ANNUAL REVIEWS

Annual Review of Political Science

Resilience to Online Censorship

Margaret E. Roberts

Department of Political Science, University of California, San Diego, La Jolla, California 92093-0521, USA; email: meroberts@ucsd.edu

ANNUAL REVIEWS CONNECT

www.annualreviews.org

- Download figures
- Navigate cited references
- Keyword search
- Explore related articles
- Share via email or social media

Annu. Rev. Political Sci. 2020. 23:401–19

First published as a Review in Advance on February 21, 2020

The *Annual Review of Political Science* is online at polisci.annualreviews.org

https://doi.org/10.1146/annurev-polisci-050718-032837

Copyright © 2020 by Annual Reviews.
This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See credit lines of images or other third-party material in this article for license information

OPEN ACCESS

Keywords

censorship, digital politics, Internet, autocracies, information manipulation, propaganda

Abstract

To what extent are Internet users resilient to online censorship? When does censorship influence consumption of information and when does it create backlash? Drawing on a growing literature on Internet users' reactions to censorship, I posit that awareness of censorship, incentives to seek out information, and resources to circumvent censorship are essential to resilience to censorship. I describe how authoritarian regimes have adapted their strategies of censorship to reduce both awareness of censorship and demand for uncensored information.

INTRODUCTION

While the Internet has long been touted as a technology that is difficult to censor, regimes around the world have adopted a wide variety of censorship technologies and online propaganda strategies to try to control it. As a result, a rich debate has emerged as to whether the Internet solidifies or undermines autocratic rule. Some scholars have called attempts to control the Internet futile because the controls can often be easily circumvented (Diamond 2010). Others have a much more dire view of the ability of governments and powerful interests to manipulate the information environment and limit their own accountability (Morozov 2011, MacKinnon 2012).

Much of the theoretical and empirical literature surrounding this debate centers on autocrats and the incentives, strategies, and tools they have to carry out censorship and propaganda online. However, an equally relevant part of the puzzle is how citizens react to government efforts to control information online. Internet activist John Gilmore once posited that the Internet was impervious to censorship because "[t]he Net interprets censorship as damage and routes around it" (Elmer-Dewitt 1993). But to what extent do Internet users actually route around censorship? The degree to which social media users recognize censorship and go out of their way to find information is central to how information spreads online in spite of information manipulation. Users' willingness to circumvent censorship depends crucially on their incentives and the circumstances surrounding censorship.

This article reviews a burgeoning literature that studies how individuals respond to online information manipulation. Drawing on this literature, I discuss the contexts in which users are resilient to censorship and the circumstances in which they are affected by it. I make two basic points about resilience to online censorship and discuss their implications for the strategies of authoritarian governments in manipulating information online. First, awareness of information manipulation is key to resilience to that manipulation. Even though some evidence suggests that under certain circumstances awareness of censorship creates chilling effects (e.g., Huang 2015, 2018; Pan & Siegel 2020), awareness of censorship also allows users to find ways to circumvent it and to update their beliefs accounting for censorship. In some cases, awareness of censorship may incentivize users to seek out information their government has concealed (e.g., Nabi 2014, Pan & Siegel 2020). In other cases, awareness of censorship can facilitate coordination, imposing costs on the government for implementing censorship (Hassanpour 2016, Boxell & Steinert-Threlkeld 2019). For those who are unaware that censorship exists and do not know what information might be censored, compensating for information manipulation is very difficult, particularly in online contexts, where censorship is masked by algorithms and the complexity of user interfaces.

Second, resilience to censorship depends on demand for the information that has been censored and the ability to pay the costs imposed by censorship (Roberts 2018, Chen & Yang 2019). Variation in users' willingness and abilities to route around censorship creates substantial heterogeneity in censorship resilience across types of information, contexts, and people. The literature suggests that entertainment is more resilient to censorship than political information (e.g., Zuckerman 2015, Hobbs & Roberts 2018). Resilience to censorship is strengthened by education, access to technology, and a wider and more diverse social network (Roberts 2018, Chen & Yang 2019). This heterogeneity in resilience to censorship could make coordination against the government more difficult because segments of the population are exposed to different information environments (Little 2017).

These insights from the literature offer some explanations for current trends in digital authoritarianism. First, the importance of awareness of censorship to resilience explains the shift in regime strategies toward more invisible forms of censorship that reduce citizens' awareness of manipulation. It also explains and forecasts an increasing use of microtargeted censorship and propaganda,

particularly for regimes with the technological capacity to aggregate and use data collected on citizens. Second, the fact that typically many users are unwilling or unable to evade censorship could explain why autocrats have largely been successful at managing the Internet even when they cannot completely control the online environment (Shadmehr & Bernhardt 2015, Chen & Xu 2017). I conclude with a discussion of how current technological advances may impact resilience to censorship and with a call for future research.

CENSORSHIP TAXES INFORMATION

To know how people can be resilient to censorship, we have to understand what censorship is. While governments have adopted a huge range of technologies to control information on the Internet, they all share a common characteristic: Censorship technologies all impose taxes on information by requiring users to incur costs for accessing or spreading information. In this way, they incentivize users to consume and spread information that the government would like them to consume because this information is less costly.

These costs manifest themselves in different ways. I organize the types of costs imposed by censorship into three categories according to the mechanism through which they affect the user: fear, friction, and flooding (Roberts 2018).[1]

Censorship through fear imposes costs on users who share and access information by threatening these users with some type of costly punishment. Countries around the world have adopted laws and regulations that prohibit users from writing or sharing particular types of information. Many governments resort to extralegal intimidation as a means of deterring users from spreading information. Iran, Russia, China, and Saudi Arabia are all examples of countries that have explicit laws that govern political speech on the Internet. A 2018 Freedom House report found that in 28 of the 65 countries surveyed, someone had been killed or physically attacked by the government or a third party for content they had posted online (Shahbaz 2018). In 43 countries, an online user was arrested or imprisoned for political or social writing. The estimated cost of these punishments and probability that these laws will be enforced enter into a user's cost-benefit calculation in deciding whether or not to spread or access information.

However, fear is not the only mechanism through which censorship imposes costs on users. Censorship can impose costs simply by making information more difficult to access, a mechanism I call friction (Roberts 2018). Methods of censorship that rely on friction do not necessarily make information explicitly off limits but instead force users to spend more time or money spreading or finding information. Governments that impose firewalls to block websites, require search engines to reorder search results, slow down access, or shut down the Internet all make spreading particular types of information more difficult. Of the 65 countries Freedom House surveyed in its 2018 report, 21 had blocked access to a social media or messaging platform for some amount of time, and 13 had shut down the Internet or mobile phones with a political motivation (Shahbaz 2018). The cost of friction is often imposed in terms of time and knowhow—friction costs users more time and effort to access particular types of information. While friction can often be circumvented, users may not always be willing to pay these costs.

Similar to friction, flooding is a method by which governments manipulate the relative costs of information, in this case by introducing massive amounts of information in a coordinated way to the digital space. Governments might flood the online space with messages that convey their beliefs or ideology, often known as propaganda. Governments are also known to coordinate

[1]Note that there are a number of other ways authors have classified censorship, for example, offline versus online (Sanovich et al. 2018) and by tool (Deibert et al. 2008, 2010, 2011).

online information—which could be positive messages, entertainment, or other miscellaneous information—in a way that is designed to divert the conversation away from information that reflects badly on the government (King et al. 2017a, Munger et al. 2019). Alternatively, they might spread false information online in order to confuse users about what is true and what is false, a practice scholars and pundits have called gaslighting (Jack 2017). [Tucker et al. (2018) provide a good review of literature that studies the types and content of political disinformation.] These efforts require users to pay more costs, in terms of time and energy, to find information about ongoing events and to separate good information from false or biased information.

While many pundits predicted that the Internet would be difficult to censor, a wide variety of research has shown that government censorship efforts can have a huge impact on information access and—at times—political belief and action. Weidmann et al. (2016) use cross-national measurements of Internet penetration to show that politically excluded groups simply have less access to the Internet overall. Boxell & Steinert-Threlkeld (2019) show large impacts of a 2018 social media tax in Uganda on overall social media use. King et al. (2013, 2014) show that the Chinese government individually removes large numbers of social media posts, often related to collective action events, and Bamman et al. (2012) show that removals of posts from Weibo in China are focused in certain geographies, massively affecting the distribution of information available online. Roberts (2018) shows large effects of China's Great Firewall on traffic coming from China to newly blocked websites and impacts of the timing of censorship on the spread of information. Chen & Yang (2019) show that college students in China who are given access to the unfiltered Internet and incentivized to use it become more skeptical about government policies and the health of the Chinese economy.

Other research has shown that government flooding of messages on the Internet can have large effects on the prevalence of content online. Verkamp & Gupta (2013) use five political events from China, Russia, Mexico, and Syria to show that coordinated spam can overwhelm a political hashtag on Twitter, diluting conversations related to the event. King et al. (2017a) estimate that the Chinese government directs the posting of 448 million social media posts per year on Chinese social media websites, largely made up of "cheerleading" posts that distract from ongoing events. Stukal et al. (2017), having developed a method to identify bots on Russian political Twitter, find that bots accounted for more than 50% of tweets on the majority of days between February 2014 and December 2015. Sanovich et al. (2018) use a similar dataset and find that a large portion of bots on Russian political Twitter promote news articles. This activity is likely to increase the search rankings of certain news websites, affecting what users read.

These findings, combined with online experiments in democracies that show huge impacts of friction on the Internet, suggest that the costs of access to information can have large effects on consumption of information and belief about politics. Epstein & Robertson (2015) use lab and online experiments to show that small changes in the order of results presented by a search engine have a large influence on the voting intentions of participants. While the experimental setting makes it difficult to know the external validity of this experiment, this "search engine manipulation effect" (Epstein & Robertson 2015) suggests that government manipulation of search engine algorithms could have large effects on political behavior. King et al. (2017b) show that the coordinated coverage in national newspapers has large impacts on the distribution of information that is discussed in social media. Participation experiments in democracies, such as Facebook experiments (Bond et al. 2012, Jones et al. 2017), have shown that an online nudge to go out and vote can significantly increase the likelihood of participation. These studies suggest that actors with power over what information reaches Internet users, and how quickly, can potentially have a large impact on what users see, what they believe, and when they decide to participate.

These findings are consistent with a substantial literature on traditional media. Government control of the print media can have a large influence on public opinion and participation when it reduces the number of outlets that individuals have access to, raising the costs of accessing information from independent sources (see Larreguy & Marshall 2019 for a review). For example, Enikolopov et al. (2011) show that the decrease in access to independent television had large and positive impacts on votes for the government party in Russia and substantial negative effects on votes for the opposition. Adena et al. (2015) use geographic and time variation in exposure to Nazi propaganda in prewar Germany to show that areas treated with propaganda were more likely to support the Nazi Party. Yanagizawa-Drott (2014) shows that exposure to radio in Rwanda advocating the killing of Tutsis increased participation in the genocide. McMillan & Zoido (2004) reveal that the Peruvian government was willing to pay millions of dollars in bribes to the media to keep information from citizens, an indication that the value of captured media and control over the distribution of information can be quite high in particular contexts.

PATHS OF INFLUENCE OF CENSORSHIP

The studies above provide evidence that censorship influences the distribution of information that citizens have access to. Models of political persuasion and coordination suggest that the distribution of information could influence citizens through several different pathways. First, it could affect citizens through agenda setting or salience, bringing issues that reflect well on the government to the forefront of citizens' attention and weighting them highly in importance (Converse 1962, McCombs & Shaw 1972). Research in political communication has shown that the prevalence of an issue reported in the media can affect how important citizens believe this issue to be (see, e.g., Iyengar 1990, Zaller 1992, Iyengar & Kinder 2010).

Second, it could affect citizens by exposing them disproportionately to biased information, which citizens could take into account to update their beliefs in ways that are more favorable to the government (for a review, see Gerber & Green 1999). Theories and experimental evidence have shown that recipients of information might update their beliefs in a Bayesian way, updating their priors to adjust to new information (Bullock 2009, Hill 2017). These theories suggest that, all other things equal, the more information a person receives that is favorable about the government, the more likely that person is to also favor the government.

Third, even if citizens' beliefs are not impacted by censorship, the distribution of information that users access might influence their propensity to coordinate against the government by influencing their evaluations of what others believe or by impacting the distribution of logistical information about protest. This logic is related to theories that citizens know that political protest can only be successful when others also coordinate; therefore, they rely on information about others' likely actions to decide their own (Kuran 1989, 1991; Lohmann 1994; Shadmehr & Bernhardt 2011). Little's (2016) model shows that communication of the logistics of protests may be the most dangerous aspect of communication technology to authoritarian regimes. These pathways may be particularly important in online protest mobilization, as research has shown that protest diffusion online follows different patterns than offline protest diffusion, allowing for quicker incorporation of a broader population (Bennett & Segerberg 2012), new subpopulations, and new causes (Earl 2010).

However, these pathways can all be undermined if users are able to notice and resist censorship. Internet users are not always passive consumers of information. On the contrary, there is substantial evidence that under certain circumstances users can be very savvy consumers of information, noticing and circumventing censorship. Despite the evidence (described above) that censorship can have a large impact on the distribution of information—and potentially on political belief and

action—there are time periods when users are willing to route around it. For the remainder of this review, I discuss the circumstances under which users are resilient to the pathways of manipulation described in this section.

For Internet users, resilience to online censorship depends on a few crucial factors. First, are users aware that censorship is happening and aware of the types of information that it is affecting? Can they use this knowledge to update their beliefs to account for censorship? Second, do users want unfiltered information enough to be willing to pay the increased costs of access imposed by censorship? Third, do users have the resources to evade censorship—in terms of the predisposition, education, money, and social networks that can facilitate its evasion? Last, do users believe that others are aware of and resilient to censorship? I describe each of these factors and the literature that informs them below.

AWARENESS OF CENSORSHIP AND BACKLASH

Because of the nature of information, censorship can be disguised, making information manipulation difficult to notice. While fear-based censorship—meant to intimidate and deter—must be visible in order to be effective, more sophisticated forms of censorship that work through friction and flooding such as blocking of websites, reordering of search results, and covert information campaigns can exert their effects without users' awareness (Roberts 2018). For this reason, information manipulation can easily go undetected, and users may not notice government influence on their information environment.

Can users resist censorship if they become aware of it? The literature has a mixed response to this question. On the one hand, awareness of censorship could make censorship even more effective. Knowledge of censorship might create chilling effects, motivating users to self-censor. Chilling effects have mostly been documented for journalists in traditional media (Stern & Hassid 2012) or in face-to-face experiments (Young 2019). However, there has been some evidence of online chilling effects as well. For example, Huang (2015, 2018) uses surveys and online survey experiments to show that individuals in China exposed to "hard propaganda" (news articles that are obviously propaganda) have a greater evaluation of state strength and reduced willingness to protest. Pan & Siegel (2020) follow Saudi Arabian Twitter users and show that those imprisoned and tortured in Saudi Arabia for their online speech are less likely to continue speaking out against the government online after they are released. Fu et al. (2013) find that some users on Sina Weibo (China's version of Twitter) decreased their discussions of political topics after the Chinese government implemented a real-name registration policy, and Tanash et al. (2017) find a high rate of self-deletion of Turkish Twitter accounts after the 2016 attempted coup, which suggests that chilling effects may be stronger when the political stakes are higher.

However, awareness of censorship can also have the opposite effect, drawing users toward the hidden information or creating backlash. Similar to the contentious politics literature, which has provided some evidence of a backlash effect with regard to offline repression (Lichbach 1987, Moore 1998, Goldstone & Tilly 2001, Francisco 2005, Davenport 2007), both Huang (2018) and Pan & Siegel (2020) document online backlash effects associated with awareness of the censorship they study, in addition to chilling effects. Huang (2018) shows that even though users have reduced willingness to protest and increased evaluation of state strength after reading hard propaganda, the propaganda also worsens individuals' opinions of the regime and makes them more likely to report wanting to move abroad. Pan & Siegel (2020) find that even though social media users who were put in prison for their online writing were chilled into reducing their criticism of the government after they were released, their arrests mobilized their followers to increase the amount of antiregime sentiment expressed on Twitter following their arrest. In

addition, Pan & Siegel (2020) find increases in Google search volume for information about those arrested after their arrests, indicating that users were becoming more aware of the dissidents' previous writings as a result of censorship.

This suggests that censorship itself can provide information to citizens that they can use to update their beliefs. Studies of traditional media show that citizens seek out more reliable information when they know that official sources are biased. Geddes & Zaller (1989) find that citizens in authoritarian Brazil can resist government propaganda when they are sophisticated enough to recognize it. Stockmann (2012) finds that citizens in China view commercial newspapers as more consistent with public opinion than official newspapers, and shows that readers strategically seek out information from commercial papers and turn to government papers when they find them informative. In a follow-up study on the search engine manipulation effect, Epstein et al. (2017) show that awareness of search engine manipulation can significantly reduce its impact on public opinion. These studies suggest that users can take into account information manipulation when they become aware of it.

Like the finding in Pan & Siegel (2020) that censorship can draw users toward that which was censored, the literature has documented several cases where censorship has caused more (rather than less) interest in the censored material. Several scholars have found that online censorship can mobilize users against the government or inspire more interest in the concealed content because of censorship. This phenomenon is often called the Streisand effect, having been observed after Barbra Streisand filed a lawsuit to remove photos of her mansion from a website owned by the California Coast Records Project. The lawsuit brought attention to the photos, which then attracted hundreds of thousands of viewers to the website (Jansen & Martin 2015).

Nabi (2014) uses case studies in Turkey and Pakistan to study how censorship of YouTube can increase the popularity of the content that is the target of censorship. The author finds that YouTube remained one of the top ten most visited websites for both Turkey and Pakistan even after it was blocked. Further, Google searches for and views of the content that was the target of censorship spiked after the content was censored, indicating that many people may have become aware of or interested in the censored content because of censorship. Nabi (2014) provides evidence that many users in both Turkey and Pakistan were obtaining software that allowed them to evade the block and were accessing the YouTube content by routing around censorship.

Boxell & Steinert-Threlkeld (2019) study the implementation of a social media tax in Uganda. Beginning on July 1, 2018, users were taxed 200 shillings per day for using any one of a set of social media applications, including Twitter, WhatsApp, and Facebook. Even though the tax decreased overall use of these applications, it also inspired backlash. Not only were authors on social media writing more about collective action after the tax, but also the incidence of protests in Uganda increased by 48%, including protests against the social media tax itself.

These studies that document censorship's backlash effect are consistent with empirical work that shows a positive relationship between connectivity and pro-government views, indicating that more censorship can sometimes be a detriment to authoritarian regimes. Kern & Hainmueller (2009) show that individuals who circumvented government censorship in East Germany to access West German television had a more positive opinion of the East German regime, likely because it allowed them access to entertainment that improved their daily lives. Hassanpour (2014, 2016) uses Internet blackouts in Egypt and Syria to argue that lapses in connectivity created "small world networks" by encouraging face-to-face interaction. The Egyptian Internet and media blackout during the Tahir Square protests increased rather than decreased coordination, perhaps because people began seeking information in face-to-face encounters and local, more dense networks (Hassanpour 2014). The Internet blackout in Egypt during the 2011 protest movement made apolitical citizens more aware of and interested in the unrest, and protests coordinated in small

local networks were more difficult for the government to control than the centrally coordinated protests had been. Similarly, Hassanpour (2016) uses a Syrian Internet blackout that originated from abroad in November 2012 to gain insight on how the Syrian conflict was affected by media disruption. He finds that the disruption increased violent incidents on the part of the rebels, in an uncoordinated way, and spread them to new locations, suggesting that there was more dispersed coordination as a result of the outage.

The literature not only shows backlash effects in response to censorship but also suggests that information manipulation might reduce the government's ability to refute misinformation, because citizens discount these refutations. Huang (2017) studies the impact of antigovernment political rumors in China through an online survey experiment. He finds that even when these rumors are rebutted, they significantly decrease citizen trust in the government. Furthermore, government rebuttal of these rumors is less effective than a government critic's rebuttal, indicating that users may discount rebuttals of antigovernment rumors that originate from government sources.

These documented cases of backlash have one thing in common: They all occurred when censorship was visible and affected many users at the same time. Censorship affected one of the most popular websites in both Turkey and Pakistan (Nabi 2014), where the inability to access YouTube likely alerted many users to censorship who had not been aware of it before. Similarly, the social media tax in Uganda (Boxell & Steinert-Threlkeld 2019) is a very visible form of censorship affecting the entire population of Internet users; it serves as a focal point for coordination. Huang (2018) shows that hard propaganda draws awareness toward censored material because of its visibility. More research could be conducted to pinpoint the conditions under which awareness of censorship creates backlash. In the next section, I discuss research that draws on these findings to explain how regimes are adapting their censorship techniques to (*a*) make censorship less visible and (*b*) remove its coordinating potential in order to harness the Internet for their own interests.

AUTHORITARIAN RESPONSES TO CENSORSHIP AWARENESS

Given that awareness of censorship can create more interest in censored material and can lead to backlash, governments have adapted their censorship strategies by only exerting partial control of the Internet through friction and flooding in an effort to hide their manipulation. Formal models of media control in authoritarian regimes point out that authoritarian regimes have incentives to mix biased and unbiased media in order to keep their citizens guessing about the reliability of the information (Gehlbach et al. 2016). Having too much (or too obvious) censorship means that citizens will not believe much of what they read. Shadmehr & Bernhardt (2015) develop a model that shows that autocrats have incentives to credibly commit to censoring less than they will in equilibrium because citizens who are aware of high levels of censorship will update their beliefs with increasingly negative opinions about the government in the absence of news.

In a similar vein, Chen & Xu (2017) show that when a government can commit to a policy of censorship in advance, it should choose a moderate amount of censorship and should commit to implementing some reforms, rather than choosing high levels of censorship. If citizens know that the government is constrained to not censor all of the time, then, when they see no other citizens supporting a measure, they are more likely to interpret that absence of information as a true reflection of public opinion rather than a product of censorship.

These models echo other theoretical work that shows that autocrats should have incentives to only partially control information. Because authoritarian governments rely on vertical flows of information to inform them of the state of the world, too much control of the information environment can be detrimental to autocrats' survival; it impedes collecting information from

and responding to citizens (Egorov et al. 2009, Lorentzen 2014, Qin et al. 2017). In addition, information manipulation is economically expensive and difficult to implement (Miller 2018), and governments may face a trade-off between investing in the economy and suppressing information (Guriev & Treisman 2019).[2]

Even though many pundits predicted that the Internet would be a death knell to authoritarianism, the implication of this literature suggests that the difficulty governments have in exerting complete control over the Internet might actually provide benefits to regimes by imparting the perception of lower censorship and freer information (Shadmehr & Bernhardt 2015). Because the authoritarian regime's partial control of the online space will occasionally bring negative news to the attention of citizens, the absence of bad news can be interpreted as a positive signal for the regime even when that absence was brought about by censorship controls. Partial control could also decrease the likelihood of backlash against censorship by reducing awareness of censorship, whereas complete control might draw attention to censorship.

Regimes with the capacity to implement subtle forms of censorship, for example by controlling the results of search engines (without bringing attention to this control) or selectively removing social media posts, may reap the benefits of online censorship without incurring the risks of awareness. Fear-based censorship relies on the awareness of users—users cannot be deterred if they are not aware of what information is off limits and of the punishment that might befall them for spreading that information. Censorship that relies on the mechanisms of friction and flooding, in contrast, does not have to be visible to users. This may be why we see a move by many regimes toward less obvious forms of censorship (Roberts 2018, Guriev & Treisman 2019).

Because the technology behind the provision of information on the Internet is complicated, manipulation can be made invisible. The ambiguity between whether a slowdown or outage results from a technical failure or some political manipulation can give the government the cover of plausible deniability. Even for experts, censorship can be hard to distinguish from a technical error. A recent Citizen Lab report details the extensive methods that experts used to determine whether the block of the Internet Archive from Jordan was a network error or a Jordanian government block (al Masri 2017). Significant efforts must be made to identify filtered images on China's WeChat or to identify when a topic is subject to search filtering on Baidu (Knockel et al. 2017). The average person's level of knowledge about the Internet is quite low. Survey data indicate that 48% of Internet users in China are unaware that the Great Firewall exists (Roberts 2018). In developing countries, some online surveys have suggested that many users are unaware that Facebook is a part of the Internet (Mirani 2015). This type of confusion about the functioning of the Internet could make it more difficult to detect government interference and subsequently route around it. It might also reduce the ability for users to view censorship as a focal point around which to coordinate.

Regimes that do not have the technological capacity to implement subtle forms of censorship and regimes whose citizens are primarily users of Western social media platforms may have a more difficult time asserting control over the digital sphere because any effort at censorship will be easily noticed by users. In Edmond's (2013) theoretical model, the degree to which autocrats have economies of scale in controlling the media is essential to whether citizens can coordinate against that control. Pan (2016) shows that social media websites used in the vast majority of countries are owned by US firms, which makes subtle forms of censorship more difficult because these governments have less control over the content on the social media platforms. While countries can request that content be removed from US social media websites, this removal can be denied and

[2] See Petrova (2011) and Gehlbach & Sonin (2014) for a discussion of economic incentives and media bias in traditional media.

is often slowly implemented, and US companies often publicize this censorship. China and Iran, where domestic firms control more than half of the social media websites, are notable exceptions to this pattern (Pan 2016). For countries without control over domestic social media websites, Internet blackouts or sudden filtering of Western social media platforms—i.e., visible forms of censorship—may be the only option.

The fact that awareness of censorship can facilitate resilience to information manipulation also may be the reason that governments and powerful interests are using flooding strategies to manipulate information online. A Freedom on the Net 2018 report estimates that, in around half of the 65 countries surveyed, progovernment commentators manipulated online discussions (Shahbaz 2018). Because of the anonymity of the Internet, these efforts may go unnoticed by unsuspecting citizens. In addition, they are relatively cheap for any government to use and therefore do not require control over domestic social media.

Data collected on the Internet may also allow authoritarian regimes to engage in more targeted repression, in order to intimidate influential users without calling attention to censorship more broadly. Gohdes (2020) discusses the trade-off that regimes face between surveillance—with an open and unaffected Internet—and censorship. Drawing on data from Syria, she finds that the regime is able to use more targeted repression when levels of Internet accessibility are high and must resort to more indiscriminate campaigns of violence when users have more limited access to the Internet. This suggests that the online environment can enable and encourage governments to engage in targeting.

Not only may the Internet allow regimes to target fear-based censorship, online microtargeting might allow them to direct friction and flooding to individual users. As vast amounts of data on users become available through social media companies, companies have the ability to personalize information to users and conduct experiments to understand how information affects user behavior. Through the advertising platforms of these companies or by investing in these companies directly, influential interests and governments increasingly have the ability to target posts toward particular types of users. Evidence suggests that the Internet Research Agency at the behest of the Russian government used Facebook's microtargeting capacity to personalize political messages to US Internet users (Lapowsky 2018). Roberts et al. (2020) show evidence that content removal in China might be personalized toward individual users. If users experience information manipulation differently, detection of this manipulation will be more difficult, and coordinated backlash may be less likely. A Pew research survey shows that most social media users in the United States do not understand the logic behind the way information is presented to them on social media and are unaware that their personal data are used in advertising (Gramlich 2019). More research is needed on how governments use large amounts of data on populations to target and personalize censorship and what effects this might have on resilience; Tufekci (2014) provides an overview of some of the important issues at stake.

HETEROGENEITY IN CENSORSHIP RESILIENCE

While awareness of censorship may be close to a necessary condition for resilience to online censorship, it is not sufficient. The backlash to censorship described in the previous sections depends on whether users who are aware of censorship are motivated and able to circumvent it and resist it. In this section, I highlight literature that has documented how demand for the censored information is necessary for resilience to censorship, and that the costs of resisting censorship also are important to resilience. I discuss how variation in the ability and willingness to circumvent censorship creates substantial heterogeneity in resilience to it. I explore what implications this might have for citizens' ability to use the Internet to coordinate against the regime.

It is well established in the political science literature that demand for political information is typically quite low. Downs (1957) calls citizens' general lack of interest in politics "rational ignorance," meaning that for the most part, people rationally should be ignorant of political issues because they are unlikely to be pivotal in those issues. Surveys have documented a very low level of political knowledge among average citizens in democracies (Converse 1964, Popkin 1994). Rational ignorance in politics may be even more likely in authoritarian contexts, where citizens have less control of the political environment than in democracies.

Even for politically interested citizens who are aware of censorship, the inability to know what is missing might make demand for circumvention low. In a field experiment in China, Chen & Yang (2019) demonstrate the importance of demand for information in censorship evasion. The authors randomly assign college students to a treatment condition where they are given free access to circumvention software for 18 months and a control condition where users are not provided with evasion software. Interestingly, even when provided with free access, treated students did not choose to use the evasion software to bypass censorship and read blocked material. However, after being given monetary incentives to read blocked news, the students did use the software to bypass censorship. More interestingly, after the experiment ended, many of the students continued to use the software and even pay for it in order to continue reading the blocked news. This suggests that while the students did not initially realize that there was censored information they were interested in accessing, once they knew what information could be accessed by routing around censorship they did want to circumvent it.

Given low demand for political information, resilience to censorship may be stronger when censorship is applied not just to political information but also to entertainment. Zuckerman's (2015) "cute cat theory of censorship" posits that while demand for political information is often low, Internet platforms that combine entertainment (like photos of cute cats) and politics may be more immune to censorship. This theory would predict general websites (such as YouTube, Facebook, and WhatsApp) that contain both entertainment and political information to be more resilient to censorship than specific websites that mostly offer political information. This argument is consistent with scholarship that shows that consumption of political information increases when it is paired with entertainment. For example, Baum (2002) shows that Americans are more likely to consume news about international politics when the news is paired with human interest stories. Pan & Roberts (2020) show that before the block of Wikipedia, mainland Chinese users largely sought out entertainment content on Wikipedia, but they ended up consuming political content because they were directed to it through the Wikipedia homepage.

Hobbs & Roberts (2018) describe a "gateway effect," where evasion of censorship motivated by demand for entertainment leads individuals to access long-blocked political information. They study the case of the block of the social media website Instagram from China during the 2014 Hong Kong protests. They show that the Instagram block (which occurred on the third day of the protests) incentivized users in China to download evasion software in order to continue to access the website. Hobbs & Roberts (2018) show that those who continue to use Instagram are largely apolitical and seem to have been drawn to circumvent censorship in order to use Instagram for entertainment. Still, once across the wall, users began engaging with long-blocked political information on Twitter and Wikipedia.

While the literature reviewed so far suggests that users may only evade censorship for entertainment, political events may also be able to garner the demand necessary to incentivize resilience. Ball-Rokeach & DeFleur (1976) argue that citizens' attention and dependence on the media are heightened during time periods of social conflict. Using survey data from Central and Eastern Europe, Loveless (2008) finds that seeking information from television, radio, and newspapers is more predominant in countries in the middle of transitions to democracy, and survey respondents

report less information seeking in democracies that are more consolidated. If information seeking increases during crises, we might expect that resilience to censorship would be stronger during crisis events. Weidmann & Rød (2019) provide evidence that the Internet solidifies authoritarian regimes during normal times but undermines them conditional on a protest event having started, which could indicate that governments have less control during times of crisis. Roberts (2018) finds that survey respondents interviewed after an explosion in Tianjin, China, were much more likely to report having evaded the Great Firewall than respondents interviewed in Tianjin before the explosion, suggesting that the crisis may have prompted users to seek out unfiltered information.

Demand for information varies not only across time periods but also, enormously, across Internet users. Roberts (2018) uses survey data to show that those who jump the Firewall in China tend to be better educated, have higher incomes, and have more interest in international politics than the average Chinese Internet user. This is consistent with Chen & Yang's (2019) findings that the political opinions of those who come from disadvantaged backgrounds and have low levels of knowledge about politics are most affected by the provision of the ability and incentive to jump the wall, indicating that they may have been most susceptible to censorship before Chen & Yang's experiment. In their analysis of the impact of the Ugandan social media tax, Boxell & Steinert-Threlkeld (2019) show that users who tweet from Apple devices in Uganda (a proxy for wealth) are basically unaffected by the social media tax, whereas users of Twitter on other devices are strongly affected by the tax. Further, users who previously discussed politics or discussed protests online are much less likely to retreat from social media than other users, indicating that those with high demand for political information are more resilient to censorship.[3] Collectively, these studies suggest that censorship may exacerbate what Norris (2001) introduced as the "digital divide"—those with less education and fewer resources not only have less access to the Internet than those with more education and resources, but also are less likely to be resilient to censorship.

Users' propensity to resist censorship might be affected by those within their social network. Chen & Yang (2019) show some transmission of information between college students provided with evasion software and their roommates. After the experiment, the roommates of those who received evasion software were more likely to answer political questions correctly, indicating that resilience to censorship could have spillover informational effects on other users. Social networks might not only facilitate transmission of information but also inspire censorship evasion. Evidence exists that in uncensored social media contexts, social pressure can impact information seeking. Bond et al. (2012) show how a social message that a subject's friends have voted increases the subject's information seeking about voting. Marshall (2019) shows that individuals in Mexico seek out costly political information in order to achieve political sophistication similar to that of their peers. Future research could seek to understand how these dynamics unfold in the context of censorship in autocracies.

The literature also indicates that pre-existing beliefs could influence the propensity for users to seek out unfiltered information and resist censorship and propaganda. Many studies have documented humans' inclination for motivated reasoning, i.e., the propensity to believe information that affirms their beliefs and reject information that is contrary to it (Kunda 1987, Taber & Lodge 2006). Theoretical work has shown that even rational individuals might be more likely to believe sources consistent with their beliefs, particularly when they have very little information about the credibility of a media source (Gentzkow & Shapiro 2006, Gehlbach & Sonin 2014).

---

[3] In a study of traditional media, Bleck & Michelitch (2017) report similar findings from a field experiment in Mali after an authoritarian coup. They find that women, who on average are less politically sophisticated than men in Mali, are more susceptible to propaganda from the new junta.

Studies of traditional media suggest that willingness to evade censorship and resist propaganda may be related to the predispositions of those who consume that media. In Ukraine, Peisakhin & Rozenas (2018) study the impacts of access to Russian television on the 2014 election. They find that while pro-Russian propaganda persuaded voters who already held favorable views toward Russia, those who held pro-Western priors were not affected by the propaganda, and in fact the propaganda may have been countereffective for this population. This work reaches similar conclusions to those of Adena et al. (2015), who note that Nazi propaganda was most effective in locations that were already leaning in the direction of that propaganda, and those of DellaVigna et al. (2014), who show that exposure to Serbian propaganda in Croatia increased anti-Serbian sentiment among Croats.

Fewer studies have sought to understand how the predispositions to beliefs might influence resistance to censorship and propaganda online. In a survey experiment conducted in China, Huang (2017) finds that respondents with pro-Western views are more susceptible to online rumors that reflect badly on the Chinese government. Similarly, Allcott & Gentzkow (2017) show that US social media users are more susceptible to false online posts when those posts support the views of their own political party. This is consistent with evidence in political science that partisans are more susceptible to misinformation that confirms their own beliefs (see Flynn et al. 2017 for a review). More work could be done to understand how censorship interacts with predispositions to impact resilience.

IMPLICATIONS OF SELECTIVE CIRCUMVENTION

What implications does the heterogeneity in censorship resilience have for authoritarianism? Authoritarian regimes could benefit from this heterogeneity. Censorship creates different information environments where those who are willing and able to evade censorship have access to a much different array of information from those who are not. Because those with more education and income are typically more able to evade censorship, this could separate elites from potential followers. Research has shown that this separation negatively impacts the propensity for collective action (Steinert-Threlkeld 2017).

Little (2017) develops a model that explains how a divided information environment could thwart coordination. In his model, some "fully credulous" citizens are susceptible to manipulated information and some "partially credulous" citizens sometimes believe it. He finds that even partially credulous citizens may decide not to protest the government when they know it is not performing well, simply because they know that fully credulous citizens will not join them in their action. This is consistent with Wedeen's (1999) study of Syria, which finds that citizens will pretend to believe propaganda because they are unsure about the beliefs of others.

In combination with the discussion in the previous section, these theories point toward the most pernicious combination of attributes of Internet censorship in authoritarian environments. Censorship that is too blunt or too sudden—such as Internet blackouts or sudden blocks of popular websites—might increase demand for evasion and unite the public by providing a focal point for action. But censorship that is subtle enough that only very sophisticated individuals notice it may divide the public while still providing the regime with the benefits of connectivity. This suggests a possible equilibrium where governments tolerate a certain level of censorship evasion, relying on the fact that most users will be susceptible to censorship.

The literature described above goes a long way in explaining why autocrats continue to emphasize political education, even in a digital era. If prior beliefs are predictive of propensity to believe propaganda—or, alternatively, to believe antistate rumors—then political education might make citizens more immune to influence by unfiltered media, if they happen to come

across it. Cantoni et al. (2017) find that changes in textbook material in China between 2004 and 2010 have had an influence on shaping underlying beliefs about politics. This work suggests that political education can have an effect even in a digital age when access to alternative sources of information is widespread.

It also explains the continued use of entertainment content in authoritarian propaganda. If dissident political information mixed with entertainment is dangerous for authoritarian regimes, as the "cute cat theory of censorship" would predict, then a "cute cat theory of propaganda" would predict that propaganda will increasingly mix with entertainment to attract users. Scholars have indeed noted an increase in this mixture in China (Brady 2008, Repnikova & Fang 2019) and Russian propaganda on Twitter (Miller 2019).

CONCLUSION AND FUTURE WORK

The literature reviewed here indicates that Internet users sometimes do, as John Gilmore thought, "view censorship as damage and route around it." Resilience to online censorship depends on awareness of censorship, demand for information that is censored, and the abilities of the Internet user. When censorship is obvious, is linked to content that users demand, or is instituted during time periods of crisis, it can facilitate backlash. Thus, censorship can be a dangerous endeavor for authoritarian regimes.

However, because authoritarian regimes have incentives to disguise censorship and because demand for political information is generally low, routing around censorship overall may be the exception rather than the rule. While a censorship experience that is common to many citizens can create a focal point for revolt, more subtle censorship may only be known to and circumvented by sophisticated users. Thus, subtle, incomplete control over the information environment can work to the autocrat's advantage, dividing the public without creating the perception of information control (Roberts 2018).

These results have implications for future work and policy in this area. First, the literature suggests that efforts to detect, expose, and publicize censorship activities are crucial to resilience to censorship. Civil society, Internet companies, computer scientists, and social scientists have done much great work in this area, and I emphasize the continued importance of revealing censorship and publicizing it as authoritarian strategies change.

Second, it suggests that the dangers of surveillance may include not only more targeted repression of dissidents but also reduced resilience to censorship more broadly. Governments are increasingly able to use data to microtarget censorship, which could reduce the awareness of censorship and the collective experience of censorship. This could make it more difficult for users to understand what others know and do not know and more difficult to coordinate against censorship. More research needs to be done on how authoritarian regimes are using the vast amount of data available on Internet users to microtarget information manipulation.

Third, the heterogeneity in censorship resilience suggests that more work should be done to understand how censorship influences inequality and political polarization in autocratic environments. Despite worries in democratic contexts that the Internet creates filter bubbles (Adamic & Glance 2005), recent work shows that social media platforms may actually expose individuals to a wider breadth of information than previously thought (Barberá et al. 2015). However, little work has been done exploring how censorship might affect these patterns in authoritarian regimes. Can censorship divide the public and polarize it? Or does it create less variation in ideas by exposing users to only a particular type of information? Understanding how censorship affects the consumption of information and the domestic politics of the countries affected will aid in our understanding of the future of authoritarianism.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The author thanks Kennedy Middleton for her excellent research assistance and comments, as well as Seth Hill, Jennifer Pan, Alex Siegel, Zachary Steinert-Threlkeld, David Yang, and Eddie Yang for their comments and feedback.

LITERATURE CITED

Adamic LA, Glance N. 2005. The political blogosphere and the 2004 US election: divided they blog. In *Proceedings of the 3rd International Workshop on Link Discovery*, pp. 36–43. New York: ACM

Adena M, Enikolopov R, Petrova M, Santarosa V, Zhuravskaya E. 2015. Radio and the rise of the Nazis in prewar Germany. *Q. J. Econ.* 130:1885–939

al Masri R. 2017. *In Jordan, the "invisible hand" blocks internet archive*. Rep., Citiz. Lab, Toronto. **https://citizenlab.ca/2017/04/jordan-invisible-hand-blocks-internet-archive/**

Allcott H, Gentzkow M. 2017. Social media and fake news in the 2016 election. *J. Econ. Perspect.* 31:211–36

Ball-Rokeach SJ, DeFleur ML. 1976. A dependency model of mass-media effects. *Commun. Res.* 3:3–21

Bamman D, O'Connor B, Smith N. 2012. Censorship and deletion practices in Chinese social media. *First Monday* 17. **https://doi.org/10.5210/fm.v17i3.3943**

Barberá P, Jost JT, Nagler J, Tucker JA, Bonneau R. 2015. Tweeting from left to right: Is online political communication more than an echo chamber? *Psychol. Sci.* 26:1531–42

Baum MA. 2002. Sex, lies, and war: how soft news brings foreign policy to the inattentive public. *Am. Political Sci. Rev.* 96:91–109

Bennett WL, Segerberg A. 2012. The logic of connective action: digital media and the personalization of contentious politics. *Inf. Commun. Soc.* 15:739–68

Bleck J, Michelitch K. 2017. Capturing the airwaves, capturing the nation? A field experiment on state-run media effects in the wake of a coup. *J. Politics* 79:873–89

Bond RM, Fariss CJ, Jones JJ, Kramer AD, Marlow C, et al. 2012. A 61-million-person experiment in social influence and political mobilization. *Nature* 489:295–98

Boxell L, Steinert-Threlkeld Z. 2019. Taxing dissent: the impact of a social media tax in Uganda. arXiv:1909.04107 [econ.GN]

Brady AM. 2008. *Marketing Dictatorship: Propaganda and Thought Work in Contemporary China*. Lanham, MD: Rowman & Littlefield

Bullock JG. 2009. Partisan bias and the Bayesian ideal in the study of public opinion. *J. Politics* 71:1109–24

Cantoni D, Chen Y, Yang DY, Yuchtman N, Zhang YJ. 2017. Curriculum and ideology. *J. Political Econ.* 125:338–92

Chen J, Xu Y. 2017. Information manipulation and reform in authoritarian regimes. *Political Sci. Res. Methods* 5:163–78

Chen Y, Yang DY. 2019. The impact of media censorship: 1984 or Brave New World? *Am. Econ. Rev.* 109:2294–332

Converse PE. 1962. Information flow and the stability of partisan attitudes. *Public Opin. Q.* 26:578–99

Converse PE. 1964. The nature of belief systems in mass publics. In *Ideology and Discontent*, ed. D Apter, pp. 206–61. New York: Free Press of Glencoe

Davenport C. 2007. State repression and political order. *Annu. Rev. Political Sci.* 10:1–23

Deibert R, Palfrey J, Rohozinski R, Zittrain J. 2011. *Access Contested: Security, Identity, and Resistance in Asian Cyberspace*. Cambridge, MA: MIT Press

Deibert R, Palfrey J, Rohozinski R, Zittrain J, Haraszti M. 2010. *Access Controlled: The Shaping of Power, Rights, and Rule in Cyberspace*. Cambridge, MA: MIT Press

Deibert R, Palfrey J, Rohozinski R, Zittrain J, Stein JG. 2008. *Access Denied: The Practice and Policy of Global Internet Filtering*. Cambridge, MA: MIT Press

DellaVigna S, Enikolopov R, Mironova V, Petrova M, Zhuravskaya E. 2014. Cross-border media and nationalism: evidence from Serbian radio in Croatia. *Am. Econ. J. Appl. Econ.* 6:103–32

Diamond L. 2010. Liberation technology. *J. Democracy* 21:69–83

Downs A. 1957. *An Economic Theory of Democracy*. New York: Harper

Earl J. 2010. The dynamics of protest-related diffusion on the web. *Inf. Commun. Soc.* 13:209–25

Edmond C. 2013. Information manipulation, coordination, and regime change. *Rev. Econ. Stud.* 80:1422–58

Egorov G, Guriev S, Sonin K. 2009. Why resource-poor dictators allow freer media: a theory and evidence from panel data. *Am. Political Sci. Rev.* 103:645–68

Elmer-Dewitt P. 1993. First nation in cyberspace. *Time*, Dec. 6

Enikolopov R, Petrova M, Zhuravskaya E. 2011. Media and political persuasion: evidence from Russia. *Am. Econ. Rev.* 101:3253–85

Epstein R, Robertson RE. 2015. The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. *PNAS* 112:E4512–E4521

Epstein R, Robertson RE, Lazer D, Wilson C. 2017. Suppressing the search engine manipulation effect (SEME). *Proc. ACM Hum.-Comput. Int.* 1(2):42

Flynn D, Nyhan B, Reifler J. 2017. The nature and origins of misperceptions: understanding false and unsupported beliefs about politics. *Political Psychol.* 38:127–50

Francisco RA. 2005. The dictator's dilemma. In *Repression and Mobilization*, ed. C Davenport, H Johnson, C Mueller, pp. 58–81. Minneapolis: Univ. Minn. Press

Fu K, Chan C, Chau M. 2013. Assessing censorship on microblogs in China: discriminatory keyword analysis and the real-name registration policy. *IEEE Internet Comput.* 17:42–50

Geddes B, Zaller J. 1989. Sources of popular support for authoritarian regimes. *Am. J. Political Sci.* 33:319–47

Gehlbach S, Sonin K. 2014. Government control of the media. *J. Public Econ.* 118:163–71

Gehlbach S, Sonin K, Svolik MW. 2016. Formal models of nondemocratic politics. *Annu. Rev. Political Sci.* 19:565–84

Gentzkow M, Shapiro JM. 2006. Media bias and reputation. *J. Political Econ.* 114:280–316

Gerber A, Green D. 1999. Misperceptions about perceptual bias. *Annu. Rev. Political Sci.* 2:189–210

Gohdes AR. 2020. Repression technology: Internet accessibility and state violence. *Am. J. Political Sci.* **https://doi.org/10.1111/ajps.12509**

Goldstone JA, Tilly C. 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*, ed. R Aminzade, J Goldstone, D McAdam, pp. 175–94. Cambridge, UK: Cambridge Univ. Press

Gramlich J. 2019. *10 facts about Americans and Facebook*. News Release, May 16, Fact Tank, Pew Res. Center, Washington, DC. **https://www.pewresearch.org/fact-tank/2019/05/16/facts-about-americans-and-facebook/**

Guriev S, Treisman D. 2019. Informational autocrats. *J. Econ. Perspect.* 33(4):100–27

Hassanpour N. 2014. Media disruption and revolutionary unrest: evidence from Mubarak's quasi-experiment. *Political Commun.* 31:1–24

Hassanpour N. 2016. *Leading from the Periphery and Network Collective Action*. Cambridge, UK: Cambridge Univ. Press

Hill SJ. 2017. Learning together slowly: Bayesian learning about political facts. *J. Politics* 79:1403–18

Hobbs WR, Roberts ME. 2018. How sudden censorship can increase access to information. *Am. Political Sci. Rev.* 112:621–36

Huang H. 2015. Propaganda as signaling. *Comp. Politics* 47:419–44

Huang H. 2017. A war of (mis)information: the political effects of rumors and rumor rebuttals in an authoritarian country. *Br. J. Political Sci.* 47:283–311

Huang H. 2018. The pathology of hard propaganda. *J. Politics* 80:1034–38

Iyengar S. 1990. The accessibility bias in politics: television news and public opinion. *Int. J. Public Opin. Res.* 2:1–15

Iyengar S, Kinder DR. 2010. *News That Matters: Television and American Opinion*. Chicago: Univ. Chicago Press

Jack C. 2017. *Lexicon of lies: terms for problematic information*. Rep., Data and Society Research Institute, New York, NY. **https://datasociety.net/pubs/oh/DataAndSociety_LexiconofLies.pdf**

Jansen S, Martin B. 2015. The Streisand effect and censorship backfire. *Int. J. Commun.* 9:656–71

Jones JJ, Bond RM, Bakshy E, Eckles D, Fowler JH. 2017. Social influence and political mobilization: further evidence from a randomized experiment in the 2012 US presidential election. *PLOS ONE* 12:e0173851

Kern HL, Hainmueller J. 2009. Opium for the masses: how foreign media can stabilize authoritarian regimes. *Political Anal*. 17:377–99

King G, Pan J, Roberts ME. 2013. How censorship in China allows government criticism but silences collective expression. *Am. Political Sci. Rev.* 107:1–18

King G, Pan J, Roberts ME. 2014. Reverse-engineering censorship in China: randomized experimentation and participant observation. *Science* 345:1251722

King G, Pan J, Roberts ME. 2017a. How the Chinese government fabricates social media posts for strategic distraction, not engaged argument. *Am. Political Sci. Rev.* 111:484–501

King G, Schneer B, White A. 2017b. How the news media activate public expression and influence national agendas. *Science* 358:776–80

Knockel J, Ruan L, Crete-Nishihata M. 2017. *We (can't) chat: "709 crackdown" discussions blocked on Weibo and WeChat*. Rep., Citiz. Lab, Toronto. **https://citizenlab.ca/2017/04/we-cant-chat-709-crackdown-discussions-blocked-on-weibo-and-wechat/**

Kunda Z. 1987. Motivated inference: self-serving generation and evaluation of causal theories. *J. Personal. Soc. Psychol.* 53:636–47

Kuran T. 1989. Sparks and prairie fires: a theory of unanticipated political revolution. *Public Choice* 61:41–74

Kuran T. 1991. Now out of never: the element of surprise in the East European revolution of 1989. *World Politics* 44:7–48

Lapowsky I. 2018. House Democrats release 3,500 Russia-linked Facebook ads. *Wired*, May 10. **https://www.wired.com/story/house-democrats-release-3500-russia-linked-facebook-ads**

Larreguy H, Marshall J. 2019. The incentives and effects of independent and government-controlled media in the developing world. In *The Oxford Handbook of Electoral Persuasion*, ed. E Suhay, B Grofman, A Trechse. Oxford: Oxford Univ. Press. **https://doi.org/10.1093/oxfordhb/9780190860806.013.13**

Lichbach MI. 1987. Deterrence or escalation? The puzzle of aggregate studies of repression and dissent. *J. Confl. Resolut.* 31:266–97

Little AT. 2016. Communication technology and protest. *J. Politics* 78:152–66

Little AT. 2017. Propaganda and credulity. *Games Econ. Behav.* 102:224–32

Lohmann S. 1994. The dynamics of informational cascades: the Monday demonstrations in Leipzig, East Germany, 1989–1991. *World Politics* 47:42–101

Lorentzen P. 2014. China's strategic censorship. *Am. J. Political Sci.* 58:402–14

Loveless M. 2008. Media dependency: mass media as sources of information in the democratizing countries of Central and Eastern Europe. *Democratization* 15:162–83

MacKinnon R. 2012. *Consent of the Networked: The Worldwide Struggle for Internet Freedom*. New York: Basic Books

Marshall J. 2019. Signaling sophistication: how social expectations can increase political information acquisition. *J. Politics* 81:167–86

McCombs ME, Shaw DL. 1972. The agenda-setting function of mass media. *Public Opin. Q.* 36:176–87

McMillan J, Zoido P. 2004. How to subvert democracy: Montesinos in Peru. *J. Econ. Perspect.* 18:69–92

Miller B. 2018. *The limits of commercialized censorship in China*. Work. Pap., London School of Economics, London. **http://eprints.lse.ac.uk/101417/1/miller_limits_of_commercialized_censorship_in_china.pdf**

Miller DT. 2019. Topics and emotions in Russian Twitter propaganda. *First Monday* 24. **https://firstmonday.org/ojs/index.php/fm/article/view/9638**

Mirani L. 2015. Millions of Facebook users have no idea they're using the internet. *Quartz*, Feb. 9. **https://qz.com/333313/milliions-of-facebook-users-have-no-idea-theyre-using-the-internet/**

Moore WH. 1998. Repression and dissent: substitution, context, and timing. *Am. J. Political Sci.* 42:851–73

Morozov E. 2011. *The Net Delusion: The Dark Side of Internet Freedom*. New York: PublicAffairs

Munger K, Bonneau R, Nagler J, Tucker JA. 2019. Elites tweet to get feet off the streets: measuring regime social media strategies during protest. *Political Sci. Res. Methods* 7(4):815–34

Nabi Z. 2014. ~~Resistance~~ censorship is futile. *First Monday* 19. **https://firstmonday.org/article/view/5525/4155**

Norris P. 2001. *Digital Divide: Civic Engagement, Information Poverty, and the Internet Worldwide*. Cambridge, UK: Cambridge Univ. Press

Pan J. 2016. How market dynamics of domestic and foreign social media firms shape strategies of internet censorship. *Problems Post-Communism* 64:167–88

Pan J, Roberts ME. 2020. Censorship's effect on incidental exposure to information: evidence from Wikipedia. *SAGE Open* 10. **https://doi.org/10.1177/2158244019894068**

Pan J, Siegel A. 2020. How Saudi crackdowns fail to silence online dissent. *Am. Political Sci. Rev.* 114(1):109–25

Peisakhin L, Rozenas A. 2018. Electoral effects of biased media: Russian television in Ukraine. *Am. J. Political Sci.* 62:535–50

Petrova M. 2011. Newspapers and parties: how advertising revenues created an independent press. *Am. Political Sci. Rev.* 105:790–808

Popkin S. 1994. *The Reasoning Voter: Communication and Persuasion in Presidential Campaigns*. Chicago: Univ. Chicago Press

Qin B, Strömberg D, Wu Y. 2017. Why does China allow freer social media? Protests versus surveillance and propaganda. *J. Econ. Perspect.* 31:117–40

Repnikova M, Fang K. 2019. Digital media experiments in China: "revolutionizing" persuasion under Xi Jinping. *China Q.* 239: 679–701

Roberts ME. 2018. *Censored: Distraction and Diversion Inside China's Great Firewall*. Princeton, NJ: Princeton Univ. Press

Roberts ME, Stewart BM, Nielsen RA. 2020. Adjusting for confounding with text matching. *Am. J. Political Sci.* In press

Sanovich S, Stukal D, Tucker JA. 2018. Turning the virtual tables: government strategies for addressing online opposition with an application to Russia. *Comp. Politics* 50:435–82

Shadmehr M, Bernhardt D. 2011. Collective action with uncertain payoffs: coordination, public signals, and punishment dilemmas. *Am. Political Sci. Rev.* 105:829–51

Shadmehr M, Bernhardt D. 2015. State censorship. *Am. Econ. J. Microecon.* 7:280–307

Shahbaz A. 2018. *Freedom on the net 2018: the rise of digital authoritarianism*. Rep., Freedom House, Washington, DC. **https://freedomhouse.org/report/freedom-net/freedom-net-2018/rise-digital-authoritarianism**

Steinert-Threlkeld ZC. 2017. Spontaneous collective action: peripheral mobilization during the Arab spring. *Am. Political Sci. Rev.* 111:379–403

Stern RE, Hassid J. 2012. Amplifying silence: uncertainty and control parables in contemporary China. *Comp. Political Stud.* 45:1230–54

Stockmann D. 2012. *Media Commercialization and Authoritarian Rule in China*. Cambridge, UK: Cambridge Univ. Press

Stukal D, Sanovich S, Bonneau R, Tucker JA. 2017. Detecting bots on Russian political Twitter. *Big Data* 5:310–24

Taber CS, Lodge M. 2006. Motivated skepticism in the evaluation of political beliefs. *Am. J. Political Sci.* 50:755–69

Tanash R, Chen Z, Wallach D, Marschall M. 2017. *The decline of social media censorship and the rise of self-censorship after the 2016 failed Turkish coup*. Paper presented at FOCI '17, 7th USENIX Workshop on Free and Open Communications on the Internet, Vancouver, BC, Aug. 14. **https://www.usenix.org/conference/foci17/workshop-program/presentation/tanash**

Tucker JA, Guess A, Barberá P, Vaccari C, Siegel A, et al. 2018. *Social media, political polarization, and political disinformation: a review of the scientific literature*. Rep., William and Flora Hewlett Found., Menlo Park, CA. **https://hewlett.org/wp-content/uploads/2018/03/Social-Media-Political-Polarization-and-Political-Disinformation-Literature-Review.pdf**

Tufekci Z. 2014. Engineering the public: big data, surveillance and computational politics. *First Monday* 19. **https://firstmonday.org/article/view/4901/4097**

Verkamp JP, Gupta M. 2013. *Five incidents, one theme: Twitter spam as a weapon to drown voices of protest*. Paper presented at FOCI '13, 3rd USENIX Workshop on Free and Open Communications on the Internet, Washington, DC, Aug. 13. **https://www.usenix.org/conference/foci13/workshop-program/presentation/verkamp**

Wedeen L. 1999. *Ambiguities of Domination: Politics, Rhetoric, and Symbols in Contemporary Syria*. Chicago: Univ. Chicago Press

Weidmann NB, Benitez-Baleato S, Hunziker P, Glatz E, Dimitropoulos X. 2016. Digital discrimination: political bias in internet service provision across ethnic groups. *Science* 353:1151–55

Weidmann NB, Rød EG. 2019. *The Internet and Political Protest in Autocracies*. Oxford, UK: Oxford Univ. Press

Yanagizawa-Drott D. 2014. Propaganda and conflict: evidence from the Rwandan genocide. *Q. J. Econ.* 129:1947–94

Young LE. 2019. The psychology of state repression: fear and dissent decisions in Zimbabwe. *Am. Political Sci. Rev.* 113:140–55

Zaller JR. 1992. *The Nature and Origins of Mass Opinion*. Cambridge, UK: Cambridge Univ. Press

Zuckerman E. 2015. Cute cats to the rescue? Participatory media and political expression. In *From Voice to Influence: Understanding Citizenship in a Digital Age*, ed. D Allen, JS Light, pp. 131–54. Chicago: Univ. Chicago Press