 24 (4): 619–34.

Davenport, Christian, Sarah A. Soule, and David A. Armstrong. 2011. "Protesting while Black? The Differential Policing of American

Activism, 1960 to 1990." *American Sociological Review* 76 (1): 152–78.

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

Earl, Jennifer. 2011. "Protest Arrests and Future Protest Participation: The 2004 Republican National Convention Arrestees and the Effects of Repression." In *Special Issue Social Movements/Legal Possibilities*, ed. Austin Sarat. Bingley, UK: Emerald Group Publishing Limited, 141–73.

Earl, Jennifer, and Alan Schussman. 2002. "The New Site of Activism: On-Line Organizations, Movement Entrepreneurs, and the Changing Location of Social Movement Decision Making." In *Consensus Decision Making, Northern Ireland and Indigenous Movements*, ed. Patrick G. Coy. Bingley, UK: Emerald Group Publishing Limited, 155–87.

Earl, Jennifer, and Katrina Kimport. 2008. "The Targets of Online Protest: State and Private Targets of Four Online Protest Tactics." *Information, Communication & Society* 11 (4): 449–72.

Earl, Jennifer, and Katrina Kimport. 2011. *Digitally Enabled Social Change: Activism in the Internet Age*. Camrbidge, MA: MIT Press.

Earl, Jennifer, Katrina Kimport, Greg Prieto, Carly Rush, and Kimberly Reynoso. 2010. "Changing the World One Webpage at a Time: Conceptualizing and Explaining Internet Activism." *Mobilization: International Quarterly* 15 (4): 425–46.

Earl, Jennifer, and Sarah A. Soule. 2010. "The Impacts of Repression: The Effect of Police Presence and Action on Subsequent Protest Rates." In *Research in Social Movements, Conflicts and Change*, ed. Lisa Leitz. Bingely, UK: Emerald Group Publishing Limited, 75–113.

Eckstein, Harry. 1965. "On the Etiology of Internal Wars." *History and Theory* 4 (2): 133–63.

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

Feierabend, Ivo K., Rosalind L. Feierabend, and Ted Robert Gurr. 1972. *Anger, Violence, and Politics: Theories and Research*. Englewood Cliffs, NJ: Prentice-Hall.

Fisher, Dana R. 1998. "Rumoring Theory and the Internet: a Framework for Analyzing the Grass Roots." *Social Science Computer Review* 16 (2): 158–68.

Francisco, Ronald A. 1996. "Coercion and Protest: An Empirical Test in Two Democratic States." *American Journal of Political Science* 40 (4): 1179–204.

Freedom House. 2017. *Freedom on the Net*. Washington D.C.: Freedom House.

Freedom House. 2018. *Saudi Arabia: Thousands Held Arbitrarily Dramatic Increase in Detention without Trial*. Washington D.C.: Freedom House. https://freedomhouse.org/report/freedom-world/2017/saudi-arabia.

Gibney, Mark, Linda Cornett, Reed Wood, Peter Haschke, and Daniel Arnon. 2016. "The Political Terror Scale 1976–2015. Political Terror Scale Website." http://www.politicalterrorscale.org/.

Glum, Julia. 2015. "Saudi Arabia's Youth Unemployment Problem Among King Salman's Many New Challenges after Abdullah's Death." *International Business Times*.

Goldstein, Robert Justin. 1978. *Political Repression in Modern America from 1870 to the Present*. Boston, MA: GK Hall & Company.

Goldstone, Jack A., and Charles Tilly. 2001. "Threat (And Opportunity): Popular Action and State Response in the Dynamics of Contentious Action." In *Silence and Voice in the Study of Contentious Politics*, eds. Ronald R. Aminzade, Jack A. Goldstone, Doug McAdam, Elizabeth J. Perry, William H. Sewell, Sidney Tarrow, and Charles Tilley. Cambridge: Cambridge University Press, 179–94.

Gurr, Ted Robert, and Raymond Duvall. 1973. "Civil Conflict in the 1960s: A Reciprocal Theoretical System with Parameter Estimates." *Comparative Political Studies* 6 (2): 135–69.

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

Hobbs, William R., and Margaret E. Roberts. 2018. "How Sudden Censorship Can Increase Access to Information." *American Political Science Review* 112 (3): 1–16.

Human Rights Watch. 2018a. *Human Rights Watch World Report 2017: Saudi Arabia*: Human Rights Watch. https://www.hrw.org/world-report/2017/country-chapters/saudi-arabia.

Human Rights Watch. 2018b. *Saudi Arabia: Thousands Held Arbitrarily Dramatic Increase in Detention without Trial*: Human Rights Watch. https://www.hrw.org/news/2018/05/06/saudi-arabia-thousands-held-arbitrarily.

Ibahrine, Mohammed. 2016. "The Dynamics of the Saudi Twitterverse." In *Political Islam and Global Media: The Boundaries of Religious Identity*, eds. Noha Mellor and Khalil Rinnawi. Abingdon, UK: Routledge, Chapter 12.

Jansen, Sue Curry, and Brian Martin. 2015. "The Streisand Effect and Censorship Backfire." *International Journal of Communication* 9: 656–71.

Jenkins, J. Craig, and Charles Perrow. 1977. "Insurgency of the Powerless: Farm Worker Movements (1946–1972)." *American Sociological Review* 42 (2): 249–68.

Khawaja, Marwan. 1993. "Repression and Popular Collective Action: Evidence from the West Bank." *Sociological Forum* 8 (1): 47–71.

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.

Kuran, Timur. 1989. "Sparks and Prairie Fires: A Theory of Unanticipated Political Revolution." *Public Choice* 61 (1): 41–74.

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

Kurzman, Charles. 1996. "Structural Opportunity and Perceived Opportunity in Social-Movement Theory: the Iranian Revolution of 1979." *American Sociological Review* 61 (1): 153–70.

Lacroix, Stéphane. 2011. "Comparing the Arab Revolts: Is Saudi Arabia Immune?" *Journal of Democracy* 22 (4): 48–59.

Lacroix, Stephane. 2014. "Saudi Arabia's Muslim Brotherhood Predicament." https://wapo.st/2UZTQVR.

Lichbach, Mark Irving. 1987. "Deterrence or Escalation? The Puzzle of Aggregate Studies of Repression and Dissent." *Journal of Conflict Resolution* 31 (2): 266–97.

Lichbach, Mark Irving. 1998. *The Rebel's Dilemma*: Ann Arbor, MI: University of Michigan Press.

Link, Perry. 2002. "The anaconda in the Chandelier: Chinese Censorship Today." In *The New York Review of Books* 49 (6). April 11.

Mackey, Robert. 2014. "Saudi Activists Call for the Release of Women Detained for Driving." https://nyti.ms/2IMzGbo.

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

Moore, Will H. 1998. "Repression and Dissent: Substitution, Context, and Timing." *American Journal of Political Science* 42 (3): 851–73.

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* 7 (4): 1–20.

Myers, Daniel J. 1994. "Communication Technology and Social Movements: Contributions of Computer Networks to Activism." *Social Science Computer Review* 12 (2): 250–60.

Nabi, Zubair. 2014. "Censorship Is Futile." *First Monday*. https://firstmonday.org/article/view/5525/4155.

Noman, Helmi, Robert Faris, and John Kelly. 2015. "Openness and Restraint: Structure, Discourse, and Contention in Saudi Twitter." *Berkman Center Research Publication* No. 2015-16, 1–48.

Oberschall, Anthony. 1973. *Social Conflict and Social Movements*. Englewood Cliffs, NJ: Prentice-Hall.

Olivier, Johan L. 1991. "State Repression and Collective Action in South Africa, 1970–84." *South African Journal of Sociology* 22 (4): 109–17.

Opp, Karl-Dieter, and Christiane Gern. 1993. "Dissident Groups, Personal Networks, and Spontaneous Cooperation: The East German Revolution of 1989." *American Sociological Review* 58 (5): 659–80.

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): 167–88.

Rasler, Karen. 1996. "Concessions, Repression, and Political Protest in the Iranian Revolution." *American Sociological Review* 61 (1): 132–52.

Rauhala, Emily. 2018. "Saudi Arabia's Spat with Canada Was a Lesson. Trump Ignored it." *The Washington Post*. Available at: https://www.washingtonpost.com/world/2018/10/17/saudi-arabias-spat-with-canada-was-lesson-trump-ignored-it/.

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

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*. New York: HarperCollins.

Stern, Rachel E., and Jonathan Hassid. 2012. "Amplifying Silence: Uncertainty and Control Parables in Contemporary China." *Comparative Political Studies* 45 (10): 1230–54.

Stukal, Denis, Sergey Sanovich, Richard Bonneau, and Joshua A. Tucker. 2017. "Detecting Bots on Russian Political Twitter." *Big Data* 5 (4): 310–24.

Sullivan, Christopher Michael, and Christian Davenport. 2017. "The Rebel alliance Strikes Back: Understanding the Politics of Backlash Mobilization." *Mobilization* 22 (1): 39–56.

The Daily Star. 2011. "Saudi Group Raps Authorities over Recent Arrests." *The Daily Star*. http://www.dailystar.com.lb/Article-Print.aspx?id=135177&mode=print.

Tilly, Charles. 1978. *Collective Violence in European Perspective*. Ann Arbor, MI: University of Michigan Press.

Tilly, Charles. 2005. *Repression, Mobilization, and Explanation*. Vol. 218. Minneapolis: University of Minnesota Press.

Tufekci, Zeynep, and Christopher Wilson. 2012. "Social media and the Decision to Participate in Political Protest: Observations from Tahrir Square." *Journal of Communication* 62 (2): 363–79.

Walter, Eugene Victor. 1969. *Terror and Resistance: A Study of Political Violence, with Case Studies of Some Primitive African Communities*. Vol. 1. New York: Oxford University Press.

Wehrey, Frederic. 2015. "Saudi Arabia's Anxious Autocrats." *Journal of Democracy* 26 (2): 71–85.

Wehrey, Frederic M. 2013. *The Forgotten Uprising in Eastern Saudi Arabia*. Washington, DC: Carnegie Endowment for International Peace.

Woolley, Samuel C., and Philip N. Howard. 2016. "Automation, Algorithms, and Politics—Political Communication, Computational Propaganda, and Autonomous Agents—Introduction." *International Journal of Communication* 10: 9.

Young, Lauren E. 2017. "Mobilization under Threat: Emotional Appeals and Dissent in Autocracy." Working Paper.