Digital Surveillance and Self-Censorship in Autocracies: Evidence from a Survey Experiment in Kazakhstan

David Karpa[1]

[1]Munich School of Politics and Public Policy, Technical University of Munich, Richard-Wagner-Straße 1, Munich, 80333, Germany.

Contributing authors: davidfkarpa@gmail.com;

Abstract

Digital surveillance technologies are increasingly employed, especially in authoritarian regimes seeking to monitor and shape online communication. Yet we know little about how such surveillance affects citizens' willingness to express political views. Theory suggests that awareness of being surveilled induces self-censorship, discouraging individuals from voicing opinions on sensitive topics. This paper tests this proposition using a survey experiment conducted in Kazakhstan in November 2023 (N = 5,025). Participants were randomly exposed to a reminder of government surveillance, an assurance of privacy, or a control condition before answering sensitive and non-sensitive questions. Exposure to the surveillance reminder reduced responses to sensitive items by about three percentage points, while the privacy assurance had no effect. The effect is strongest among respondents who consume foreign media, suggesting that politically informed individuals are more responsive to surveillance cues. These findings provide experimental evidence that perceived surveillance discourages political expression and reinforces authoritarian stability.

Keywords: information control; surveillance; privacy; political repression; media; autocracy

1 Introduction

Over 75 countries worldwide use surveillance tools that are associated with artificial intelligence, including over 50% of advanced democracies (Feldstein 2019a). For example, Ethiopia, with its long-standing network of in-person surveillance, was a quick adopter and transitioned to digital surveillance despite initially having a low percentage of the population with access to the internet (Feldstein 2021). Many have raised the need to critically reflect on surveillance practices in contemporary societies, because of ongoing human rights violations.[1] Beyond ethical and human rights concerns, mass surveillance has been shown to have an effect on human behavior by undermining autonomy and well-being, and inducing self-censorship (Büchi et al. 2022). Surveillance practices lead to a "spiral of silence", where people are deterred from exchanging opinions (online), particularly concerning sensitive topics (Stoycheff 2016). The rise of pre-emptive and conformist behavior is in direct conflict with the essential components of deliberative democratic frameworks and represents a significant challenge to the healthy functioning of participatory societies (Penney 2022; Kappeler et al. 2023).

In an autocratic context, anticipatory and conformist behaviors are not a side-product but the main means to secure power. Many of the long ruling autocrats like Russia's Putin or Turkey's Erdoğan have increasingly resorted to using violence on protesters, repressing dissidents, and imprisoning journalists, as a means to consolidating power (Pan and Siegel 2020; Egorov and Sonin 2024). Once feared, dictators strategically signal their surveillance and repression capabilities in order to enforce self-disciplining behavior (Gohdes 2023). This self-disciplining behavior can come in many forms, but first and foremost, it results in self-censorship concerning political topics (Roberts 2018). Surveillance thus contributes to undermining collective action, to democratic backsliding, and to authoritarian stability (Carothers and Press 2022). However, some scholars have argued that autocrats refrain from directly repressing their population because of its net negative consequences (Guriev and Treisman 2019), and instead try to control the informational environment by co-opting the elite and media (Guriev and Treisman 2020). Here, manipulation control is a *substitute* mass repression, whereas others argue that repression and information control *complement* each other (Gehlbach et al. 2022; Lamberova and Sonin 2023; Gohdes 2023).

Yet, how perceptions of surveillance translate into concrete behavioral responses remains largely assumed rather than empirically demonstrated. The main aim of this paper is thus to provide an empirical assessment of how surveillance practices shape individual behavior, specifically whether they induce self-censorship among citizens, to what magnitude, and under what conditions. To this end, a survey experiment with 5,025 participants was conducted in Kazakhstan in November 2023. Kazakhstan is a country where the government has repeatedly deployed mass surveillance technology at the internet service provider level (Raman et al. 2020), where the press is mostly state-controlled (FreedomHouse 2024), and where targeted repression against journalists and dissidents is common.[2] Kazakhstan is hence an ideal environment

[1] Spyware and surveillance: Threats to privacy and human rights growing, UN report warns

[2] See, for example: Amnesty International Kazakhstan Report 2023, Human Rights Watch Kazakhstan 2023 Report

to investigate how citizens adapt their expressive behavior in the face of pervasive surveillance.

Participants in the study were asked sensitive questions on domestic and geopolitical topics after exposure to either a control, surveillance, or privacy condition.[3] The main results show that reminders of government surveillance increase self-censorship by an average of 3.3 percentage points.[4] This effect is not symmetrical: experimentally increasing perceived privacy had no statistically detectable impact. Importantly, the effect of surveillance cues varies across citizens. Respondents who regularly consume foreign media – the more informed elite – are more responsive to surveillance reminders (6.1%) than those without such exposure (0.9%). While these citizens are generally more willing to express political opinions, their heightened awareness of politically sensitive topics makes them more cautious when reminded of potential surveillance.

This study contributes to research on the behavioral consequences of perceived surveillance, showing how subtle reminders of monitoring can alter citizens' willingness to voice opinions. More broadly, it advances theoretical understanding of how perceptions of surveillance shape political behavior in digitalized societies, following the call of Büchi et al. (2022). In addition, it contributes to public opinion research by quantifying the extent to which self-censorship may lead to an overestimation of politically desirable attitudes in autocracies (Corstange 2012; Frye et al. 2017, 2023; Robinson and Tannenberg 2019; Tannenberg 2022). Finally, it adds to the literature on authoritarianism by showing how perceptions of surveillance can sustain control over the informational environment (King et al. 2017; Roberts 2018; Guriev and Treisman 2019, 2020; Feldstein 2021; Gohdes 2023; Egorov and Sonin 2024). The following section reviews the relevant literature from which the hypotheses are derived. Section 3 embeds the hypotheses in the research design and details the methodology. Section 4 presents the findings, and Section 5 concludes with their broader implications.

2 Literature

Social scientists who study digital surveillance sometimes call it *covert repression* (Earl et al. 2022), *dataveillance* (Festic 2022; Büchi et al. 2022; Kappeler et al. 2023; Lee 2023), *fear-based censorship* (Roberts 2018, 2020), or embed it into a broader discussion of *digital authoritarianism* (Feldstein 2019b, 2021; Jones 2022; Gohdes 2023). The literature distinguishes between research on digital surveillance in different types of regimes, because there is an important difference. In theory, government surveillance in democracies is an unintended side effect, a necessary evil of anti-terror or COVID measures. Independent institutions are supposed to monitor each other and keep power in check to protect civil liberties and individual rights. In the literature on autocracies, surveillance is a crucial tool in the state's repertoire of survival strategies, to the extent that it is strategically signaled to the population (Roberts 2018; Gohdes 2023). Accordingly, research on digital surveillance in autocracies tends to understand it as a form of state repression strategically deployed by autocrats to stay in power. This research is complemented by a political economy perspective that focuses on

[3] Nazarbayev University Institutional Research Ethics Committee (NU-IREC) reviewed and approved the experiment (771/25092023).
[4] 4.2% on issues concerning domestic politics and between 2.5 and 3.1% on geopolitical topics.

the mutual benefits of a private-public partnership in the development of surveillance technologies in autocracies (Liu 2019; Beraja et al. 2022, 2023b,a; Huang et al. 2022).

2.1 Repression, fear and chilling effects

The importance of surveillance in authoritarian states can also be explained by the information dilemma of the authoritarian government. As a result of censorship, media control, and the absence or manipulation of elections, the regime does not know the true sentiments of its citizens (Edmond 2013; Xu 2021; Egorov and Sonin 2024). Consequently, the efficient allocation of resources to co-opt regime opponents remains impossible, as the regime is uncertain about which actors require co-optation and which actors can be better controlled through repression. Such targeted co-optation or repression is necessary, however, because large-scale mass repression is rarely used in contemporary dictatorships (Guriev and Treisman 2019; Xu 2021), partly because of the disadvantages of international backlash in a globalized economy, but also because visible repression can signal regime weakness (Guriev and Treisman 2020). Surveillance of social media helps to identify protests early and monitor local governments and officials (Qin et al. 2017).

When dissidents were identified through surveillance, targeted repression of regime dissidents discourages and deters the participation of larger segments of the population (Roberts 2018; Xu 2021; Gohdes 2023). In autocracies, political expression and discussion are possible but very limited (King et al. 2017). By *taxing information* through propaganda, distraction, and censorship, free debate on political issues is hindered (Roberts 2018). Thus, political participation takes the form of protests or revolts because of the absence of meaningful elections and the censorship of grievances. More surveillance can lead to more repression since the authorities can act on the collected information (Earl et al. 2022). In sum, there are two functions of surveillance: (1) it enables targeted repression by increasing the information available to regimes, and (2) it signals repressive capacities that induce *fear*, leading to self-censorship (Roberts 2018).

In the discourse on surveillance in democracies, a related phenomenon has been referred to as *chilling effect*. Chilling effects – the deterrence of lawful behavior out of fear that it is suspect – have been studied by several scholars (Schauer 1978; Penney 2016, 2017; Stoycheff 2016; Stoycheff et al. 2019; Büchi et al. 2022). The core of democracy can be considered to be the freedom to hold and express any political views. The discussion of political issues has increasingly moved to online spaces such as social media and text messengers, and while in online environments these expressions and debates of political opinion are vulnerable to surveillance. Theoretical studies of digital surveillance argue that *salience shocks*[5] of digital surveillance lead to inhibited digital communication behavior (Büchi et al. 2022). Recent research has suggested a common denominator in research on surveillance in autocracies and democracies: surveillance induces self-discipline (mostly self-censorship) due to the *fear* of repression (Roberts 2018; Manokha 2018; Tannenberg 2022; Stoycheff 2022; Oz and Yanik 2022). Citizens – when aware of surveillance practices – have an increased expectation of negative outcomes and will self-censor. In this vein, the first hypothesis is formulated as:

[5] One such shock was Edward Snowden's revelations about the NSA's ongoing surveillance of US citizens.

Hypothesis 1: The perception of digital surveillance induces self-censorship in politically sensitive topics.

2.2 Mass surveillance in Kazakhstan

Kazakhstan is a resource-rich Central Asian country bordering China and Russia. After the collapse of the Soviet Union, of which Kazakhstan was a part, the country gained independence and was ruled authoritatively for nearly three decades by former Party Secretary Nursultan Nazarbayev. Nazarbayev followed the model of the modern autocrat of the late 20th century, who didn't oppress his people with brutal force, but rather told the story of a man of the people while ensuring an acceptable minimum of living conditions (Guriev and Treisman 2019). In 2019, the country's leadership changed as Nazarbayev appointed a predecessor, Kassym-Jomart Tokayev. While this transition of power was initially successful, Tokayev eventually struggled with perceptions of illegitimacy (Kudaibergenova and Laruelle 2022; Silvan 2024). Growing protests culminated in the so-called "Bloody January" of 2022 – mass protests against corruption and economic inequality on an unprecedented scale were followed by a state of emergency and fighting between the military and protesters, with thousands arrested and hundreds killed (FreedomHouse 2023a). There have been reports of torture of protesters, activists, and journalists.[6]

The government has broad powers to control the digital infrastructure, deriving its authority from laws and weak legal resistance. From controlling the content of websites through legal pressure to outright blocking of websites, to punishing journalists, there is widespread censorship (FreedomHouse 2023b). In addition, laws make anonymity online impossible, VPNs are cracked down on, and SIM cards – the access point to the internet for most of the population – must be registered with an ID. In 2019, Kazakhstan became the first country to force its population to install a custom root certificate capable of decrypting content running through the country's largest internet service provider. These surveillance capabilities have primarily targeted social media and communications services, making them seemingly a political rather than a security endeavor (Raman et al. 2020). While the root certificate was only active for about three weeks, it set a precedent and signaled the government's capabilities to the population. In addition to mass surveillance on the internet service provider level, government agencies monitor social media and communication apps targeting journalists, dissidents, and minorities (FreedomHouse 2023b). All this potentially culminates in self-censorship on a large scale, especially when it comes to the two most important political issues – the "Bloody January" and Russia's invasion of Ukraine. Leveraging this context enhances the study's external validity, as the conditions of surveillance, repression, and self-censorship – prevalent in Kazakhstan – can be authentically simulated within an experimental framework.

Other studies suggest that behavioral adaptations to surveillance include increased use of privacy-preserving technologies to cope with surveillance (Büchi et al. 2022; Kappeler et al. 2023). Büchi et al. (2022) model *recovery potential* for inhibited digital communication behavior occurs over time as the salience of surveillance practices diminishes. Others found that censorship in the form of blocked websites is being

[6] Human Rights Watch: Longing for Justice in Kazakhstan

bypassed with circumvention tools, leading to renewed access by citizens and increased interest in blocked content (Hobbs and Roberts 2018). Sparse empirical work further highlights that privacy-enhancing technologies (PETs) and anonymity tools can mitigate self-censorship by providing users with a protected space for expression: a recent study finds that online anonymity is positively associated with individuals' willingness to express true opinions and engage in discussions on sensitive topics, demonstrating that PETs can foster more open communication even in the presence of potential risks (Prajith and Bhuyan 2025). Moreover, research on voice anonymization in civic dialogue platforms shows that anonymizing participants' voices increases their willingness to express themselves freely, reinforcing the idea that privacy technologies can empower users to overcome the chilling effects of surveillance (Kang et al. 2024).[7]

Building on this literature, this study theorizes that effective encryption mechanisms should *recover* digital communication behavior. Given the baseline of digital surveillance in contemporary societies, particularly in Kazakhstan (Raman et al. 2020), the potential for recovering digital communication behavior is significant. Hence, this study proposes that the recovery of digital communication behavior can not only occur on its own over time (as in Büchi et al. (2022)), but also immediately through the use of PETs. Specifically, the theoretic expectation is that citizens may believe the negative consequences of surveillance can be reduced by adopting privacy-enhancing technologies. Correspondingly, the second hypothesis proposes that:

Hypothesis 2: The perception of privacy-enhancing technology reduces self-censorship in politically sensitive topics.

A literature that studies the effects of media and propaganda in autocracies highlights the importance of controlling the informational environment in order to control the population (Enikolopov et al. 2011; Adena et al. 2015; Zhuravskaya et al. 2020; Guriev and Treisman 2022). While some theories interpret propaganda and repression as *substitutes* (Guriev and Treisman 2019, 2020), recent developments suggest that such a substitution is incomplete. The level of information control through propaganda and censorship has remained high or even increased, while the expected decline in coercive repression has not materialized. Instead of trying to decide between these theoretical positions, this study looks at the consequences of perceived surveillance on behavior. Specifically, it examines how reminders of monitoring influence citizens' willingness to express their opinions.

Theoretical work nonetheless provides useful insights into who may be most sensitive to surveillance. Models of authoritarian control argue that citizens learn from visible signals of censorship or repression and adjust their behavior accordingly (Gehlbach et al. 2022; Gohdes 2023; Egorov and Sonin 2024). Citizens differ in how they interpret and react to such surveillance signals. Those with access to foreign media or higher education tend to follow politics more closely and hold more differentiated, and sometimes more critical, opinions about the government. Because they

[7] The UN Special Rapporteur on freedom of expression explicitly connects encryption and anonymity to reduced self-censorship, arguing that such technologies "create a zone of privacy to protect opinion and belief," especially in environments where individuals might otherwise self-censor due to fear of surveillance or repression (Kaye 2015).

are better informed, they also have a clearer sense of which topics are politically sensitive and how the state responds to dissent. Their exposure to alternative information sources and critical perspectives enables them to recognize when a question touches upon politically charged or risky subjects. Precisely because they understand the sensitivity of these topics, such individuals may exercise greater caution when expressing their views, anticipating that their responses could be observed or misinterpreted. In other words, knowledge sharpens perception of danger: people who better understand the political environment also better recognize when silence is safer. As a result, well-informed citizens may self-censor not because they are inherently more fearful, but because they can more accurately gauge the boundaries of acceptable expression and the potential costs of crossing them. Less informed citizens, by contrast, may not perceive these boundaries as clearly and therefore feel fewer reasons to withhold their answers. This mechanism links education and information exposure to self-censorship through awareness of political sensitivity and perceived personal risk. In Kazakhstan, where repression is often targeted toward high-profile individuals such as journalists or activists (FreedomHouse 2024), such heightened awareness among informed citizens is particularly plausible.[8]

Accordingly, the following hypotheses are proposed:

Hypothesis 3: The effect of surveillance reminders on self-censorship will be stronger among citizens who consume foreign media.

Hypothesis 4: The effect of surveillance reminders on self-censorship will be stronger among citizens with higher education.

This study draws on this literature to examine (1) whether surveillance induces self-censorship, (2) whether this effect can be mitigated by a privacy-preserving treatment, and (3) which citizens are most responsive to reminders of surveillance.

3 Method and data

To test the hypotheses, an online survey experiment with 5,025 respondents was conducted in November 2023 in Kazakhstan. The survey was pre-registered[9] and carried out by NAC Analytica, a leading Kazakh sociological and public opinion research organization.[10] Participants were recruited through advertisements in social media, and a weighting-scheme was applied to make the sample nationally representative.

Before being randomly assigned to either a control group or one of the two treatment conditions, participants answered a range of demographic questions. The treatments were text-based information on the security of participants data. The treatment conditions differ with the control condition in that they either point out the possibility of the government being able to access information on online activity (surveillance condition) or ensure confidentiality by encryption (privacy condition).

[8] "Media independence is severely limited in Kazakhstan. [...] Independent outlets and journalists are routinely shut down or harassed, and self-censorship is common", see: Freedom in the World Report Kazakhstan 2024.

[9] https://aspredicted.org/BVT_9Z3. See also appendix A.4 for a discussion of the pre-analysis plan and divergences from that.

[10] https://nacanalytica.com/en/

Section A.1 in the Appendix presents the control and treatment scenarios. The control condition consists only of a standard experimental instruction without additional information.

After having faced either treatment, participants were asked four questions in random order, three of which are politically sensitive, and one that is not sensitive and acts as a placebo. The sensitive questions concerned domestic politics (*In your opinion, is participating in protests for political change generally justified or not justified?*) and geopolitics (*In your opinion, is helping Russia avoid Western sanctions generally justified or not justified?* and *In your opinion, is Russia's Special Military Operation/ invasion of Ukraine generally justified or not justified?*). The framing *Special Military Operation* and *invasion of Ukraine* was assigned at random, in order to balance invoked framing effects. Arguably, the way one describes Russia's invasion of Ukraine gives away their view on this war and thus invokes demand effects and social desirability bias. A neutral stance between the two mutually exclusive narratives of an illegitimate invasion or a "Special Military Operation" is hard to find. Question 4 acted as a placebo, in order to control for design effects (*In your opinion, is working more than 50 hours per week generally justified or not justified?*). Answer options for the outcome variables were *Justified*, *Not justified*, and *Prefer not to answer*.

Quality controls included attention checks (two questions on respondents age had to match), speeding filters (minimum of 200 seconds), allowing only two completes per IP address, and allowing phone numbers to participate only once (payment was carried out by phone number). Out of 28,201 participants, 5,025 completed the survey, passed quality checks, were unique respondents, and were compensated 700 Tenge (approx. 1.50 USD).[11] 25 respondents left the experiments after having faced the control (7), surveillance (7) or privacy (11) condition, respectively. Most of the participants that left the survey before finishing did so in the very first pages of the survey.

Table A1 presents summary statistics for all variables. Categorical variables were transformed to scales or dummies. The sample was 48.7% male and 42.7 years old (SD=16.1), on average. Participants were asked on a 1 - 5 scale about their financial situation (M=2.85, SD=1.14), with the mean corresponding to the answer option *We have enough money for food and clothes, but buying durable goods, such as a TV or refrigerator, is difficult.* 23.2 % of participants reported having received higher education (SD=0.42),[12] their residency (where 22% (SD=0.41) reported living in either of the two large cities Astana or Almaty), and being ethnically Kazakh (M=0.72, SD=0.45). 32.6% (SD=0.47) of participants reported consuming news sources from abroad.[13] 10.9% (SD=0.31) of the participants work for some government organization

[11] Figure A1 shows when and how participants left the survey.

[12] Figure A4 shows the distribution of responses.

[13] Figure A2 and A3 show the distribution of responses regarding media sources and their origin. Respondents were first asked to identify up to four main sources from which they learn about events outside Kazakhstan (e.g., television, websites, social media platforms such as YouTube, Telegram, or Instagram). For each selected source, follow-up questions recorded (1) frequency of use, (2) specific channels or groups, and (3) the country of origin of those outlets. If a respondent indicated that any of the channels, groups, or websites they follow are based outside Kazakhstan (e.g., Russia, Western countries, Turkey, or China), the binary variable was coded as 1, and 0 otherwise. This measure therefore distinguishes between respondents who are exclusively exposed to domestic information sources and those who consume at least some foreign media content. Qualitative assessment of the Russian channel names participants indicated revealed that roughly one third of channels are explicitly opposition channel.

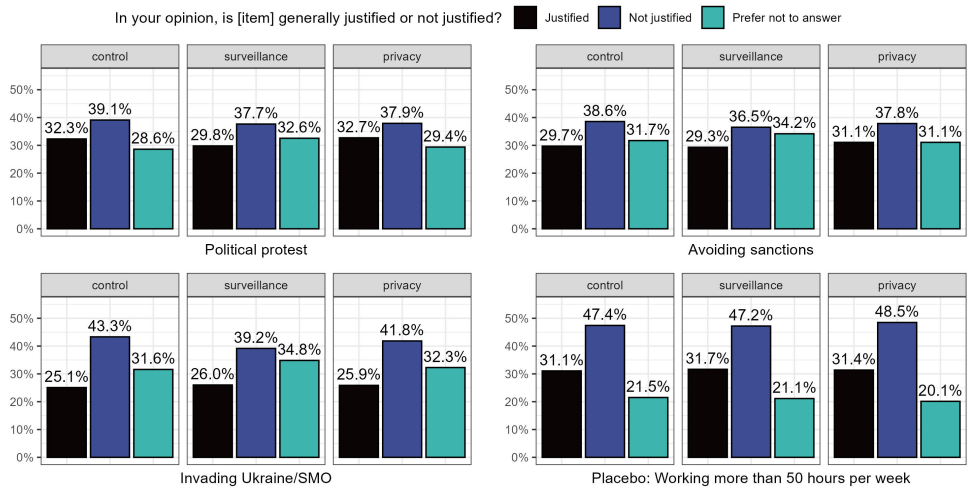
and 10.9% (SD=0.31) use a virtual private network (vpn). Participants were asked about the number of children they have (M=2.26, SD=1.46). Weights for age and sex were applied in regression models to make the sample nationally representative.

4 Results

4.1 Average treatment effects

Figure 1 shows the proportion of responses in percent by treatment condition. For the first item, *participating in protests for political change* 32.3% responded *justified*, 39.1% *not justified*, and 28.6% *prefer not to answer*. In the surveillance treatment, these numbers changed by -2.51%, -1.44% and 3.95% and in the privacy treatment by 0.41%, -1.19%, and 0.78%, respectively. The second item, *helping Russia to avoid Western sanctions* has a justification rate of 29.72%, whereas 38.57% responded *not justified*, and 31.71% *prefer not to answer*. In the surveillance treatment, these numbers changed by -0.41%, -2.05% and 2.46% and in the privacy treatment by 1.36%, -0.72%, and -0.63%, respectively. The third item, whether *Russia's Special Military Operation/ invasion of Ukraine* was justified, found 25.09% of supporters, whereas 43.32% responded *not justified*, and 31.59% *prefer not to answer*. In the surveillance treatment, these numbers changed by 0.92%, -4.16% and 3.24% and in the privacy treatment by 0.79%, -1.48%, and 0.7%, respectively. Generally, self-censorship was the lowest in the question revolving around avoiding sanctions, and the highest in the question corresponding to domestic politics.

Fig. 1 Responses to dependent variables



In order to assess whether these differences are meaningful statistical deviations, i.e., whether treatments have significant effects, linear regression models were estimated with the response option *prefer not to answer* as dependent variables, and treatment dummies as dependent variables. Weights for age and gender were applied

to make the sample representative of the population. A treatment effect, as defined here, is the increase or decrease in the response option *prefer not to answer* to a sensitive question. Table 1 shows the average treatment effects (ATE) resulting from these models in column 1.[14]

The surveillance treatment results in an increase in the *prefer not to answer* option for sensitive items by 3.3% on average being statistically significant at the 99% level. This increase provides some evidence in favor of the first hypothesis, that is, participants are more likely to self-censor when reminded of governmental surveillance capacities. The privacy condition does not yield a measurable effect (0.3%) and is not significant, hence not providing evidence in favor of the second hypothesis. Priming participants for an increase in privacy does not motivate them to increasingly share their opinions, on average. The privacy interactions (columns 3 and 4) – as well as the simple privacy dummy – have no measurable effect. The privacy treatment does not appear to have a detectable effect. This is consistent with the other specifications and does not support the acceptance of hypothesis 2.

4.2 Conditional average treatment effects

The remaining two hypotheses concern the question of *who* self-censors. As laid out in section 2, predictions from the literature suggest that an informed elite self-censors. This is because they have a) more complete information, and b) awareness of divergence of this information from other public information. Following this logic, hypothesis 3 posits that citizens who consume media from abroad are more likely to self-censor. Turning to results in table 1 can inform this hypothesis. In column 2, the overall association between international media consumption and self-censorship is negative (-12%) and statistically significant when controlling for the full set of control variables. It seems that, on average, consuming international media informs citizens in a way that allows them to answer questions more easily, which is somewhat intuitive. On the surface this seems to oppose the expected effect formulated in hypothesis 3.

While citizens who consume international media are generally less likely to self-censor – perhaps because broader exposure equips them with more information and confidence to answer sensitive questions – the surveillance treatment alters this relationship. The positive and significant interaction term in column 3 indicates that the effect of the surveillance reminder is stronger among these respondents: they reduce their willingness to answer sensitive items more than those who rely only on domestic media. In other words, well-informed citizens are not generally more prone to self-censor, but they are more sensitive to reminders of surveillance. This aligns with the idea that awareness of surveillance and sensitivity to political risks increase with information and sophistication. An alternative interpretation is that citizens with lower media exposure already self-censor at high baseline levels, leaving little room for the surveillance cue to further increase nonresponse – a ceiling effect that this design

[14] Table A2 presents the same models for each dependent variable separately. While the placebo question shows no treatment effects, the disaggregated results indicate that the overall surveillance effect is primarily driven by increased non-response to the protest question. Treatment effects for the sanction evasion question are statistically insignificant, and the effect for the invasion question is only marginally significant at the 10% level. This pattern suggests that the surveillance reminder mainly affected willingness to answer the protest question, which is also the most politically sensitive item in the set.

Table 1 Linear regression models

	"Prefer not to answer"			
	(1)	(2)	(3)	(4)
Privacy	0.003	−0.006	−0.011	−0.013
	(0.009)	(0.009)	(0.010)	(0.010)
Surveillance	0.033**	0.027**	0.009	0.023*
	(0.009)	(0.009)	(0.010)	(0.010)
International Media		−0.120**	−0.142**	−0.120**
		(0.009)	(0.014)	(0.009)
Higher Education		−0.057**	−0.057**	−0.072**
		(0.009)	(0.009)	(0.015)
Privacy × International Media			0.013	
			(0.019)	
Surveillance × International Media			0.053**	
			(0.019)	
Privacy × Higher Education				0.029
				(0.021)
Surveillance × Higher Education				0.016
				(0.021)
Region FE	No	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes
Observations	15,075	15,075	15,075	15,075
R^2	0.002	0.187	0.187	0.189
Adjusted R^2	0.001	0.184	0.184	0.187

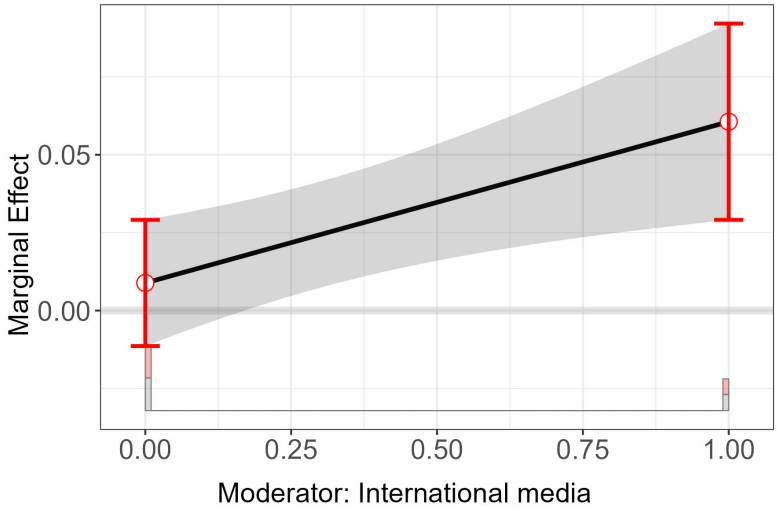
Notes: The dependent variable is answering *prefer not to answer* to a sensitive question. Robust standard errors in parentheses. Weighting was applied. The list of control variables includes: age group, gender, financial situation, city size, VPN usage, ethnicity, number of children, Russian language proficiency, government employment status, region, religious affiliation, and consumption of Kazakh media.
$^{+}$p<0.1; *p<0.05; **p<0.01

cannot fully disentangle. Accordingly, Figure 2 should be read as showing that the marginal effect of surveillance is strongest among the well-informed, rather than as evidence that these citizens are most likely to self-censor overall. This more cautious interpretation preserves the main insight of Hypothesis 3 – that information exposure conditions the behavioral response to surveillance – while acknowledging that foreign media consumers are not inherently more self-censoring.

Figure 2 shows the marginal effects for the interaction of the surveillance treatment dummy and the consumption of international media. The estimation follows the same procedure as in column 3 of table 1 but with recommended diagnostics. Those who not consume international media self-censor with a likelihood of 0.9%, compared to those who do with 6.1%. The possibility that the surveillance treatment effect might be driven by those who consume international media receives further support. In light

of the theory this can be interpreted as an elite (access to international media) that censors because of the concrete threat of repression (surveillance treatment). This finding supports hypothesis 3.

Fig. 2 Interaction effect international media and the surveillance treatment dummies

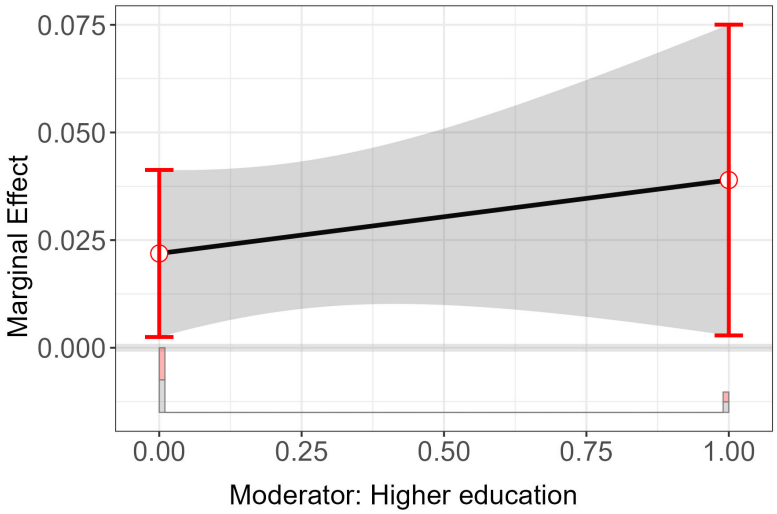


Notes: Comparisons of control and surveillance treatment groups; linear regression yielding point estimates with 95% level confidence intervals; interaction between the surveillance treatment and consuming international media. The estimation was carried out following Hainmueller et al. (2019). The estimation is independent of but corresponds to column 3 of table 1. Standard errors are robust. Weighting was applied. The list of control variables includes: age group, gender, higher education, financial situation, city size, VPN usage, ethnicity, number of children, Russian language proficiency, government employment status, region, religious affiliation, and consumption of Kazakh media.

Similarly, albeit with a different definition of elites – having access to higher education – hypothesis 4 suggests that informed citizens self-censor more. Following this logic, column 2 of table 1 and figure 3 report results corresponding to this hypothesis. First, the higher education dummy is also negative (-5.7%) and statistically significant. This suggests that citizens with higher education are more *less* likely to self-censor, on average. Although against the theoretical prediction, an alternative interpretation is that citizens are simply better informed and hence better able (and willing) to answer these sensitive questions. Turning to column 4 of table 1, both interaction effects are small and not significant. More than that, the size of the surveillance dummy decreases only slightly, suggesting no (or very small) heterogeneity for having access to higher education. Figure 3 confirms this suspicion, showing that that non-highly educated citizens self-censor with a likelihood of 2.2% and highly educated citizens with a likelihood of 3.89%, which is not a statistically significant difference between the groups. Hypothesis 4 does not find enough evidence to be supported. This study might be not powered enough to detect a small effect like this (Kane 2024). Moreover, it appears

that there are heterogeneous effects regarding the association between being elite (in either definition) and self-censorship. On the one hand, access to a richer media environment and education increases the answering of political questions. On the other hand, receiving a surveillance signal – here experimentally induced – reverses this effect and consistently leads to self-censorship. This confirms the theoretical intuition that an informed elite knows when to self-censor Gehlbach et al. (2022).

Fig. 3 Interaction effect higher education and the surveillance treatment dummies



Notes: Comparisons of control and surveillance treatment groups; linear regression yielding point estimates with 95% level confidence intervals; interaction between the surveillance treatment and having a higher education degree. The estimation was carried out following Hainmueller et al. (2019). The estimation is independent of but corresponds to column 4 of table 1. Standard errors are robust. Weighting was applied. The list of control variables includes: age group, gender, financial situation, city size, VPN usage, ethnicity, number of children, Russian language proficiency, government employment status, region, religious affiliation, and consumption of Kazakh and international media.

In summary, the main results correspond to: (1) the surveillance treatment – compared to the control group – lead to an increase in the likelihood to answer *prefer not to answer*. In the context of this study, this is evidence for political self-censorship. (2) This effect is not symmetrical: the privacy condition did not lead to a decrease in political self-censorship. Finally, (3) the effect is driven a by an informed elite that chooses to self-censor. This elite is best defined by international media consumption, and to a much more limited extent by having access to higher education.

While the measured effect of self-censorship (3.2%) is seemingly modest compared to other studies (Robinson and Tannenberg 2019), these numbers did not decrease when participants were treated with the privacy condition. In other words, while people tend to increase their self-censorship in the face of salient surveillance practices, no decrease in existing self-censorship in the face of encryption technology could be

found. This study thus provides evidence for the theoretical prediction of Büchi et al. (2022), which suggests an erosion of digital communication behavior over time, with an increasing aggregate chilling effect that does not diminish. More specifically, it was shown that the potential for *immediate* recovery is very low, if not non-existent, and that the only recovery possible is one in which the salience of surveillance practices declines over time.

An alternative interpretation is that, since the loss of privacy reduces communication behavior much more than the gain of privacy increases it, seemingly, citizens are loss averse concerning privacy. In other words, losses of privacy affect citizens more than gains in privacy do, as expressed in their behavioral adaptations. Assuming symmetry in the strength of the experimental treatments, this asymmetry in measured effects suggests asymmetric preferences, corresponding to what is known as loss-aversion (Schmidt and Zank 2005). This also means that – for policies that aim to enhance the political discourse – privacy-preserving technologies are no solution for increasing surveillance capabilities, first because they are costly and access is unequally distributed, and second because they are simply not as effective – because of the aforementioned loss-aversion. In addition, the privacy treatment may have had less valence than the surveillance treatment, i.e., it may have conveyed less urgency because it was semantically closer to the control group.

Table 2 Hypotheses overview

Hypothesis	**Support**
1) Surveillance ↑	Yes
2) Privacy ↓	No
3) International media ↑	(Yes)
4) Higher education ↑	No

Notes: Summary of the hypotheses and corresponding evidence in this study. The arrows indicate the expected direction of self-censorship associated with the variable. *Surveillance* and *Privacy* correspond to treatment conditions and are causal hypotheses. *International media* and *higher education* correspond to correlates and interactions with the treatments. Hypothesis 3 finds partial support. The surveillance treatment has the largest marginal effect among international media consumers, but these respondents are overall less likely to self-censor in the absence of surveillance cues.

5 Concluding discussion

This study contributes to the literature on authoritarianism by showing how surveillance reduces digital communication behavior in autocracies. It is one of the very few studies that directly measures and assesses self-censorship in a randomized and controlled experiment. By experimentally demonstrating that citizens self-censor when reminded of digital surveillance, this study effectively recreates a key mechanism of contemporary authoritarianism in a controlled environment, which strengthens the generalizability of its findings to real-world contexts. This has many theoretical and practical implications: Self-censoring citizens do not express their opinions on political issues, which contributes to the chilling of political discussions and the further

depoliticization of individuals, or in other words, to the stabilization of the hegemonic power of the state over public opinion. Without knowledge of peers' preferences on political issues, political opposition to incumbents has difficulty organizing, a key reason why autocrats resort to censorship (King et al. 2017). New surveillance technologies can thus directly bolster the autocrat's power before unrest forms, which in turn can be suppressed through the use of facial recognition surveillance technology (Beraja et al. 2023a).

Previous research has focused on the acceptance of new (surveillance) technologies (Kostka 2019; Kostka and Antoine 2020; Kostka et al. 2021; Kostka and Habich-Sobiegalla 2022; Xu et al. 2022; Kalmus et al. 2024; Kostka et al. 2023; Karpa and Rochlitz 2024), measurements of *opinions* towards surveillance (Davis and Silver 2004; Dietrich and Crabtree 2019; Alsan et al. 2023), or behavioral *intentions* in order to cope with surveillance (Stoycheff 2016; Stoycheff et al. 2019; Stoycheff 2022; Büchi et al. 2022; Xu 2022). The correlation between *approval* or *intentions* towards a specific technology and *behavioral adaptations* because of this exact technology might not be linear nor homogeneous. More specifically, approval or tolerance for state surveillance does not singularly translate into no self-censorship, or conversely, high self-censorship. In China, there are exceptionally high approval rates of state surveillance (Su et al. 2022), while there are also high rates of self-censorship Robinson and Tannenberg (2019). In Kazakhstan, the approval of state surveillance is much lower,[15] and self-censorship rates are also smaller, yet substantial. It appears as if approving or tolerating state surveillance might be a coping mechanism to deal with the cognitive and emotional stress of surveillance, an argument also suggested in the context of China (Ollier-Malaterre 2023). As Ollier-Malaterre (2023) documents, living with digital surveillance intertwines cultural, psycho-social, and economic factors, resulting in multifaceted behavior not free of contradictions.

In this study, the average treatment effect was primarily driven by an informed elite who strategically withhold their opinions when confronted with potential repression – a rational adaptation to an authoritarian context such as Kazakhstan. While this mechanism is rooted in theories of authoritarian information control (Roberts 2018; Gehlbach et al. 2022; Egorov and Sonin 2024), self-censorship in response to surveillance is not confined to autocracies. In most societies, expressing politically sensitive views carries some perceived risk, and awareness of being monitored – whether by state authorities or social peers – can prompt caution or silence. The chilling effect of surveillance, as others have argued, reflects conformity to perceived social or political norms (Penney 2022, p. 1520).

This study contributes to this broader understanding by examining how digital surveillance cues affect self-censorship within an autocratic setting. The heterogeneous effect by foreign media consumption should, however, be interpreted with caution: citizens exposed to foreign information are generally more willing to express opinions, yet they respond more strongly to surveillance reminders. Greater awareness and sophistication appear to heighten sensitivity to surveillance, whereas less informed citizens

[15] 31.6% of Kazakh people say the government should definitely or probably have the right to monitor all emails and any other information exchanged on the Internet, whereas this number is 60.6% in China. Source: World value survey wave 7.

may already self-censor at high baseline levels, leaving limited room for additional effects.

Finally, there are cognitive components behind behavioral adaptations that remain opaque to the design of this study. The present study identified average behavioral responses and further investigated which groups are more sensitive to self-censoring as a behavioral response, but by design neglected an investigation of cognitive mechanisms. There are different promising offers in the literature providing avenues for further research; the economics of privacy literature suggests the involvement of an evolutionary "sense" of privacy related to congenital processes of impression management (Acquisti et al. 2022), or, the literature on chilling effects of digital surveillance, which suggests including digital surveillance "imaginaries", i.e., the cognitive understanding of humans subject to surveillance processes, which substantially shape behavioral responses (Kappeler et al. 2023). If anything, this study has helped to shed light on the need for qualitative studies or mixed-methods designs that complement and enhance the findings of quantitative studies of digital surveillance.

Acknowledgements

I am grateful to Michael Rochlitz, Andrey Tkachenko, Meruyert Tatkeyeva, Quintin Beazer, Jessica Birkholz, Philip Kerner, Dinara Pisareva, Denis de Crombrugghe, and Bermond Scoggings, among others, for their valuable comments and suggestions on earlier versions of this manuscript. I also thank participants of the IERP Seminar at the University of Bremen, the First and Second Conferences on "New Advances in the Political Economy of Development in Eurasia," and the meeting of the "Research Network Innovation Eco, AI & Data Science" for their comments, as well as colleagues at the University of Bremen and Nazarbayev University for their helpful feedback. Any remaining errors are my own.

Conflict of Interest

The author declares that there are no conflicts of interest relevant to the content of this article.

Data Availability

Replication materials, including anonymized survey data, analysis code, and all scripts necessary to reproduce the results, will be uploaded upon publication.

References

Acquisti A, Brandimarte L, Hancock J (2022) How privacy's past may shape its future. Science 375(6578):270–272. https://doi.org/10.1126/science.abj0826

Adena M, Enikolopov R, Petrova M, et al (2015) Radio and the rise of the nazis in prewar germany. Quarterly Journal of Economics 130:1885–1939. https://doi.org/10.1093/qje/qjv030

Alsan M, Braghieri L, Eichmeyer S, et al (2023) Civil liberties in times of crisis. American Economic Journal: Applied Economics 15(4):389–421. https://doi.org/10.1257/app.20210736

Barceló J, Kubinec R, Cheng C, et al (2022) Windows of repression: Using covid-19 policies against political dissidents? Journal of Peace Research 59(1):73–89. https://doi.org/10.1177/0022343321106238

Beraja M, Yang DY, Yuchtman N (2022) Data-intensive innovation and the state: Evidence from ai firms in china. The Review of Economic Studies 90(4):1701–1723. https://doi.org/10.1093/restud/rdac056

Beraja M, Kao A, Yang DY, et al (2023a) AI-tocracy*. The Quarterly Journal of Economics 138(3):1349–1402. https://doi.org/10.1093/qje/qjad012

Beraja M, Kao A, Yang DY, et al (2023b) Exporting the Surveillance State via Trade in AI. Working Paper 31676, National Bureau of Economic Research, https://doi.org/10.3386/w31676

Büchi M, Festic N, Latzer M (2022) The Chilling Effects of Digital Dataveillance: A Theoretical Model and an Empirical Research Agenda. Big Data & Society 9(1):205395172110653. https://doi.org/10.1177/20539517211065368

Carothers T, Press B (2022) Understanding and responding to global democratic backsliding. Tech. rep., CEIP: Carnegie Endowment for International Peace

Corstange D (2012) Vote Trafficking in Lebanon. International Journal of Middle East Studies 44(3):483–505. https://doi.org/10.1017/S0020743812000438

Davis D, Silver B (2004) Civil Liberties vs. Security: Public Opinion in the Context of the Terrorist Attacks on America. American Journal of Political Science 48(1):28–46

Dietrich N, Crabtree C (2019) Domestic Demand for Human Rights: Free Speech and the Freedom-Security Trade-Off. International Studies Quarterly 63(2):346–353. https://doi.org/10.1093/isq/sqz011

Earl J, Maher TV, Pan J (2022) The digital repression of social movements, protest, and activism: A synthetic review. Science Advances 8(10):eabl8198. https://doi.org/10.1126/sciadv.abl8198

Edmond C (2013) Information Manipulation, Coordination, and Regime Change. The Review of Economic Studies 80:1422–1458. https://doi.org/10.1093/restud/rdt020

Egorov G, Sonin K (2024) The Political Economics of Non-democracy. Journal of Economic Literature 62(2):594–636. https://doi.org/10.1257/jel.20221494

Enikolopov R, Petrova M, Zhuravskaya E (2011) Media and political persuasion: Evidence from russia. The American Economic Review 101(7):3253–3285. https://doi.org/10.1257/aer.101.7.3253

Feldstein S (2019a) The global expansion of AI surveillance. Carnegie Endowment for International Peace Washington, DC

Feldstein S (2019b) The Road to Digital Unfreedom: How Artificial Intelligence Is Reshaping Repression. Journal of Democracy 30(1):40–52

Feldstein S (2021) The Rise of Digital Repression: How Technology Is Reshaping Power, Politics, and Resistance. Oxford University Press

Festic N (2022) Same, Same, But Different! Qualitative Evidence on How Algorithmic Selection Applications Govern Different Life Domains. Regulation & Governance 16(1):85–101

FreedomHouse (2023a) Kazakhstan: Freedom in the world 2023 country report. URL https://freedomhouse.org/country/kazakhstan/freedom-world/2023

FreedomHouse (2023b) Kazakhstan: Freedom on the net 2023 country report. URL https://freedomhouse.org/country/kazakhstan/freedom-net/2023

FreedomHouse (2024) Kazakhstan: Freedom in the world 2023 country report. URL https://freedomhouse.org/country/kazakhstan/freedom-world/2024

Frye T, Gehlbach S, Marquardt KL, et al (2017) Is Putin's Popularity Real? Post-Soviet Affairs 33(1):1–15. https://doi.org/10.1080/1060586X.2016.1144334

Frye T, Gehlbach S, Marquardt KL, et al (2023) Is Putin's Popularity (Still) Real? A Cautionary Note on Using List Experiments to Measure Popularity in Authoritarian Regimes. Post-Soviet Affairs pp 1–10. https://doi.org/10.1080/1060586X.2023.2187195

Gehlbach S, Luo Z, Shirikov A, et al (2022) A model of censorship, propaganda, and repression. Tech. rep., Society for Institutional & Organizational Economics Working Paper

Gitmez AA, Sonin K (2023) The dictator's dilemma: A theory of propaganda and repression. SSRN Working Paper 67:1–31. https://doi.org/10.2139/ssrn.4451613

Gohdes A (2023) Repression in the Digital Age: Surveillance, Censorship, and the Dynamics of State Violence. Disruptive Technology and International Security Series, Oxford University Press

Grasse D, Pavlik M, Matfess H, et al (2021) Opportunistic repression: Civilian targeting by the state in response to covid-19. International Security 46(2):130–165. https://doi.org/10.1162/isec_a_00419

Guriev S, Treisman D (2019) Informational Autocrats. Journal of Economic Perspectives 33(4):100–127. https://doi.org/10.1257/jep.33.4.100

Guriev S, Treisman D (2020) A Theory of Informational Autocracy. Journal of Public Economics 186(104158):1–11. https://doi.org/10.1016/j.jpubeco.2020.104158

Guriev S, Treisman D (2022) Spin Dictators: The Changing Face of Tyranny in the 21st Century. Princeton University Press

Hainmueller J, Mummolo J, Xu Y (2019) How much should we trust estimates from multiplicative interaction models? simple tools to improve empirical practice. Political Analysis 27(2):163–192. https://doi.org/10.1017/pan.2018.46

Hobbs WR, Roberts ME (2018) How sudden censorship can increase access to information. American Political Science Review 112(3):621–636. https://doi.org/doi:10.1017/S0003055418000084

Huang G, Hu A, Chen W (2022) Privacy at Risk? Understanding the Perceived Privacy Protection of Health Code Apps in China. Big Data & Society 9(2):1–13. https://doi.org/10.1177/20539517221135132

Jones M (2022) Digital Authoritarianism in the Middle East: Deception, Disinformation and Social Media. Hurst

Kalmus V, Bolin G, Figueiras R (2024) Who is afraid of dataveillance? attitudes toward online surveillance in a cross-cultural and generational perspective. New Media & Society 26(9):5291–5313. https://doi.org/10.1177/14614448221134493

Kane JV (2024) More than meets the itt: A guide for anticipating and investigating nonsignificant results in survey experiments. Journal of Experimental Political Science p 1–16. https://doi.org/10.1017/XPS.2024.1

Kang W, Hughes MA, Roy D (2024) Anonymization of voices in spaces for civic dialogue: Measuring impact on empathy, trust, and feeling heard. Proceedings of the ACM on Human-Computer Interaction 8(CSCW2):Article 482. https://doi.org/10.1145/3687021

Kappeler K, Festic N, Latzer M (2023) Dataveillance imaginaries and their role in chilling effects online. International Journal of Human-Computer Studies 179:103120.

https://doi.org/https://doi.org/10.1016/j.ijhcs.2023.103120

Karpa D, Rochlitz M (2024) Authoritarian surveillance and public support for digital governance solutions. Comparative Political Studies 58(10): 2237-2265. https://doi.org/10.1177/00104140241290208

Kaye D (2015) Report on encryption, anonymity, and the human rights framework. United Nations Human Rights Council 29(32). https://www.ohchr.org/en/documents/thematic-reports/ahrc2932-report-encryption-anonymity-and-human-rights-framework

King G, Pan J, Roberts ME (2017) How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, not Engaged Argument. American Political Science Review 111(3):484–501. https://doi.org/doi:10.1017/S0003055417000144

Knutsen CH, Marquardt KL, Seim B, et al (2024) Conceptual and measurement issues in assessing democratic backsliding. PS: Political Science & Politics 57(2):162–177. https://doi.org/10.1017/S104909652300077X

Kostka G (2019) China's Social Credit Systems And Public Opinion: Explaining High Levels of Approval. New Media & Society 21(7):1565–1593. https://doi.org/10.1177/1461444819826

Kostka G, Antoine L (2020) Fostering Model Citizenship: Behavioral Responses to China's Emerging Social Credit Systems. Policy & Internet 12(3):256–289. https://doi.org/https://doi.org/10.1002/poi3.213

Kostka G, Habich-Sobiegalla S (2022) In Times of Crisis: Public Perceptions Toward COVID-19 Contact Tracing Apps in China, Germany, and the United States. New Media & Society 0(0). https://doi.org/10.1177/14614448221083285

Kostka G, Steinacker L, Meckel M (2021) Between Security and Convenience: Facial Recognition Technology in the eyes of citizens in China, Germany, the United Kingdom, and the United States. Public Understanding of Science 30(6):671–690. https://doi.org/10.1177/096366252110015

Kostka G, Steinacker L, Meckel M (2023) Under Big Brother's Watchful Eye: Cross-country Attitudes Toward Facial Recognition Technology. Government Information Quarterly 40(1):101761. https://doi.org/10.1016/j.giq.2022.101761

Kudaibergenova DT, Laruelle M (2022) Making sense of the january 2022 protests in kazakhstan: failing legitimacy, culture of protests, and elite readjustments. Post-Soviet Affairs 38(6):441–459. https://doi.org/10.1080/1060586X.2022.2077060

Lamberova N, Sonin K (2023) Information manipulation and repression: A theory and evidence from the covid response in russia. University of Chicago, Becker Friedman

Institute for Economics Working Paper 101(2022-101):1–70

Lee HJ (2023) I've left enough data: Relations between people and data and the production of surveillance. Big Data & Society 10(1):20539517231173904. https://doi.org/10.1177/20539517231173904

Little AT, Meng A (2024) Measuring democratic backsliding. PS: Political Science & Politics 57(2):149–161. https://doi.org/10.1017/S104909652300063X

Liu KZ (2019) Commercial-state Empire: A Political Economy Perspective on Social Surveillance in Contemporary China. The Political Economy of Communication 7(1)

Lührmann A, Lachapelle J, Grahn S, et al (2020) Pandemic backsliding: Democracy and disinformation seven months into the covid-19 pandemic. V-dem Policy Brief 25:1–6

Manokha I (2018) Surveillance, Panopticism, and Self-Discipline in the Digital Age. Surveillance & Society 16:219–237. https://doi.org/10.24908/ss.v16i2.8346

Ollier-Malaterre A (2023) Living with Digital Surveillance in China: Citizens' Narratives on Technology, Privacy, and Governance. Routledge studies in surveillance, Routledge

Oz M, Yanik A (2022) Fear of surveillance: Examining turkish social media users' perception of surveillance and willingness to express opinions on social media. Mediterranean Politics 0(0):1–25. https://doi.org/10.1080/13629395.2022.2046911

Pan J, Siegel AA (2020) How saudi crackdowns fail to silence online dissent. American Political Science Review 114(1):109–125. https://doi.org/10.1017/S0003055419000650

Penney JW (2016) Chilling Effects: Online Surveillance and Wikipedia Use. Berkeley Technology Law Journal 31(1):117–182. https://doi.org/10.15779/Z38SS13

Penney JW (2017) Internet Surveillance, Regulation, and Chilling Effects Online: a Comparative Case Study. Internet Policy Review 2(6):1–39. https://doi.org/10.14763/2017.2.692

Penney JW (2022) Understanding chilling effects. Minnesota Law Review 106:1451–1530

Prajith N, Bhuyan B (2025) The role of online anonymity in fostering self-expression: An analysis of individual and contextual dynamics. International Journal of Indian Psychology 13(1):155–162. https://doi.org/10.25215/1301.014

Qin B, Strömberg D, Wu Y (2017) Why Does China Allow Freer Social Media? Protests versus Surveillance and Propaganda. Journal of Economic Perspectives

31(1):117–40. https://doi.org/10.1257/jep.31.1.117

Raman RS, Evdokimov L, Wurstrow E, et al (2020) Investigating large scale https interception in kazakhstan. In: Proceedings of the ACM Internet Measurement Conference. Association for Computing Machinery, New York, IMC '20, p 125–132

Roberts ME (2018) Censored: Distraction and Diversion Inside China's Great Firewall. Princeton: Princeton University Press

Roberts ME (2020) Resilience to Online Censorship. Annual Review of Political Science 23(1):401–419. https://doi.org/10.1146/annurev-polisci-050718-032837

Robinson D, Tannenberg M (2019) Self-censorship of Regime Support in Authoritarian States: Evidence From List Experiments in China. Research & Politics 6(3):1–9. https://doi.org/10.1177/2053168019856449

Schauer F (1978) Fear, Risk, and the First Amendment: Unraveling the "Chilling Effect.". Boston University Law Review 58(0):685—732

Schmidt U, Zank H (2005) What is loss aversion? Journal of Risk and Uncertainty 30(2):157–167

Silvan K (2024) Three Levels of Authoritarian Legitimacy: Successor Designation and Peaceful to Non-Peaceful Leadership Transition in Kazakhstan. Communist and Post-Communist Studies pp 1–23. https://doi.org/10.1525/cpcs.2024.1996171

Stoycheff E (2016) Under Surveillance. Journalism & Mass Communication Quarterly 93(2):296–311. https://doi.org/10.1177/1077699016630255

Stoycheff E (2022) Cookies and Content Moderation: Affective Chilling Effects of Internet Surveillance and Censorship. Journal of Information Technology & Politics pp 1–12. https://doi.org/10.1080/19331681.2022.2063215

Stoycheff E, Liu J, Xu K, et al (2019) Privacy and the Panopticon: Online Mass Surveillance's Deterrence and Chilling Effects. New Media & Society 21(3):602–619. https://doi.org/10.1177/1461444818801317

Su Z, Xu X, Cao X (2022) What Explains Popular Support for Government Monitoring in China? Journal of Information Technology & Politics 19(4):377–392. https://doi.org/10.1080/19331681.2021.1997868

Tannenberg M (2022) The Autocratic Bias: Self-censorship of Regime Support. Democratization 29(4):591–610. https://doi.org/10.1080/13510347.2021.1981867

Xu X (2021) To Repress or to Co-opt? Authoritarian Control in the Age of Digital Surveillance. American Journal of Political Science 65(2):309–325. https://doi.org/10.1111/ajps.12514

Xu X (2022) The unintrusive nature of digital surveillance and its social consequences. Tech. rep., Working paper

Xu X, Kostka G, Cao X (2022) Information Control and Public Support for Social Credit Systems in China. Journal of Politics 84(2):2230–2245. https://doi.org/10.1086/718358

Zhuravskaya E, Petrova M, Enikolopov R (2020) Political effects of the internet and social media. Annual Review of Economics 12(1):415–438. https://doi.org/10.1146/annurev-economics-081919-050239

Appendix A Supplementary Material

A.1 Treatment design

control	privacy	surveillance
In the next section, you will be asked your opinion on economic and political issues directly. Your answers will remain confidential.	In the next section, you will be asked your opinion on economic and political issues directly. Your answers will remain confidential. Our encryption mechanisms make it completely impossible to track your data.	In the next section, you will be asked your opinion on economic and political issues directly. Your answers will remain confidential. However, as you may be aware, the government of Kazakhstan may access information about your online activity directly from your Internet Service Provider.

A.2 Additional Tables

Table A1 Summary statistics

Variable	N	Mean	SD
agegroup			
... 18-24	5025	0.13	0.337
... 25-34	5025	0.216	0.411
... 35-44	5025	0.211	0.408
... 45-54	5025	0.169	0.375
... 55+	5025	0.275	0.446
male	5025	0.487	0.5
higher_ed	5025	0.232	0.422
financial_situation_scale	5025	2.85	1.14
large_city	5025	0.22	0.414
vpn_user	5025	0.109	0.312
ethnicity			
... Kazakh	5025	0.722	0.448
... Other	5025	0.033	0.179
... Russian	5025	0.245	0.43
children	5025	2.26	1.46
language_russian	5025	0.622	0.485
government_employee	5025	0.109	0.312
media_international	5025	0.326	0.469

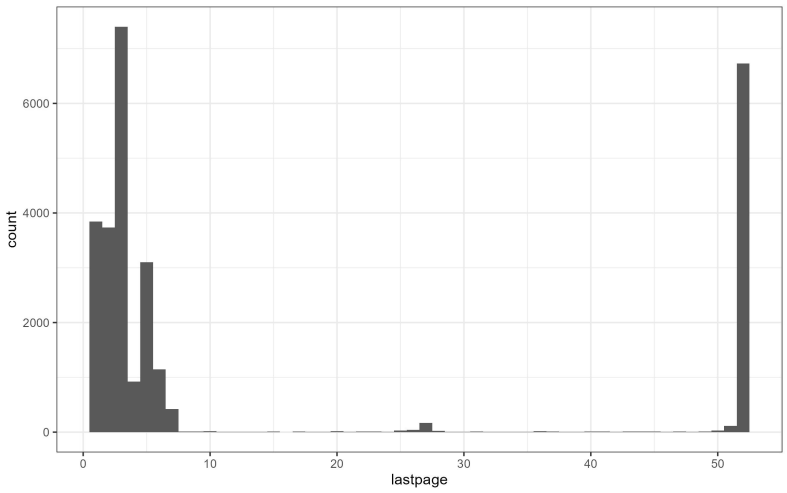
Table A2 Linear regression models for all dependent variables separately

	Dependent variable:			
	Protest	Sanction evasion	Invasion	Placebo
	(1)	(2)	(3)	(4)
Privacy	0.011	−0.010	0.007	−0.021
	(0.016)	(0.016)	(0.016)	(0.014)
Surveillance	0.042**	0.025	0.031+	−0.004
	(0.016)	(0.016)	(0.016)	(0.014)
Constant	0.285**	0.317**	0.318**	0.217**
	(0.011)	(0.011)	(0.011)	(0.010)
Observations	5,025	5,025	5,025	5,025
R^2	0.002	0.001	0.001	0.0005
Adjusted R^2	0.001	0.001	0.0004	0.0001
Residual Std. Error (df = 5022)	0.459	0.467	0.470	0.406
F Statistic (df = 2; 5022)	3.843*	2.446+	2.019	1.199

Notes: Linear regression models for all sensitive questions and the placebo question. Treatment dummies served as independent variables. Weighting for age and gender applied. The dependent variable is answering *prefer not to answer* in the four different questions described in section 3. +p<0.1; *p<0.05; **p<0.01

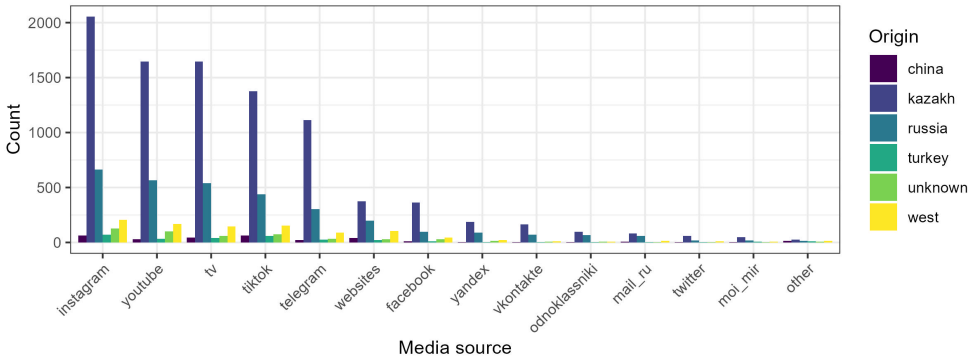
A.3 Additional Figures

Fig. A1 Last page of the survey



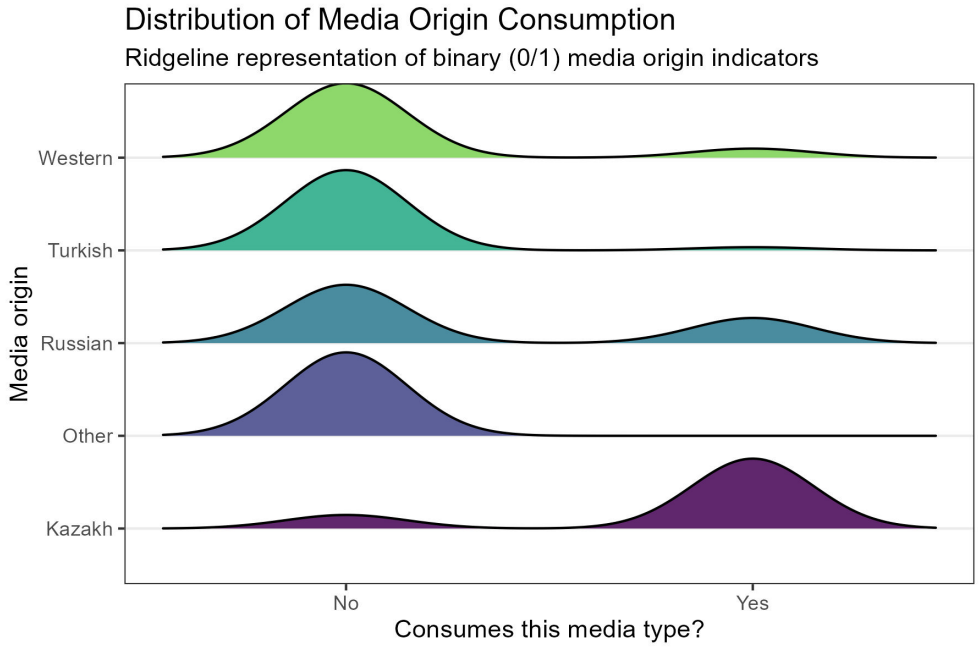
Notes: This figure shows which page of the survey was the last page a respondent saw. The very last page was the payment page, which was reached by 6,727 participants. 5,025 of these were successfully compensated, the others were not due to either failing quality checks, or entering a phone number that has already been used for payment. This indicated a duplicate entry by a single individual which is against the terms of the survey. The small bump around question 27 failed the matching age question quality check. The left tail of the distribution, i.e., those who left before page ten failed quality checks repeatedly (speeding, duplicate IP, under 18 years old) and/or indicated living outside of Kazakhstan. Many attempts to enter and complete the survey multiple times were prevented by the quality checks. The compensation of 700 Tenge posed as an incentive to complete the survey multiple times.

Fig. A2 Media consumption patterns



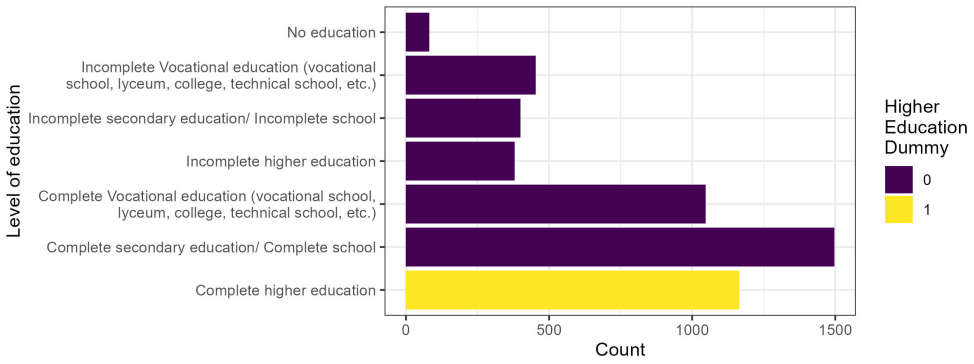
Notes: This figure shows which media sources respondents indicated as their "most important source of information for political and economic events". They were able to select up to four out of a list with all the options shown in randomized order. The next question asked them about the origin of each news source, which is indicated in the colors of the bars.

Fig. A3 Distribution of Foreign Media Consumption by Region of Origin



Notes: Each density ridge shows the proportion of respondents who reported consuming political or economic news from a given region (Kazakh, Russian, Western, Turkish, or Other). The variable underlying these distributions equals 1 if the respondent reported at least one source from that region and 0 otherwise. Ridgelines represent smoothed densities for visual clarity. About one-third of respondents reported consuming at least some foreign media (Western, Russian, Turkish, or Other).

Fig. A4 Education levels



Notes: This figure shows which is the highest level of education respondents indicated they received. Those with a completed higher education degree have been as "elite", corresponding to hypothesis 4 and regressions in table 1.

A.4 Review process and divergence from the pre-analysis plan

The pre-registration can be found on https://aspredicted.org/BVT_9Z3. The major departures from the planned analysis – particularly with respect to not including portions of the data collected – are outlined below. The author is grateful to the excellent anonymous reviewers who provided very useful suggestions for improving the paper. Much of the deviation from the pre-registered plan is due to reviewer recommendations.

A.4.1 Revised Estimation Approach

The original plan relied on a list experiment with indirect measures analyzed via maximum likelihood and logistic/multinomial regressions. This approach has been replaced by an estimation strategy that defines self-censorship solely as "prefer not to answer" responses, with linear (OLS) regression and robust standard errors—including specifications with control variables and regional fixed effects – to ensure a more intuitive and consistent interpretation. The initial pre-registration suggested to compare direct questions with list experiment ("indirect") responses. This approach has been neglected in favor of a more concise analysis. Assessing a "baseline" of opinions toward a particular question through list experiments – while empirically successful – did not improve the theoretical contribution with respect to self-censorship because of surveillance. Assessing the effect of the treatments on the responses to the direct questions proved to be more crucial in this regard.

A.4.2 Streamlined Analysis of Conditional Effects

Originally, heterogeneous effects were examined through multiple subgroup analyses using median sample splits, resulting in a large number of tests without sufficient theoretical guidance. The literature review has since been refined to better motivate the empirical design, resulting in new, focused hypotheses regarding the heterogeneous effects of surveillance on self-censorship. Furthermore, the estimation of conditional average treatment effects now uses interaction models, providing clearer insights into moderation effects.