Silenced Voices: State Surveillance and Self-Censorship in Kazakhstan

David Karpa

University of Bremen*

November 21, 2024

Abstract

How do authoritarian states maintain their hegemony over public opinion beyond state-run media and outright repression of journalists? Theory and previous research suggest that surveillance practices discourage citizens from engaging in legitimate digital communication behaviors, such as expressing opinions online. Drawing on an original survey experiment conducted in Kazakhstan in November 2023 (N=5,025), this study is able to show that citizens exposed to a text-based surveillance treatment reduce their response rate to sensitive questions by 2.5% to 4%, while this effect is not triggered for non-sensitive questions. By comparing subgroups, heterogeneous treatment effects are identified, showing that treatment effects are much larger (9%) or non-existent for certain groups. In particular, older citizens, those living in rural areas, those with a high intensity of internet use, and those consuming media from Russia are prone to self-censorship. By comparing indirect and direct questions within a list experiment, the baseline level of self-censorship is calculated by which allows to distinguish pre-existing self-censorship from treatment effects. This study contributes to the literature on digital authoritarianism by showing how state surveillance practices undermine political discourse on the Internet, which in turn contributes to authoritarian stability.

Keywords: big data, surveillance, privacy, political repression, democracy, autocracy

* davidfkarpa@gmail.com

Nazarbayev University Institutional Research Ethics Committee (NU-IREC) approved the experiment 771/25092023

CONFERENCE PAPER; DO NOT CIRCULATE

1 Digital mass surveillance

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 behaviour 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 behaviour is in direct conflict with the essential components of deliberative democratic frameworks and represents a significant challenge to the healthy functioning of democratic societies (Penney 2022; Kappeler et al. 2023).

In contrast, within an autocratic context, the autocrat seeks anticipatory and obedient behaviors. 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). However, 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 behaviour (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 democratic deliberation processes, democratic backsliding, and to authoritarian stability (Carothers and Press 2022).

The main aim of this paper is to test these theories and investigate whether surveillance practices indeed induce self-censorship among citizens, to which magnitude, and what factors, if any, moderate this effect. To this end, a survey experiment with 5,025 participants was conducted in Kazakhstan, a country where the government has repeatedly deployed mass surveillance technology at the internet service provider level (Raman et al. 2020). Partici-

[1] https://www.ohchr.org/en/press-releases/2022/09/spyware-and-surveillance-threats-privacy-and-human-rights-growing-un-report

pants in the study were asked sensitive questions on domestic and geopolitical topics, after exposure to either a control, surveillance or privacy condition. The main results of this study are that participants in the surveillance condition indeed self-censor, 4% on items concerning domestic politics and between 2.5 and 3.2% on geopolitical topics, whereas exposure to the privacy treatment had no effects. Furthermore, strong heterogeneity in the surveillance treatment effects was detected, with effect sizes increasing up to three times the size, or diminishing entirely, for some demographic groups. This study adds to the literature, by following the call of Büchi et al. (2022) in experimentally investigating self-censorship induced by digital surveillance and estimating its magnitude. In addition, this study contributes to public opinion research by estimating a usually undetected baseline of self-censorship that leads 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 (digital) authoritarianism by showing how autocrats control the informational environment with digital tools (King et al. 2017; Roberts 2018; Guriev and Treisman 2019, 2020; Feldstein 2021; Gohdes 2023; Egorov and Sonin 2024). The following Section provides an overview of the relevant literature from which the hypotheses are derived. Section 3 embeds the hypotheses in the research design and elaborates on the methodological details of the study. Section 4 presents the results, while Section 5 concludes with a discussion of the results.

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

A literature agnostic to the institutional background revives Bentham's and Foucault's metaphors of the panopticon (Manokha 2018; Stoycheff et al. 2019), in which Bentham (2011) paints a picture of a prison called *Panopticon*, where a central tower oversees cells in the form of a ring around the tower in the centre. In essence, the theory of the panopticon involves three main assumptions (Manokha 2018): First, the omnipresence of the inspector, guaranteed by his total invisibility; second, the universal visibility of the objects of surveillance; and third, the *assumption* of constant observation of those being watched. Under this regime, inmates infer that they are under constant surveillance and thus exercise self-control and self-discipline. Coercion becomes unnecessary, except for a few rare instances of disobedience. Based on this design, Foucault (2012) developed his theory of self-discipline through *assumed* surveillance. The metaphor of the panopticon helps understanding the effects of state surveillance, as it involves a centralised entity (the watchtower/state) watching over the inmates/citizens. Even if in practice the actual surveillance is much more distributed than the metaphor suggests, the perception of a centralised state may dominate because of the at least partially opaque and sometimes highly technical processes involved. This perception of surveillance, similar to the original design of Bentham's prison, encourages self-discipline. Indeed, empirical studies explicitly used the metaphor of the panopticon in the context of digital environments (Stoycheff et al. 2019), while others implicitly described the mechanism of self-discipline (mostly self-censorship) due to the *fear* of repression (Roberts 2018; Manokha 2018; Tannenberg 2022; Stoycheff 2022; Oz and Yanik 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). As a result, 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, surveillance enables targeted repression, and the mere possibility of repression, in turn, induces *fear*, which leads 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*[2] 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:

Hypothesis 1: Digital surveillance induces self-disciplining behavior in the form of 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,

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

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[3].

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 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.

Other studies suggest that behavioural adaptations to surveillance include increased use of privacy-preserving technologies to cope with surveillance (Büchi et al. 2022; Kappeler et al. 2023). Censorship in the form of blocked websites is being bypassed with circumvention tools, leading to renewed access by citizens and increased interest in blocked content (Hobbs and Roberts 2018). In the same way, effective encryption mechanisms should *recover* digital communication behaviour. Given the baseline of digital surveillance in contemporary societies, particularly in Kazakhstan, the potential for recovering digital communication behaviour is significant (see Section 4.3). Correspondingly, the second hypothesis proposes that:

Hypothesis 2: Privacy-enhancing technology reduces self-disciplining behavior in the form of self-censorship in politically sensitive topics.

[3] https://www.hrw.org/news/2024/01/31/longing-justice-kazakhstan

This study draws on this literature and investigates (1) whether digital surveillance induces self-censorship, (2) whether this effect can be reversed by a privacy-preserving technology, and (3) which role demographic factors play in moderating these effects.

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[4] and carried out by NAC Analytica, a leading Kazakh sociological and public opinion research organization.[5] 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). 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. Figure 10 in the appendix shows the difference between the two framings in terms of the outcome variable. While there are differences in the proportions, the dynamics of self-censorship develop analogously across treatments. Question 4 acted as a placebo, in order to control for design effects (*In your opinion, is working more than 50 hours per week*

[4] https://aspredicted.org/BVT_9Z3

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

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). 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 1 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*. Participants reported having received education on a scale ranging from 1 - 6 (M=4.55, SD=1.75), 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). 30.1% (SD=0.46) of participants reported consuming news sources from Russia. Participants were asked on a 1 - 4 scale about their trust in government institutions (M=2.99, SD=0.94) and whether they think most people can be trusted (M=2.57, SD=0.87). 10.9% (SD=0.31) of the participants work for some government organization and 10.9% (SD=0.31) use a virtual private network (vpn). Using a 1 - 5 scale, participants were asked about the frequency of using the internet (M=4.47, SD=1.8), and about the intensity of religious practices (M=3.77, SD=1.25). Weights for age and sex were applied to make the sample nationally representative.

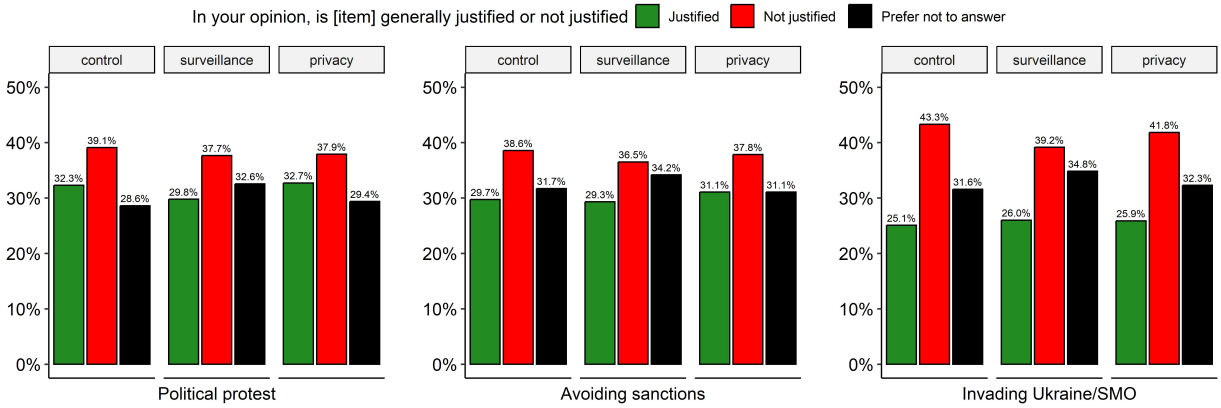
4 Results

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.

Figure 1: Responses to dependent variables

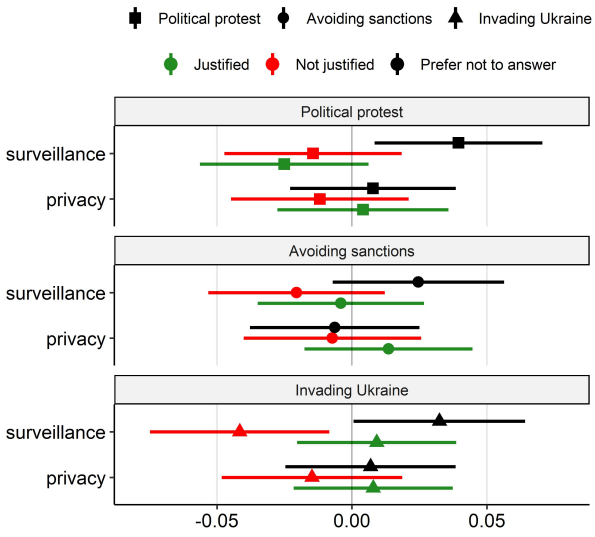


In order to assess whether these differences are meaningful statistical deviations, i.e., whether treatments have significant effects, multinomial logistic regression models were estimated with the three response options as unordered dependent variables, and treatment dummies as dependent variables. Weights for age and gender were applied to make the sample representative of the population. Figure 2 shows the average treatment effects (ATE) resulting from these models. A treatment effect, as defined here, would cause a change in proportions in the response options between treatments. The sensitive items show treatment effects being significant at the 95% level. The surveillance treatment results in an increase in the 'prefer not to answer' option for sensitive items, by 2.46 to 3.95%. The differences are significant at the 95% level in the questions involving participating in political protests and Russia's invasion of Ukraine. Additionally, the surveillance treatment decreased the rate of the 'not justified' option by 1.44 to 4.16%, on average. This difference is significant at the 95% level in the question involving Russia's invasion of Ukraine. The average differences between control and treatment groups are modest but hint at systematic self-censorship, thus lending some support for hypothesis 1. The differences in the option 'justified' in the surveillance condition were not significant at the 95% level.

The privacy condition did not yield any statistically meaningful comparison. Noteworthy, however, is that treatment effects seem to be systematic, in that in every question the number

of respondents who answered *justified* increased (between 0.41 to 1.36%) and the number of those who answered *not justified* decreased (between 0.72 to 1.48%). This study is perhaps not powered to detect effects this small (Kane 2024), hence not providing evidence in favour of the second hypothesis.

Figure 2: Average treatment effects



Note: Comparisons of control and treatment groups; multinomial logistic regression model point estimates with 95% level confidence intervals.

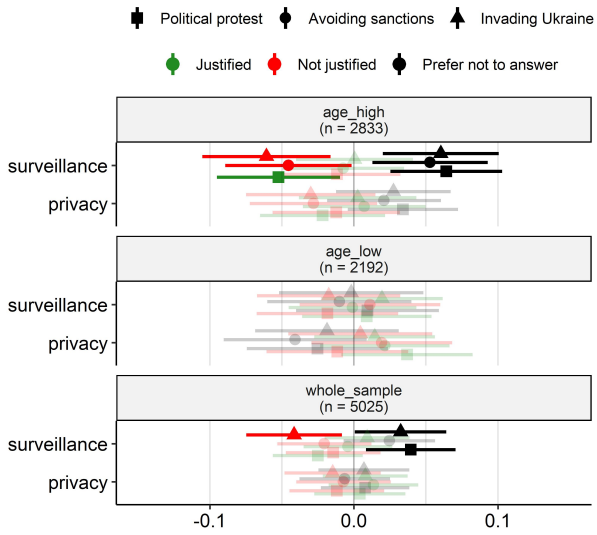
4.2 Conditional average treatment effects

Following the pre-registered analysis plan, conditional average treatment effects are calculated for the following moderators: age, education, trust in government, ethnicity, government employment, trust in government, income, Internet use intensity, place of residence, gender, religiosity, Russian media consumption, social trust, and VPN use. Some of these are discussed in detail here, while for readability the reader is referred to Figure 11 in the appendix for an overview. Scales were split at the median, whereas binary variables were analysed following their natural split at 1 and 0. Fully visible point estimates are significant at the 95% level, while insignificant comparisons are transparent.

First, considering results in Figure 11, and consistent with the results in 4.1, a comparison of the estimates between the control and privacy conditions yields only two significant comparisons (out of 98 comparisons). The confidence intervals are not larger compared to the surveillance condition, which could indicate greater uncertainty or nonlinearity in the effects. In sum, there is insufficient evidence to conclude a reduction in self-censorship or,

conversely, an increase in preference disclosure due to the treatment of respondents in the privacy condition. Therefore, hypothesis 2 is not accepted.

Figure 3: Conditional treatment effects: age



Second, across the majority of splits, the surveillance condition increased self-censorship on sensitive items relative to the control group, mostly through an increase in the proportion of *prefer not to answer* responses on sensitive items. Interestingly, this effect seems to depend on both item and individual characteristics, as different patterns emerge on both dimensions. Older participants self-censor by switching from saying *participating in a political protest* can be justified to not revealing their opinion (see Figure 3). Regarding the second and third items corresponding to the invasion of Ukraine, older subjects tend to hide their preferences by avoiding disagreement and not revealing their opinion. Younger participants do not seem to be sensitive to the surveillance treatment at all, as reflected by very small estimates and insignificant comparisons. This finding challenges previous empirical studies measuring self-censorship in autocracies finding self-censorship to be greater among younger respondents (Robinson and Tannenberg 2019) or, less dramatically, suggests different effects depending on the context.

This pattern is repeated for more educated individuals, as they when compared to those with less education, exert a higher level of self-censorship. Precisely, the proportions of respondents answering *prefer not to answer* to the sensitive questions in the whole population of 3.9%, 2.5%, and 3.2% increase to 5.6%, 3.1%, and 5.3%, respectively, for the subgroup of more educated respondents.

Figure 4: Conditional treatment effects: education

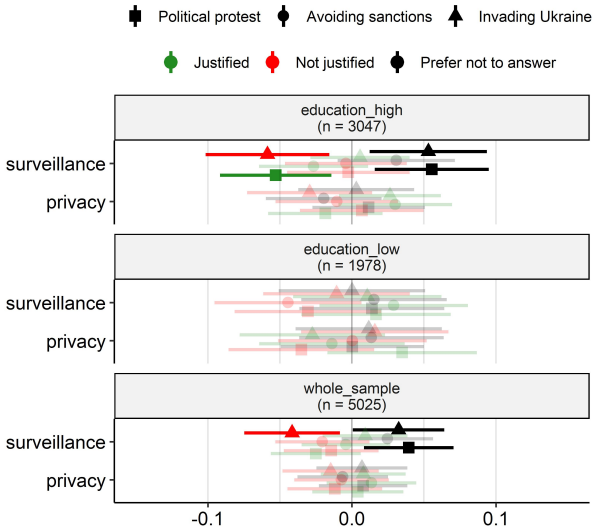


Figure 5: Conditional treatment effects: place of residence

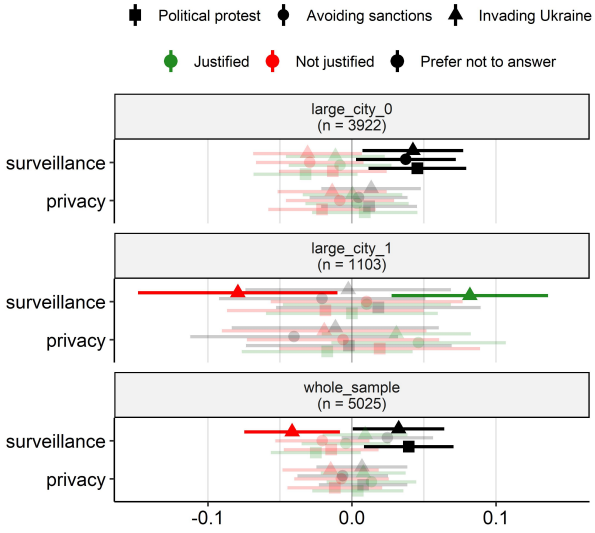
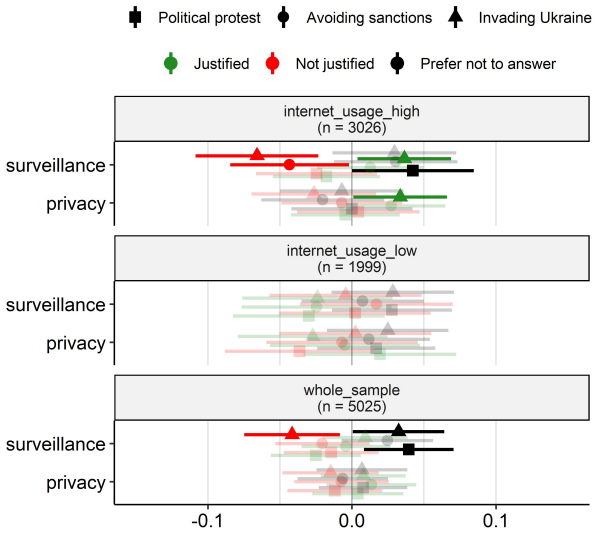


Figure 5 shows comparisons where the sample is split in respondents residing in Almaty or Astana (*large_city_1*), or not. In the rural population (not living in either metropolitan area), an increase in the *prefer not to answer* response was found, not in stating an opinion to this question. On the contrary, respondents from either of the two large cities – when comparing the surveillance to the control group – responded *justified* more often (8.2%) and *not justified* less often (-7.9%) to the question of whether invading Ukraine was justified. This behavior is unique in that it was only found for the subgroups of large city residents

and high-intensity internet users (see Figure 6)[6], and for the item corresponding to Russia's invasion of Ukraine. Similarly large effects were found for respondents who consumed media sources where they indicated Russia as the country of origin, see Figure 7. In the surveillance condition, respondents increased the proportion of *prefer not to answer* by between 8.5 and 8.9% when they also consumed media sources from Russia. Conversely, those who did not consume any media sources from Russia chose *prefer not to answer* between -0.04% and 1.9% less or more often, without statistical significance. This finding raises the question of what it is that makes respondents who choose to consume Russian media also self-censor strongly. Or, perhaps, what it is about consuming Russian media that makes people self-censor more intensively. Other studies found strong effects of domestic propaganda consumption on approval rates of surveillance technology in China (**?**) and Russia (Karpa and Rochlitz 2023).[7] This study adds to this by finding what appears to be a 'spill-over' effect of Russian propaganda on the Kazakh population[8].

Figure 6: Conditional treatment effects: internet use



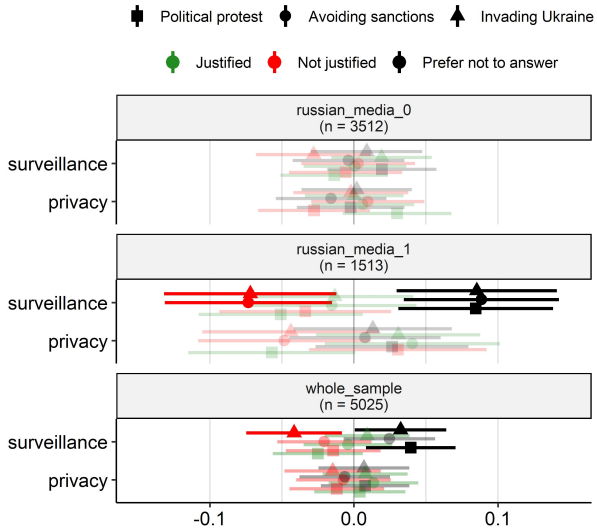
When focusing on the second item, whether to *help Russia avoid Western sanctions*, results show that older subjects tend to hide their preferences by avoiding to disagree and

[6]There is a large overlap between the two subgroups: 847 out of 1103 respondents (77%) in the large city group were also high-intensity internet users.

[7]All of these studies including this one, however, provide only correlational evidence.

[8]Out of 1513 respondents who indicated consuming Russia media, 939 are ethnically Kazakh, 494 Russian, and 80 preferred not to reply to this question.

Figure 7: Conditional treatment effects: Russian media consumption



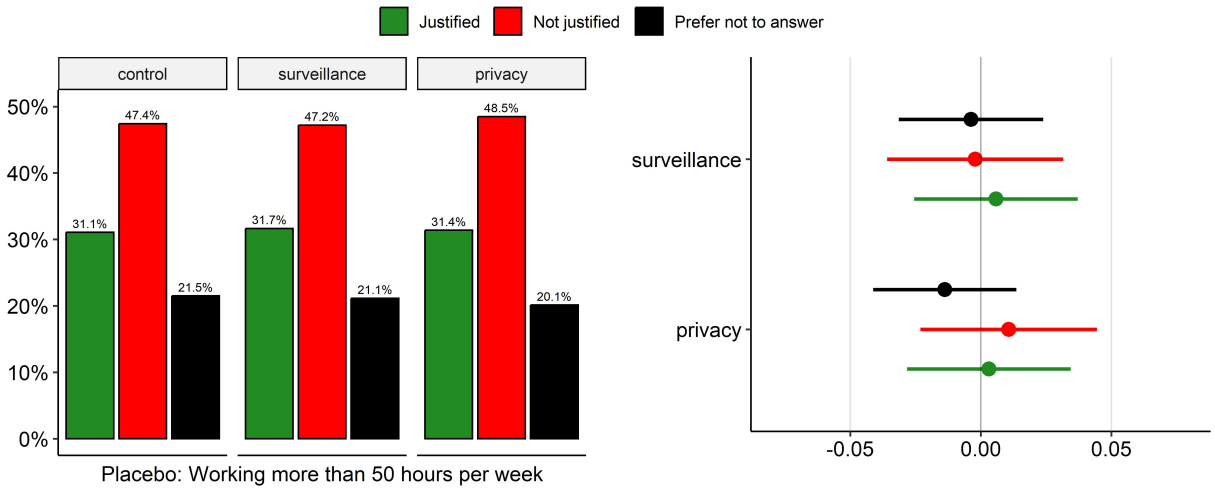
instead not revealing their opinion, which seems to also be the case for heavy internet users, those exhibiting a lower trust in the government, and highly educated subjects. The same pattern emerges for the third item, whether the *Russian invasion of Ukraine/SMO* was justified. Older, more educated, and those with low social trust hide their preferences by avoiding to disagree and instead not revealing their opinion. For those who earn better, heavy internet users and those with high social trust the same pattern of response shifts emerge, albeit appearing less often with statistically significant estimates. In general, the surveillance treatment made respondents significantly less likely to say *not justified* when asked whether Russia's invasion of Ukraine was justified. The only two subgroups that increased their approval of the invasion in the surveillance condition were respondents employed by the government (3.6%) and ethnically Russian respondents (0.2%), both of which effects are statistically insignificant. Other studies found that people are less likely to *say* they would like to trade civil liberties for security when they are disadvantaged compared to peers (Davis and Silver 2004; Dietrich and Crabtree 2019; Alsan et al. 2023). In this study, those with lower income *act* close to average, whereas those with lower education seem unresponsive to the surveillance treatment.

In summary, the main results correspond to: (1) the surveillance treatment mostly leads those to who agree or disagree with an item to instead chose *prefer not to answer*, and (2) far less often it induces preference falsification by participants (dis)agreeing with the opposite of the stated opinion in the control group. Furthermore, (3) the privacy treatment had small and thus undetectable effects. Finally, (4) effect sizes dramatically increase for

some subgroups (up to three times the size) or diminish entirely, clearly indicating strong heterogeneity in treatment effects.

4.3 Robustness tests and baseline self-censorship

Figure 8: Placebo Item

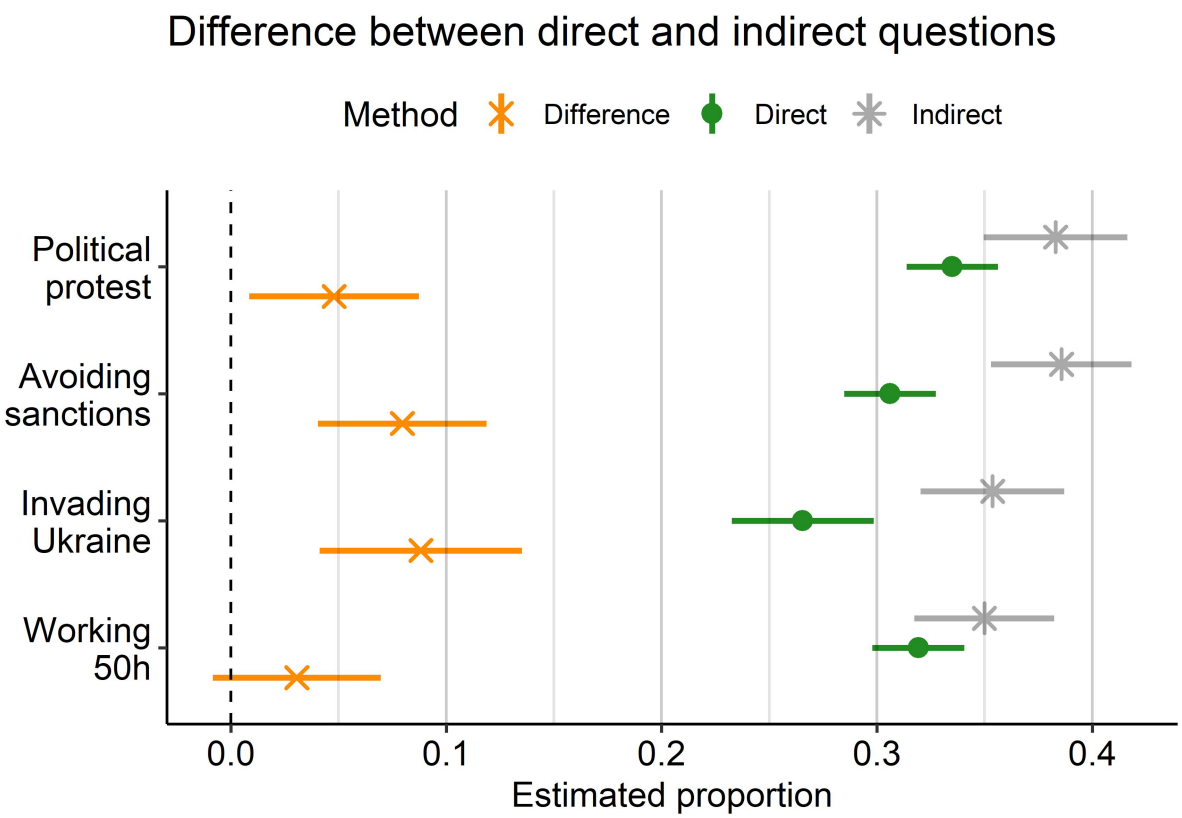


Note: Left: Responses to dependent variable by treatment condition. Right: Comparisons of control and treatment groups; multinomial model point estimates with 95% level confidence intervals.

In order to assess the quality of effects described above different robustness checks were performed. First, Figure 8 shows the results for a placebo item (*In your opinion, is working more than 50 hours per week generally justified?*) that was run beside the three sensitive items. First, the rate of non-responses, i.e. subjects who chose "prefer not to answer", is about 30% for the sensitive items (figure 1), while it is about 20% for the placebo item. On average, preference falsification is higher in the politically sensitive items, which lends support for the assumption that participants perceived the questions as sensitive and adapted accordingly. Second, the estimates for treatment effects in both conditions for the placebo question are very small, do not show systematic variation, and are insignificant at the 95% level. This serves as an indication that the study does not suffer from this specific design effect, which would correspond to self-censorship regardless of the sensitivity of the item.

How does the context in which this study was conducted affect its results? When asking about politically sensitive issues – e.g., whether political protests are justified – the context itself is a powerful treatment. Other studies have shown that there is substantial self-censorship on such questions in autocracies, with up to a quarter of respondents falsifying their preferences when asked directly (Robinson and Tannenberg 2019). Arguably,

Figure 9: Baseline self-censorship



Note: The shape of the point estimate indicates whether the response was given to an indirect question (list experiment) or whether it was given to a direct question with the options *Justified / Not justified / Prefer not to answer* (logistic regression on "Justified"). Confidence intervals are obtained using Monte Carlo simulations and are given at the 95% level. Fitted values were obtained using control variables including region fixed-effects, following Blair and Imai (2012). For more methodological details on the list experiment, see Section A.4

the treatment effects in this study's experiment are conservative because participants are *already* treated, simply by using the Kazakh Internet. Thus, what was measured in 4.1 and 4.2 are the *additional* treatment effects, and these should be smaller, assuming a diminishing marginal effect[9].

To account for this, direct and indirect questioning techniques are combined and a baseline of self-censorship is calculated and shown in Figure 9. The differences between the direct and indirect questions are, from top to bottom, 4.8%, 8.0%, 8.8% and 3.0%, with the difference being significant at the 95% level for the sensitive items, but not for the placebo

[9]While the assumption of diminishing marginal effects seems reasonable, it cannot be ruled out that participants become unresponsive to an additional treatment or that there are increasing marginal effects, nor would it make sense to assume homogeneous effects across different demographic groups.

item, again supporting the assumption of no design effects. In the indirect question, non-response was both not possible and not strategically necessary, as the indirect question does not compromise the respondent's privacy. Similar to the treatment effects in Figure 8, there is no also no *baseline* self-censorship in the placebo condition. The differences between the direct and indirect questions thus include the amount of respondents who hid their preferences behind *prefer not to answer* and, arguably about 3% of participants who chose the option because it was less cognitively demanding. While seemingly modest compared to other studies (Robinson and Tannenberg 2019), these numbers increased when respondents were treated with the surveillance condition and, notably, 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 and imperfect recovery. 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.

5 Concluding discussion

This study contributes to the literature on digital authoritarianism by showing how surveillance reduces digital communication behavior. 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. 2023b).

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; **?**; Kalmus et al. 2022; Karpa and Rochlitz 2023; Kostka et al. 2023), measurements of *opinion* 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 towards state surveillance is much lower[10], and self-censorship rates are also smaller, yet substantial. Furthermore, in this study, the average treatment effect was driven by older respondents and previous research on mass surveillance found a high tolerance among the older population, when it comes to state surveillance in post-soviet countries (Kalmus et al. 2022). 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.

The complexity of behavioral adaptations concerning self-censorship is reflected by finding *directed* and *undirected* self-censorship in this study. Directed self-censorship corresponds to falsifying preferences by giving a specific answer where socially desirable behavior is known. Conversely, undirected self-censorship relates to denying to state any opinion where socially desirable behavior is not known or not deductible. This can result from a lack of information or a lack of political literacy to evaluate relevant information. Citizens who are intensively using the internet showed directed self-censorship, whereas generally undirected self-censorship in the form of non-responses was found. This finding begs the question of how respondents inferred that they *should* answer that the invasion was justified and answered accordingly in the surveillance condition. Those with a higher degree of education also resort to *preferring*

[10] 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.

not to answer, while those characterized by a lower level of education tend to behave unresponsive under the effect of the treatments, on average. Assuming that political literacy correlates with higher levels of education, it seems that it is the informational difference rather than a difference in political literacy that matters when it comes to inferring socially desirable behavior. Thus, frequent use of the Internet and living in one of Kazakhstan's two large cities produce an informational difference relative to others that translates into different behavioral responses. Others found higher self-censorship among urban citizens (Robinson and Tannenberg 2019), a finding that is complemented here by showing a more nuanced differentiation in self-censorship patterns. In other words, the complexity of response options as well as the complexity of questions does seem to matter. When socially desired behavior is not easily deductible from the context, responses mirror this complexity in that they tend to be non-responses (i.e., *prefer not to answer*), instead of what is believed to be the correct response. In a similar vein, others have argued that chilling effects induced by surveillance "can best be understood as an act that conforms to, or is in compliance with, social norms in that context" (Penney 2022, p.1520).

Finally, there are cognitive components behind behavioural 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 dataveillance, which suggests including 'dataveillance imaginaries', i.e., the cognitive understanding of humans subject to (data) surveillance processes, which substantially shape behavioral responses (Kappeler et al. 2023). If anything, this study has contributed to shed light on the necessity for qualitative studies or mixed-method designs that complement and enhance results of quantitative studies.

6 Bibliography

Acquisti, A., Brandimarte, L., and Hancock, J. (2022). How privacy's past may shape its future. *Science*, 375(6578):270–272.

Agerberg, M. and Tannenberg, M. (2021). Dealing with measurement error in list experiments: Choosing the right control list design. *Research & Politics*, 8(2):20531680211013154.

Ahlquist, J. S. (2018). List Experiment Design, Non-Strategic Respondent Error, and Item Count Technique Estimators. *Political Analysis*, 26(1):34–53.

Alsan, M., Braghieri, L., Eichmeyer, S., Kim, M. J., Stantcheva, S., and Yang, D. Y. (2023). Civil liberties in times of crisis. *American Economic Journal: Applied Economics*, 15(4):389–421.

Bentham, J. (2011). *The Panopticon Writings*. Radical Thinkers. Verso Books.

Beraja, M., Kao, A., Yang, D., and Yuchtman, N. (2023a). Exporting the Surveillance State via Trade in AI. *Working Paper*.

Beraja, M., Kao, A., Yang, D. Y., and Yuchtman, N. (2023b). AI-tocracy*. *The Quarterly Journal of Economics*, 138(3):1349–1402.

Beraja, M., Yang, D. Y., and Yuchtman, N. (2022). Data-intensive Innovation and the State: Evidence From AI Firms in China. *The Review of Economic Studies*.

Blair, G., Coppock, A., and Moor, M. (2020). When to Worry About Sensitivity Bias: A Social Reference Theory and Evidence From 30 Years of List Experiments. *American Political Science Review*, 114(4):1297–1315.

Blair, G. and Imai, K. (2012). Statistical Analysis of List Experiments. *Political Analysis*, 20(1):47–77.

Blair, G., Imai, K., and Lyall, J. (2014). Comparing and Combining List and Endorsement Experiments: Evidence From Afghanistan. *American Journal of Political Science*, 58(4):1043–1063.

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

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

Coffman, K. B., Coffman, L. C., and Ericson, K. M. M. (2013). The Size of the LGBT Population and the Magnitude of Anti-Gay Sentiment are Substantially Underestimated. Working Paper 19508, National Bureau of Economic Research.

Corstange, D. (2012). Vote Trafficking in Lebanon. *International Journal of Middle East Studies*, 44(3):483–505.

Davis, D. and 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. and Crabtree, C. (2019). Domestic Demand for Human Rights: Free Speech and the Freedom-Security Trade-Off. *International Studies Quarterly*, 63(2):346–353.

Earl, J., Maher, T., and Pan, J. (2022). The Digital Repression of Social Movements, Protest, and Activism: A Synthetic Review. *Science advances*, 8.

Edmond, C. (2013). Information Manipulation, Coordination, and Regime Change. *The Review of Economic Studies*, 80(4 (285)):1422–1458.

Egorov, G. and Sonin, K. (2024). The Political Economics of Non-democracy. *Journal of Economic Literature*, Forthcoming.

Feldstein, S. (2019a). The Global Expansion of AI Surveillance. *Carnegie Endowment for International Peace*.

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: 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.

Foucault, M. (2012). *Discipline and Punish: The Birth of the Prison*. Vintage. Knopf Doubleday Publishing Group.

FreedomHouse (2023a). Kazakhstan: Freedom in the world 2023 country report.

FreedomHouse (2023b). Kazakhstan: Freedom on the net 2023 country report.

Frye, T., Gehlbach, S., Marquardt, K. L., and Reuter, O. J. (2017). Is Putin's Popularity Real? *Post-Soviet Affairs*, 33(1):1–15.

Frye, T., Gehlbach, S., Marquardt, K. L., and Reuter, O. J. (2023). Is Putin's Popularity (Still) Real? A Cautionary Note on Using List Experiments to Measure Popularity in Authoritarian Regimes. *Post-Soviet Affairs*, pages 1–10.

Glynn, A. N. (2013). What Can We Learn with Statistical Truth Serum?: Design and Analysis of the List Experiment. *Public Opinion Quarterly*, 77(S1):159–172.

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, Incorporated.

Guriev, S. and Treisman, D. (2019). Informational Autocrats. *Journal of Economic Perspectives*, 33(4):100–127.

Guriev, S. and Treisman, D. (2020). A Theory of Informational Autocracy. *Journal of Public Economics*, 186(104158).

Hobbs, W. R. and Roberts, M. E. (2018). How sudden censorship can increase access to information. *American Political Science Review*, 112(3):621–636.

Huang, G., Hu, A., and Chen, W. (2022). Privacy at Risk? Understanding the Perceived Privacy Protection of Health Code Apps in China. *Big Data & Society*, 9(2).

Imai, K. (2011). Multivariate regression analysis for the item count technique. *Journal of the American Statistical Association*, 106(494):407–416.

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

Kalmus, V., Bolin, G., and Figueiras, R. (2022). Who is Afraid of Dataveillance? Attitudes Toward Online Surveillance in a Cross-cultural and Generational Perspective. *New Media & Society*.

Kane, J. V. (2024). More than meets the itt: A guide for anticipating and investigating nonsignificant results in survey experiments. *Journal of Experimental Political Science*, page 1–16.

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

Karpa, D. and Rochlitz, M. (2023). Authoritarian Surveillance and Public Support for Digital Governance Solutions. *SSRN*, 4454627.

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

Kostka, G. (2019). China's Social Credit Systems And Public Opinion: Explaining High Levels of Approval. *New Media & Society*, 21(7):1565–1593.

Kostka, G. and Antoine, L. (2020). Fostering Model Citizenship: Behavioral Responses to China's Emerging Social Credit Systems. *Policy & Internet*, 12(3):256–289.

Kostka, G. and 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).

Kostka, G., Steinacker, L., and 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.

Kostka, G., Steinacker, L., and Meckel, M. (2023). Under Big Brother's Watchful Eye: Cross-country Attitudes Toward Facial Recognition Technology. *Government Information Quarterly*, 40(1):101761.

Kramon, E. and Weghorst, K. (2019). (Mis)Measuring Sensitive Attitudes with the List Experiment: Solutions to List Experiment Breakdown in Kenya. *Public Opinion Quarterly*, 83(S1):236–263.

Kudaibergenova, D. T. and 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.

Kuhn, P. M. and Vivyan, N. (2022). The Misreporting Trade-Off Between List Experiments and Direct Questions in Practice: Partition Validation Evidence from Two Countries. *Political Analysis*, 30(3):381–402.

Lee, H. J. (2023). 'i've left enough data': Relations between people and data and the production of surveillance. *Big Data & Society*, 10(1):20539517231173904.

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

Manokha, I. (2018). Surveillance, Panopticism, and Self-Discipline in the Digital Age. *Surveillance & Society*, 16:219–237.

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

Oz, M. and 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.

Pan, J. and Siegel, A. A. (2020). How saudi crackdowns fail to silence online dissent. *American Political Science Review*, 114(1):109–125.

Penney, J. W. (2016). Chilling Effects: Online Surveillance and Wikipedia Use. *Berkeley Technology Law Journal*, 31(1):117–182.

Penney, J. W. (2017). Internet Surveillance, Regulation, and Chilling Effects Online: a Comparative Case Study. *Internet Policy Review*, 2(6).

Penney, J. W. (2022). Understanding chilling effects. *Minnesota Law Review*, 106:1451.

Qin, B., Strömberg, D., and Wu, Y. (2017). Why Does China Allow Freer Social Media? Protests versus Surveillance and Propaganda. *Journal of Economic Perspectives*, 31(1):117–40.

Raman, R. S., Evdokimov, L., Wurstrow, E., Halderman, J. A., and Ensafi, R. (2020). Investigating large scale https interception in kazakhstan. In *Proceedings of the ACM Internet Measurement Conference*, IMC '20, page 125–132, New York, NY, USA. Association for Computing Machinery.

Roberts, M. E. (2018). *Censored: Distraction and Diversion Inside China's Great Firewall*. Princeton: Princeton University Press.

Robinson, D. and Tannenberg, M. (2019). Self-censorship of Regime Support in Authoritarian States: Evidence From List Experiments in China. *Research & Politics*, 6(3):2053168019856449.

Rosenfeld, B., Imai, K., and Shapiro, J. N. (2016). An Empirical Validation Study of Popular Survey Methodologies for Sensitive Questions. *American Journal of Political Science*, 60(3):783–802.

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

Schmidt, U. and 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*, pages 1–23.

Stoycheff, E. (2016). Under Surveillance. *Journalism & Mass Communication Quarterly*, 93(2):296–311.

Stoycheff, E. (2022). Cookies and Content Moderation: Affective Chilling Effects of Internet Surveillance and Censorship. *Journal of Information Technology & Politics*, pages 1–12.

Stoycheff, E., Liu, J., Xu, K., and Wibowo, K. (2019). Privacy and the Panopticon: Online Mass Surveillance's Deterrence and Chilling Effects. *New Media & Society*, 21(3):602–619.

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

Tannenberg, M. (2022). The Autocratic Bias: Self-censorship of Regime Support. *Democratization*, 29(4):591–610.

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.

Xu, X. (2022). The Unintrusive Nature of Digital Surveillance and Its Social Consequences. *Working Paper*.

A Appendix

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.

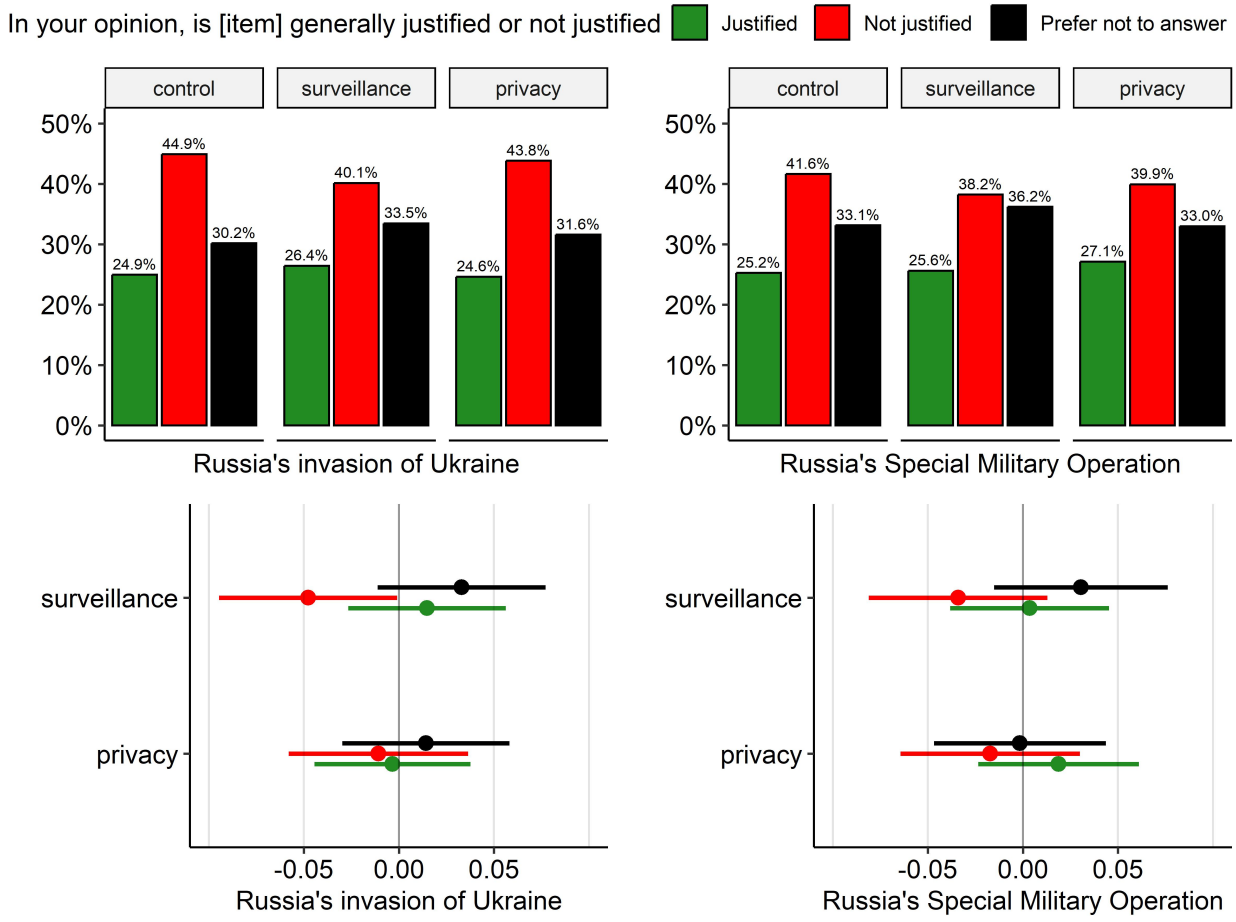
A.2 Additional Tables

Table 1: Summary statistics

Variable	N	Mean	SD
age	5025	42.7	16.1
education_scale	5025	4.55	1.75
ethnicity_kazakh	5025	0.722	0.448
government_employee	5025	0.109	0.312
government_trust_scale	4340	2.99	0.937
financial_situation_scale	5025	2.85	1.14
internet_scale	5025	4.47	1.8
large_city	5025	0.22	0.414
male	5025	0.487	0.5
religiosity_scale	3816	3.77	1.25
russian_media_consumption	5025	0.301	0.459
social_trust_scale	4485	2.57	0.87
vpn_user	5025	0.109	0.312
framing_invasion			
... invasion	5025	0.502	0.5
... SMO	5025	0.498	0.5

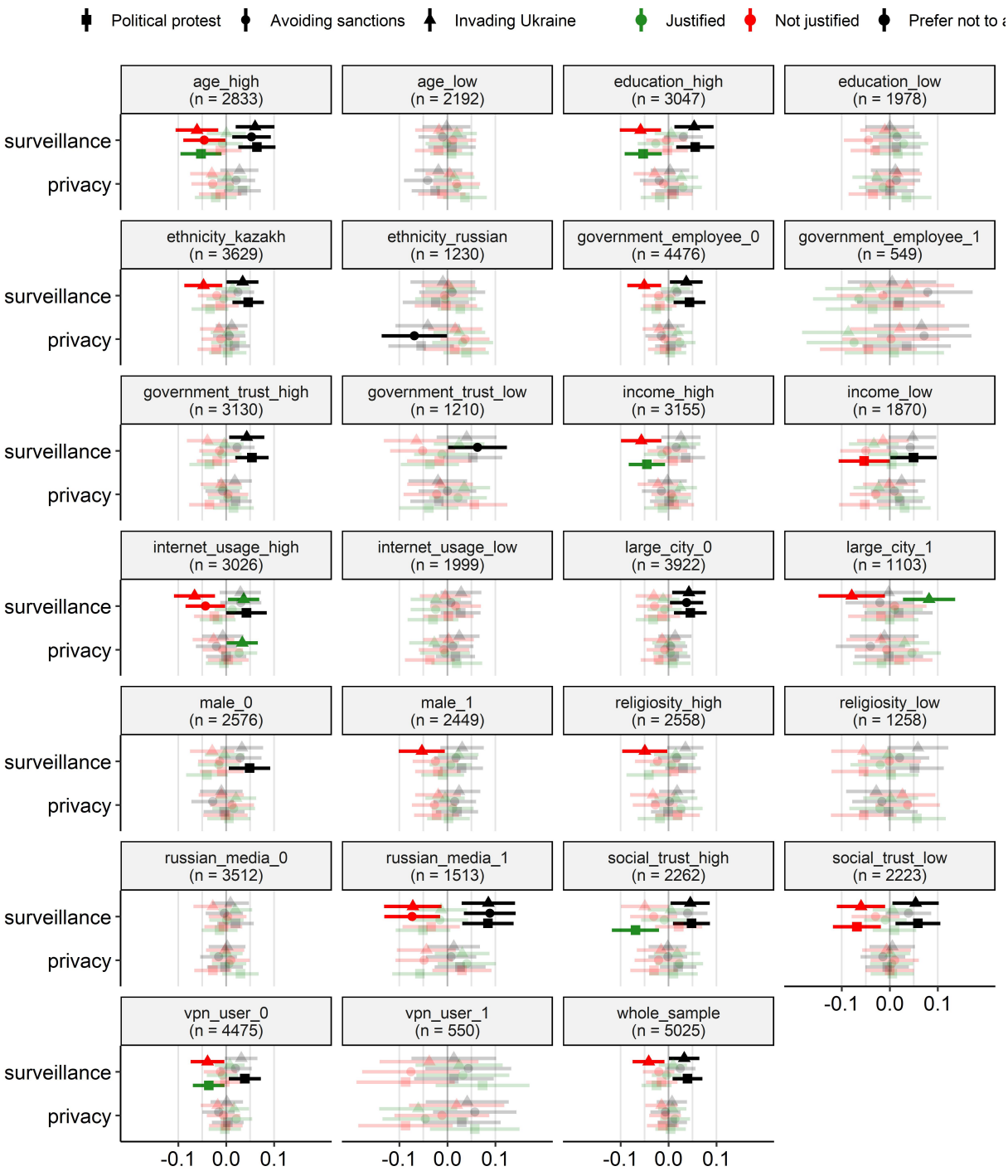
A.3 Additional Figures

Figure 10: Differences between framings of the invasion of Ukraine



Note: Top row: Responses to dependent variable by treatment condition. Bottom row: Comparisons of control and treatment groups; multinomial model point estimates with 95% level confidence intervals.

Figure 11: Conditional average treatment effects



Note: The shape of the points indicates the item posing as the dependent variable of the multinomial model, while the colors indicate the response options. Fully visible point estimates with confidence intervals are statistically significant different from the control condition at the 95% level, while transparent coefficients are not. Splits are made at the median or between 1 and 0 for binary variables.

A.4 List experiment design

List experiments – also known as the item count technique – have been successfully used, for example, to study support for authoritarian leaders (Blair et al. 2014; Frye et al. 2017; Robinson and Tannenberg 2019; Frye et al. 2023), estimating the size of LGBT population (Coffman et al. 2013), and vote trafficking in Lebanon (Corstange 2012).

Participants are exposed to either J or $J + 1$ items and then asked to count the number of items that apply to them, with the additional (J + 1th) item being the sensitive item of interest.[11] The premise of list experiments is that when a sensitive question is asked indirectly, respondents are more likely to give a truthful answer, even if social norms encourage them to answer the question in a particular way (Blair and Imai 2012). Fear of being judged or punished by others leads to a change in behavior best known as social desirability bias, a subset of what is known as sensitivity bias (Blair et al. 2020). Using list experiments as well as direct questions, the difference between honest and self-censored responses can be estimated (Robinson and Tannenberg 2019). In other words, comparing the means of direct and indirect responses can reveal strategic misreporting. More sophisticated statistical methods allow analysis beyond mean comparisons so that sensitivity bias can be shown, but also which sociodemographic factors and personality traits play a role (Imai 2011; Blair and Imai 2012).

In list experiments, there is a trade-off between being able to circumvent or measure *strategic* misreporting, and inevitable *nonstrategic* misreporting (Kuhn and Vivyan 2022). Nonstrategic measurement errors are, for example, floor and ceiling effects (Ahlquist 2018), and complexity (Kramon and Weghorst 2019). Floor and ceiling effects appear when either all or none of the statements in the list are true, and survey participants thus must reveal their preference over the sensitive item in order to remain giving truthful answers. In this case, one would have to expect strategic misreporting, i.e., lying. List experiments thus help to elicit true preferences, but, by design, suffer from overwhelming or confusing study participants. By careful consideration of the design, these measurement errors can be minimized (Glynn 2013; Blair et al. 2020), for example by choosing a suitable control list (Agerberg and Tannenberg 2021). Is it furthermore common practice to validate experiments before going into the field (Rosenfeld et al. 2016).

This study's list experiment takes this literature into account and is designed accordingly. The experiment tries to measure self-censorship by comparing the means of direct and indirect questioning, with respect to three different questions on politically sensitive items

[11] J is commonly equal to 3 or 4. Higher values substantially increase complexity, while lower values are risking to compromise the privacy of the participants (floor effect).

and one placebo item, which can be found below in Table 2. Non-strategic measurement errors were tried to be minimized according to suggestions from the literature:

First, floor and ceiling effects were avoided by (1) choosing items where the prevalence among the population was known beforehand, and (2) choosing control list items that are negatively correlated. For example, the items 'homosexuality' and 'holding on to religious values' are strongly negatively correlated. It is very unlikely that respondents agree with both items. Furthermore, it is unlikely that none of the items are true, statistically. Items were taken and adapted from the World Value Survey and Central Asian Barometer. Thus, ceiling effects and floor effects could, by design, be minimized. Second, sensitive control list items were chosen, in order to conceal the item of interest and not evoke demand effects. Third, a placebo item is used in order to test whether there are design effects (Frye et al. 2017, 2023). Fourth, a pre-test was run with 400 subjects in order to validate the lists. Last, a double list experiment was employed in order to enhance statistical power. In a double list experiment the sample is split into two groups, and each half acts as a control group for the other half, see Table 2.

Table 2: List experiment design

In the next section, you will be presented with a number of items. Please read the list of items carefully. After you have read the items, please indicate HOW MANY of the items (things or actions) below are generally justified in your opinion. We will not be able to know which items are justified for you.	
[randomization into one of the two columns below with equal probability]	
In your opinion, HOW MANY of the things or actions below are generally justified? [selection, randomization of list item order] Prioritizing industrial development above environmental safety Protecting the environment even if costs are high Prostitution Participating in protests for political change 0 1 2 3 4	In your opinion, HOW MANY of the things or actions below are generally justified? [selection, randomization of list item order] Prioritizing industrial development above environmental safety Protecting the environment even if costs are high Prostitution 0 1 2 3
In your opinion, HOW MANY of the things or actions below are generally justified? [selection, randomization of list item order] Homosexuality Helping Russia avoid Western sanctions Full time work for women Holding on to religious values 0 1 2 3 4	In your opinion, HOW MANY of the things or actions below are generally justified? [selection, randomization of list item order] Homosexuality Full time work for women Holding on to religious values 0 1 2 3
In your opinion, HOW MANY of the things or actions below are generally justified? [selection, randomization of list item order] Suicide Being proud of national traditions Aspiring to Western values 0 1 2 3	In your opinion, HOW MANY of the things or actions below are generally justified? [selection, randomization of list item order, randomization of framing: SMO or invasion - variable framing34] Suicide Being proud of national traditions Russia's Special Military Operation in Ukraine/ Russia's invasion of Ukraine Aspiring to Western values 0 1 2 3 4
In your opinion, HOW MANY of the things or actions below are generally justified? [selection, randomization of list item order] Death penalty Violating traffic rules Banning smoking in public places 0 1 2 3	In your opinion, HOW MANY of the things or actions below are generally justified? [selection, randomization of list item order, randomization of framing: SMO or invasion - variable framing34] Death penalty Violating traffic rules Working more than 50 hours per week Banning smoking in public places 0 1 2 3 4