Endogenous Popularity: How Perceptions of Support Affect the Popularity of Authoritarian Regimes

Noah Buckley, Kyle L. Marquardt, Ora John Reuter, Katerina Tertytchnaya

March 2022

Working Paper

SERIES 2022:132

THE VARIETIES OF DEMOCRACY INSTITUTE

UNIVERSITY OF GOTHENBURG

DEPT OF POLITICAL SCIENCE

Please address comments and/or queries for information to:

V–Dem Institute

Department of Political Science

University of Gothenburg

Sprängkullsgatan 19, PO Box 711

SE 40530 Gothenburg

Sweden

E-mail: contact@v-dem.net

V–Dem Working Papers are available in electronic format at www.v-dem.net.

Endogenous Popularity: How Perceptions of Support Affect the Popularity of Authoritarian Regimes*

Noah Buckley†
Kyle L. Marquardt‡
Ora John Reuter§
Katerina Tertytchnaya¶

*Paper presented at ASEEES 2021, SPSA 2022 and the CPEP working group at UCL. We commend Israel Marques for facilitating our work with the POADSRR surveys, and thank Henry Hale and David Szakonyi for their comments on earlier drafts.

†Trinity College Dublin
‡University of Bergen
§University of Wisconsin–Milwaukee
¶University College London

Abstract

Autocracies with popular leaders tend to survive longer. A growing body of scholarship therefore focuses on the factors that influence authoritarian popularity. However, it is possible that the perception of popularity itself breeds popularity in nondemocratic regimes, impacting incumbent approval. Here we use framing experiments embedded in four recent surveys in Russia to examine the extent to which information about the support an authoritarian leader enjoys influences the level of support respondents report for him. We find that negative information about the Russian president's popularity decreases support for him, while positive information has no effect. Additional analyses, which rely on a novel combination of framing and list experiments, provide evidence that these changes are not due to preference falsification. This study has implications for research on the origins of incumbent approval and dramatic defection cascades in nondemocratic regimes.

Tangible evidence that an autocratic leader is popular—such as favorable opinion polls or election victories—can prevent voters and elites alike from defecting from the regime. Such public evidence of support can thus bolster regime control (Geddes 1999, Reuter & Szakonyi 2021, Hale & Colton 2017, Tertytchnaya 2020) and is one important reason that autocracies with popular leaders tend to be more long-lived than those without such popular approval.

While authoritarian popularity can come from agreement with the autocrat's political positions (Colton & Hale 2009, Hale & Colton 2017) or positive assessments of their performance (Magaloni 2006, Treisman 2011), the *perception* of an autocrat's popularity may itself influence their popularity. Individuals may be more likely to express support for leaders if opinion polls and elections suggest that support for the authorities is widespread. In a similar vein, individuals may be less likely to report support when evidence suggests that support for the regime is in decline. A respondent's support could be sincere—they may truly approve of the leader more (or less) when they are perceived to be popular (or unpopular) —or insincere due to preference falsification.

Here we examine the extent to which perceptions of an autocrat's popularity influence their approval ratings. We do so with a framing experiment that presents respondents truthful information about popular support for Russia's authoritarian leader, President Vladimir Putin. We implemented this framing experiment between 2020 and 2021, using three nationally-representative surveys (two face-to-face and one online) and one regionally-representative online survey. Our framing experiment takes advantage of a unique circumstance: while a majority of Russians expressed support for Putin in public opinion surveys during our survey waves, Putin's ratings had recently sunk to historic lows. We were thus able to experimentally portray Putin's approval ratings in either a positive or negative light without deception. Across all survey waves, we find that inducing respondents to consider Putin's ratings as relatively low leads to lower levels of support for him. Showing respondents a frame that prompts them to consider Putin's approval as relatively high, however, does not influence their support for him.

Our research design also allows us to examine whether changes in support for the Russian president are driven by sincere preference updating or preference falsification. Specifically, we take advantage of the large sample size in the regionally-representative survey to pair our framing experiment with a list experiment. This methodologically innovative combination provides no evidence of preference falsification: list experiment results indicate that respondents in the negative treatment appear to be sincerely reporting lower support for the president.

Our findings have important implications for the study of both Russian politics and autocracy more generally. First, the results highlight the importance of social consensus—widespread perceptions of incumbent approval—to both Putin's rule (Greene & Robertson 2019) and authoritarian rule more broadly (Kuran 1995). In the Russian context, our findings imply that Putin's popularity is based not just on ideological attachment (Colton & Hale 2009), personal affinity (Hale & Colton 2017), and performance evaluations (Treisman 2011), but also on widespread perceptions about Putin's very popularity. Autocratic popularity is thus endogenous to itself.

1 The Popularity of Autocrats

Most mid-twentieth century dictatorships relied on to ensure social control. By contrast, many contemporary autocrats rely on manipulation and persuasion to ensure their control. To maintain a veneer of legitimacy, most contemporary autocrats also try to be popular (Guriev & Treisman 2019).

Autocrats can draw popularity from some of the same sources as democratic leaders: citizens may support the leader's ideological positions, programmatic positions, or character traits (Colton & Hale 2009, Hale & Colton 2017). The public may also support them because they believe the autocrat is competent and performing well in office (Magaloni 2006, Treisman 2011). Contemporary authoritarian regimes also try to *shape* citizen perceptions of the regime. Through their control of the media, electoral subversion, and the suppression of opposition voices, dictators elevate their own real and perceived popularity (Guriev & Treisman 2019).

Attempts by governments to portray an autocrat as popular may have important consequences for the leader's popularity. Specifically, though the phenomenon is understudied, it is plausible that perceptions of the regime's popularity can reproduce or dampen support for the regime. Simpser (2013) argues that perceptions of incumbent popularity can persuade potential challengers that the regime is invincible and that resistance to it is futile. In the case of Russia, recent work advances the notion that Putin's popularity is, in part, founded on social pressures to conform (Greene & Robertson 2017, Greene & Robertson 2019): Russians with socially conformist tendencies are overrepresented in Putin's support base. In a similar vein, Hale (2021) shows that the need to conform with the socially acceptable view could account for rally-round-the-flag effects.

This scholarship suggests that individuals are more likely to report supporting autocratic leaders when the leader is perceived to be popular. However, it is unclear to what extent these mechanisms reflect sincere support for an autocrat. Evidence that the autocrat enjoys widespread social support can help individuals infer information about the leader's quality. For example, opinion polls suggesting majority support for an incumbent may lead citizens to infer that the leader is competent and trustworthy. In this case, updating approval for an autocrat based on new public opinion data could reflect a sincere change in beliefs.

New information might also lead to sincere preference changes by communicating the dominant, socially desirable view in society (Lohmann 1994, Hale & Colton 2017, p.324). Indeed, a long line of research in sociology shows that individuals susceptible to social pressure are sincerely willing to conform with the views held by their fellow citizens, even if this means that they have to discount personal beliefs. By appearing to be in harmony with an important and meaningful reference group—here the rest of society—individuals could derive some positive utility (Edwards 1957, Hale 2021, p.2). In the political realm, evidence that the regime is popular may encourage some individuals to adopt and report more favorable assessments of the incumbent. In a similar vein, information that regime support is in decline or the perception that expressing opposition to the authorities is becoming the socially desirable view could encourage individuals to update their evaluations of the incumbent downwards and report lower support. In both cases, updating is sincere. Respondents align their preferences and beliefs with those of the perceived majority.

However, desire to conform with the majority may also encourage individuals to misreport their true views of the regime. When individuals express public preferences different from their private preferences, they engage in preference falsification. Individuals could report public views that contrast with their private beliefs for a number of reasons. They could do so because they strive for social approval (Tourangeau & Yan 2007), sense that their private views are not socially or politically desirable (Hale 2021), wish to maintain a socially favorable self-presentation (see, e.g. Stocket 2007, Tourangeau and Yan 2007), or are concerned about negative social sanctions, such as a sense of disapproval from the survey enumerator in the context of an interview. Indeed, across a range of contexts, social desirability considerations routinely lead people to either report views or to engage in behaviors that do not align with their private beliefs (Hale 2021, Maass & Clark 1983, Blair, Coppock & Moor 2020).

Finally, reputational cascade models hold that new information about regime support may encourage individuals who falsely reported support for the authorities to reveal their true preferences, believing that their preferences are more widely shared than previously thought (Lohmann 1994). For example, opinion polls suggesting that opposition to the regime is widespread could encourage individuals who privately disapprove of the authorities' performance to reveal their sincere preferences. Believing that support for the authorities is waning could encourage preference falsifiers to think that expressing opposition to the regime is now possible. On the flip side, evidence that the regime enjoys widespread support could encourage individuals who previously reported opposition to the authorities to reveal their support.

2 Authoritarian Popularity in Russia

Contemporary Russia is one of the world's most prominent autocracies; most observers agree that President Vladimir Putin's popularity is fundamental to the stability of the

Russian regime (Hale 2014, Greene & Robertson 2019). Since taking office in 2000, Putin has enjoyed popularity ratings that have never dropped below 60 percent. There is also substantial evidence that this support is sincere (Frye, Gehlbach, Marquardt & Reuter 2017, Frye, Gehlbach, Marquardt & Reuter 2021, Greene & Robertson 2019).

Although Putin's approval ratings have historically been quite high, they dropped to their lowest level in 2020-2021. After persisting above 80% for almost four years following the annexation of Crimea in 2014, Putin's approval rating dropped dramatically in early 2018 following an unpopular pension reform. Since then Vladimir Putin's popularity has hovered just above 60%. As we discuss below, this allows one to view Putin's popularity in both a 'positive' and 'negative' light. On the one hand, opinion polls indicate that a sizeable majority still supports him. On the other hand, his popularity has declined dramatically in recent years, having sunk to its lowest level ever.

3 Research Design

Our goal is to estimate how perceptions of Putin's popularity affect support for him. In principle, a researcher could simply ask respondents directly whether their support for Putin is influenced by such considerations. Indeed, the Levada Center, Russia's most respected polling agency, routinely includes this consideration in a list of options of respondents can select as reasons they support Putin. While assessments of Putin's experience, decisiveness, leadership and perceived accomplishments routinely top the list, the president's perceived popularity also matters. In multiple surveys in the 2000s, for example, 12–17% of respondents note that they support Putin because he "has the respect of people around me."

While such responses are interesting, they cannot form the basis for reliable inferences about how adherence to social norms drives Putin's popularity. For one, respondents who sincerely adhere to social norms about supporting Putin are likely to rationalize their support by identifying concrete reasons that they support Putin. Indeed, these social norms themselves—e.g., a socially accepted belief that Putin is a strong leader—may lead respondents to choose those concrete justifications for supporting Putin. Moreover, respondents might loathe to admit that they are so easily swayed by the opinion of those around them. This concern would be especially valid for respondents who misrepresent their true preferences when they think it is socially desirable to do so.

Another way of addressing this question is to look at the association between support for Putin and a respondent's beliefs about Putin's popularity. We were only able to identify two instances in which this question was posed: in March 2015, when respondents were asked about perceptions of Putin's support levels among their friends and family; and in August 2018, when respondents were asked to estimate Putin's popularity in society. In both cases, support for Putin is very strongly associated with believing that Putin is popular.[1]

However, respondents may have drawn conclusions about Putin's popularity based on their own support: respondents might generalize their own views on Putin's popularity to broader swathes of society. Similarly, support for Putin and beliefs about support for Putin are co-determined by a litany of unobserved social and spatial factors that bedevil any attempt to make causal inferences on the basis of association. Finally, there is the concern that responses about the share of Russians that support the Russian president may be subject to preference falsification, like the questions about personal support.

To exogenously manipulate respondents' beliefs about Putin's popularity, we therefore employ a framing experiment that attempts to shift respondents' perceptions about the popularity of the regime. To our knowledge, this is the first effort by scholars to explicitly examine the effects of different frames of societal approval levels on respondents' own reported support for the regime. Our approach leverages the fact that current levels of support for Putin in Russia are objectively high, but still much lower than in recent memory. This makes it possible to frame Putin's popularity in both positive and negative light, without deceiving respondents or fabricating numbers. The phrasing of the survey experiment is given in Figure 1.

Figure 1: Framing experiment

Control: On the whole, how much do you support the activities of the President of Russia?

Positive frame: Sociological surveys unanimously show that, on the whole, 2/3 of Russians support the activities of the President of Russia. The President enjoys stable support from the population- a strong majority of Russians support the activities of the President of Russia. On the whole, how much do you support the activities of the President of Russia?

Negative frame: Sociological surveys unanimously show that only 2/3 of Russians support the activities of the President of Russia. This is the lowest level of support for the President of Russia in recent years. On the whole, how much do you support the activities of the President of Russia?

- Completely do not support
- Mainly do not support
- Mainly support
- Completely support

Both the positive and negative frame provide the respondents with the same information: close to 67% of Russians have reported support for Putin in recent surveys when asked directly (63% in our November 2020 pilot survey).[2] The positive frame notes that this quantity represents a strong and stable majority, while the negative frame notes that

[1] In 2018, the modal estimate of Putin's popularity among opponents of Putin was 30-40%. Among Putin supporters, the estimate was approximately 65%.

[2] In the initial pilot conducted by the Levada Center, we referred to the president by name, i.e. 'Vladimir Putin, the President of Russia'; the framing wording also used 'social' as opposed to 'sociological' and the scales were slightly different. Given the broad similarity in results between the pilot and the other three surveys, we infer that these differences are only marginal and do not drive our results.

only that many Russians support Putin and that his approval rating is lower than it has been in recent years.

As noted in the previous sections, respondents who update in response to these experimental primes may be doing so because they sincerely update their preferences for Putin, or because they are misrepresenting (or ceasing to misrepresent) their true preferences. In order to investigate whether this updating is driven by a sincere change in preferences, we directly followed the framing experiment with a list experiment in a large-scale online survey. List experiments allow respondents to reveal support for a political figure in aggregate without doing so individually. Specifically, respondents are exposed to either a control or treatment list and asked to report the *number*—not *which*—of items pertain to them (Blair, Coppock & Moor 2020). The lists are identical, save that the treatment list includes the sensitive item in addition to the other items on the control list. Since respondents do not report the items themselves respondents in the treatment group do not reveal if the sensitive item pertains to them. However, since the only difference between the control and treatment lists is the presence of the sensitive item, the average difference between control and treatment responses should reflect the prevalence of the sensitive item in the population.

Figure 2 illustrates our list experiment phrasing, for which we used the lists in Frye et al. (2017) as a template. The control list includes international political figures, to whom we refer using their title. The treatment list also includes the Russian president.

Figure 2: List experiment

Take a look at this list of politicians and tell me for how many you generally support their activities:

- The President of the USA
- The Chancellor of Germany
- The President of Belarus
- **The President of Russia**

Support: 0 1 2 3 **4**

Crucially, because the list experiment comes directly after the framing experiment, the framing effects should spill over into the list experiment. As a result, we can estimate the degree to which the frames effect support for the president in the list responses, as well as in the direct responses. Specifically, the difference between estimated support for the president in the negative and positive frames vs. the control should reflect actual changes in support, as opposed to changes induced by preference falsification. As follows from our experimental design, to estimate levels of preference falsification we compare the mean estimate of support for the President of Russia in the above list experiment with the level of support indicated in an identically formulated direct question. The setup allows us to assess whether or not observed differences between framing conditions could be attributable to changes in levels of preference falsification, as opposed to sincere updating.

Theory leads us to expect that the positive frame should increase support for Putin and the negative frame should dampen his support. These changes may be either sincere or insincere. The positive frame could persuade some citizens that Putin is competent, so that they sincerely report (greater) support for him. It could also induce encourage preference falsification by signaling that support for Putin is the desirable response, causing respondents to insincerely report support for him.

The negative frame could persuade respondents that Putin is less competent than they believed, causing them to sincerely lower their support for him. The negative frame could also induce preference falsification in two ways. First, by suggesting that respondents are not alone in disapproving of the authorities, it might give insincere regime supporters 'permission' to voice dissent. In this context, the frame reduces preference falsification. Alternatively, the negative frame could signal to regime supporters that support is socially undesirable, causing them to insincerely report opposition to Putin.

If results from the combined framing and list experiment are similar to those from the framing experiment alone, it is evidence that the frames result in a sincere change in preferences.

4 The Data

We analyze data from four surveys fielded in Russia between November 2020 and September 2021. The Levada and Russian Election Study (RES) surveys are nationally representative face-to-face surveys implemented by the Levada Center. The Public Opinion on Analog and Digital Services in Russia's Regions (POADSRR, `osf.io/rp7b5/`) surveys are nationally and regionally representative, respectively; they were fielded online using a sample frame provided by a well-regarded online polling center. Both the Levada and POADSRR nationally-representative survey were pilots for the RES and POADSRR regionally-representative surveys. Since the changes between the pilots and pre-registered surveys were minimal we report the results together.[3] All surveys included the framing experiment. Only the POADSRR surveys included the framing × list experiment. Since the nationally-representative POADSRR survey was severely underpowered for this framework (and not pre-registered), we only report framing and list results from the regionally-representative survey.

5 Models and Results

To estimate the direct effect of the negative and positive frames on support for President Putin, we dichotomize the 4-pt Likert scale support for Putin (President of Russia)

[3]Pre-registration available at `osf.io/8fj2q/`.

question, coding the top two categories as 1 ("support") and the bottom two categories as 0 ("do not support"). We use a linear probability model to regress this outcome on dichotomous indicators for the Negative and Positive frame, leaving the control condition as the reference category:[4]

$$y_i = \alpha_1 + \alpha_2 Negative_i + \alpha_3 Positive_i + \epsilon_i \tag{1}$$

To estimate framing effects in the list experiment, we use standard linear regression.[5] Specifically, we regress the number of political figures a respondent reports supporting on 1) an indicator for the list experiment treatment, 2) indicators of the framing treatments, and 3) the interaction of the experimental treatments:

$$y_i = \beta_1 + \beta_2 Negative_i + \beta_3 Positive_i + \alpha_1 List_i + \alpha_2 List_i \times Negative_i + \alpha_3 List_i \times Positive_i + \epsilon_i \tag{2}$$

Here, the quantities of interest are denoted by α. α_1 represents the estimated proportion of the population which supports Putin in the framing control condition. α_2 and α_3 represent the difference in this proportion between the control and the negative and positive framing conditions, respectively. β represents coefficients pertaining to the control list, and are of little substantive interest.

Table 1 reports the results from these analyses, which are remarkably consistent across survey waves. Columns 1–4 show the direct effect of the two experimental frames on support for Putin, while Column 5 estimates indirect framing effects estimated in the framing $\times$ list experiment. The top row in Table 1 shows the estimated prevalence of support for the Russian president in the control condition (α_1), while the second and third rows report the effect of the positive and negative frames on this proportion (α_2 and α_3); the last three rows show the corresponding statistics for the control list (β_1–β_3) in the list experiment (Column 5).

In all survey waves, the positive frame shows little substantive effect and is statistically insignificant. In contrast, the negative frame shows a consistently significant and substantively strong effect across direct responses: a 6-11% decrease in estimated support. These results are consistent across both the direct estimates (Columns 1–4) and the indirect (list) estimate (Column 5). The fact that the list experiment yielded similar results to those with directly-stated outcomes constitutes strong evidence that results from the framing experiment are attributable to sincere changes in preferences, as opposed to preference

[4]We use dichotomized outcomes so that the results are 1) comparable to those we obtain in the framing x list experiment and 2) easily interpretable on standard dichotomous popularity scale. Results are largely robust if we analyze the full outcome range using an ordered probit models (Appendix A.5).

[5]We also implement a pre-registered algorithm to clean the list experiment data, reported in Appendix A.1.

Table 1: Framing effects on support for President Putin

	Levada National Nov 2020	POADSRR National Jun 2021	RES National Sep 2021	POADSRR Regional Aug 2021	POADSRR (List) Regional Aug 2021
			Support for the president		
Constant	0.63***	0.52***	0.67***	0.56***	0.56***
	(0.02)	(0.02)	(0.02)	(0.01)	(0.03)
Positive	−0.02	0.01	−0.02	−0.002	−0.05
	(0.03)	(0.03)	(0.03)	(0.01)	(0.04)
Negative	−0.08**	−0.06*	−0.07**	−0.11***	−0.12***
	(0.03)	(0.03)	(0.03)	(0.01)	(0.04)
			Control list		
Constant					1.00***
					(0.02)
Positive					0.02
					(0.03)
Negative					0.01
					(0.03)
Observations	1,554	1,503	1,277	16,324	14,577
R^2	0.004	0.003	0.004	0.01	0.06

Note: *p<0.1; **p<0.05; ***p<0.01

All analyses use linear regression(dichotomized outcome for Columns 1–4). The control list constant is the number of items respondents report supporting in the control condition.

falsification.[6] These results constitute strong evidence that, when respondents are exposed to novel and negative information about the Putin's popularity, a substantial proportion sincerely revises their support for him downward.

In addition to these main results, there is one other important consistency across survey waves: treatment effects are largely concentrated in the bottom three categories (Table 2). That is, the proportion of respondents who 'completely' support President Putin is largely consistent across framing treatments. Much of the experimental effects involves a shift in respondents from the 'Mainly support' to the 'Mainly do not support' category. This result is evidence that, although negative information can reduce the probability respondents report support for the president, this effect is largely limited to those individuals with weaker preferences.

Table 2: Change in distribution of support for Russian president across framing conditions

	Completely do not support	Mainly do not support	Mainly support	Completely support
POADSRR Control	0.17	0.27	0.43	0.13
POADSRR Positive frame	0.17	0.27	0.43	0.12
POADSRR Negative frame	0.22	0.32	0.34	0.11
RES Control	0.15	0.19	0.52	0.14
RES Positive frame	0.12	0.23	0.49	0.16
RES Negative frame	0.14	0.27	0.44	0.15

POADSRR data from regionally-representative survey.

6 Conclusion

Autocrats in the 21st century are media-savvy. In place of overt repression, they spin the news to convince the masses that they are popular (Guriev & Treisman 2020, Guriev & Treisman 2019). Here we examine a potential reason why this spinning may be particularly important: perceptions of incumbent popularity might themselves breed popularity in nondemocratic regimes, inflating incumbents' approval. To the best of our knowledge, this study provides one of the first experimental tests of the degree to which perceptions of incumbent approval influence public opinion in these regimes.

The empirical analysis uses a series of framing experiments, embedded in four surveys of public opinion in Russia. We find that a frame revealing relatively low support Putin makes respondents less likely to report support for him; the frame revealing relatively high

[6] We investigate the possibility of preference falsification across experimental conditions more rigorously, and find little evidence of this phenomenon. Similarly, we conduct analyses of heterogenous treatment effects and find little evidence of such effects. (Appendix A.1)

levels of support has no significant effect. A combined framing-list experiment suggests that the results from the framing experiment are, in fact, due to sincere updating of preferences.

These results imply that shaping perceptions—through propaganda, indoctrination, schools, the media, and, indeed, opinion polls—is an important element of authoritarian popularity and thus stability.[7] However, our results also suggest limits on illiberal regimes' ability to bolster support: information about Putin's approval as both stable and high does not bolster support for him.

Finally, we note that support that relies on perceptions is fragile. Indeed, our results show that relatively mild negative information can reduce support for an autocrat by 6-11%. This fragility has important implications for regime stability. According to Greene & Robertson (2017), when unanimity or social consensus breaks down, regimes can dissolve very quickly. Such cascades are likely to be even more abrupt when consensus rests on perceptions, as opposed to when consensus is manufactured through intimidation, normative congruence or ideological agreement (Easton 1975). Individuals who support the authorities because they think that the authorities are popular may withdraw support when they think that others around them have begun to do the same, leading to cascade effects.

References

Bartels, Larry. 1988. *Presidential Primaries and the Dynamics of Public Choice*. Princeton, NJ: Princeton UP.

Blair, Graeme, Alexander Coppock & Margaret Moor. 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.

Colton, Timothy J. & Henry E. Hale. 2009. "The Putin Vote: Presidential Electorates in a Hybrid Regime." *Slavic Review* 68(3):473–503.

Easton, David. 1975. "A Re-Assessment of the Concept of Political Support." *British Journal of Political Science* 5(4):435–457.

Edwards, Allen L. 1957. *The Social Desirability Variable in Personality Assessment and Research*. Ft Worth, TX: Dryden Press.

[7]Note that our results may not hold in settings where attitudes toward the regime are hardened by entrenched class cleavages (e.g., Venezuela), ethnic divides (e.g., Malaysia), or legacies of conflict (e.g., Mozambique, Zimbabwe).

Farjam, Mike. 2021. "The Bandwagon Effect in an Online Voting Experiment With Real Political Organizations." *International Journal of Public Opinion Research* 33(2):412–421.

Frye, Timothy, Scott Gehlbach, Kyle L. Marquardt & Ora John Reuter. 2017. "Is Putin's Popularity Real?" *Post-Soviet Affairs* 33(1):1–15.

Frye, Timothy, Scott Gehlbach, Kyle L. Marquardt & Ora John Reuter. 2021. "Measuring Support for Political Elites under Autocracy: The Promise and Pitfalls of Item-Count Techniques.".

Geddes, Barbara. 1999. "What Do We Know About Democratization After Twenty Years?" *Annual Review of Political Science* 2(1):115–144.

Gerber, Alan S., Donald P. Green & Christopher W. Larimer. 2008. "Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment." *American Political Science Review* 102(1):33–48.

Glynn, Carroll J. & Jack M. McLeod. 1984. "Public Opinion Du Jour: An Examination of the Spiral of Silence." *Public Opinion Quarterly* 48(4):731–740.

Greene, Samuel & Graeme Robertson. 2017. "Agreeable Authoritarians: Personality and Politics in Contemporary Russia." *Comparative Political Studies* 50(3):1802–1834.

Greene, Samuel & Graeme Robertson. 2019. *Putin v. the People*. New Haven, CT: Yale UP.

Guriev, Sergei & Daniel Treisman. 2019. "Informational Autocrats." *Journal of Economic Perspectives* 33(4):100–127.

Guriev, Sergei & Daniel Treisman. 2020. "A Theory of Informational Autocracy." *Journal of Public Economics* 186:104158.

Hale, Henry E. 2014. *Patronal Politics: Eurasian Regime Dynamics in Comparative Perspective*. Cambridge: Cambridge University Press.

Hale, Henry E. 2021. "Authoritarian Rallying as Reputational Cascade? Evidence from Putin's Popularity Surge after Crimea." *American Political Science Review* pp. 1–15.

Hale, Henry E. & Timothy J. Colton. 2017. "Who Defects? Unpacking a Defection Cascade from Russia's Dominant Party 2008–12." *American Political Science Review* 111(2):322–337.

Klar, Samara, Christopher R. Weber & Yanna Krupnikov. 2016. "Social Desirability Bias in the 2016 Presidential Election." *The Forum* 14(4):433–443.

Knack, Stephen. 1992. "Civic Norms, Social Sanctions, and Voter Turnout." *Rationality and Society* 4(2):133–156.

Kuran, Timur. 1995. "The inevitability of future revolutionary surprises." *American Journal of Sociology* 100(6):1528–1551.

Lohmann, Susanne. 1994. "The Dynamics of Informational Cascades: The Monday Demonstrations in Leipzig, East Germany, 1989-91." *World Politics* 47(1):42–101.

Maass, Anne & Russell D. Clark. 1983. "Internalization versus Compliance: Differential Processes Underlying Minority Influence and Conformity." *European Journal of Social Psychology* 13(3):197–215.

Magaloni, Beatriz. 2006. *Voting for Autocracy: Hegemonic Party Survival and Its Demise in Mexico.* Cambridge Studies in Comparative Politics Cambridge: Cambridge University Press.

Reuter, Ora John & David Szakonyi. 2021. "Electoral Manipulation and Regime Support: Survey Evidence from Russia." *World Politics* 73(2):275–314.

Simon, Herbert A. 1954. "Bandwagon and Underdog Effects and the Possibility of Election Predictions." *The Public Opinion Quarterly* 18(3):245–253.

Simpser, Alberto. 2013. *Why Governments and Parties Manipulate Elections: Theory, Practice, and Implications.* Cambridge: Cambridge University Press.

Tertytchnaya, Katerina. 2020. "Protests and Voter Defections in Electoral Autocracies: Evidence From Russia." *Comparative Political Studies* 53(12):1926–1956.

Tourangeau, Roger & Ting Yan. 2007. "Sensitive Questions in Surveys." *Psychological Bulletin* 133(5):859–883.

Treisman, Daniel. 2011. "Presidential Popularity in a Hybrid Regime: Russia under Yeltsin and Putin." *American Journal of Political Science* 55(3):590–609.

A The POADSRR regionally-representative survey

Given the large sample size of the POADSRR regionally-representative survey (16,342), we conducted analyses of these data to both estimate preference falsification across framing experiment conditions and investigate heterogeneous treatment effects across these conditions. We pre-registered these analyses based on results from the nationally-representative POADSRR survey.

A.1 List experiment cleaning algorithm

Analyses of the POADSRR nationally-representative (pilot) survey indicated that a substantial proportion of respondents in the online setting falsify their responses in the treatment condition. Specifically, many respondents reported supporting only one or fewer of the political figures in direct questions, but reported supporting the maximum number of figures (four) in the treatment list.[8] This pattern results in drastic inflation of estimated support for the Russian President.

Based on these results, we pre-registered a cleaning algorithm that we then implemented in the POADSRR regionally-representative survey. Specifically, we clean the dataset such that respondents in the control group can only report ± 1 the number of direct figures they report supporting in the control list, while respondents in the treatment group can only report only one fewer figures and two more. We removed respondents who violated these conditions from the cleaned dataset.

In principle, this procedure might inflate the estimates of the sensitive item (some people who are two more on the treatment are doing so in error, not because of support). On the other hand, this approach might be underestimating support because it remove respondents who clearly support the President (those who reported 0-1 figures in the control directs and four in treatment).

In the text, we report only analyses from the cleaned dataset. However, in this appendix we report results from from both the full dataset and a dataset in which we remove probable list falsifiers (the "cleaned" dataset). We analyze both datasets—as opposed to just the cleaned dataset—for the sake of robustness, though evidence of systematic trends in those who engage in preference falsification means that the cleaned dataset should take precedence in the case of discrepancies.

A.1.1 List experiment diagnostics

Prior to proceeding to the analyses, we provide some diagnostics related to the cleaning algorithm. First, Table A.1 shows the most important diagnostic. Rows represent the number of political figures a respondent reported supporting in direct questions, while columns represent the number they report supporting in list. Italics are on the diagonal (in the case of the treatment list, both the diagonal and diagonal plus one are italicized), showing respondents who report this number with error. Bold denotes the problem values: respondents who reported supporting 4 figures on the treatment list, and 0-1 in the direct questions.

[8] Prior to the list experiment, respondents were asked to directly report whether or not they supported the activities of each of the three control list figures: 1) the President of the USA, 2) the Chancellor of Germany, and 3) The President of Belarus. The sum of these three responses is the number of figures a respondent directly supports.

Table A.1: Number of figures supported directly vs. in list

Control list

	0	1	2	3
0	*2376*	253	167	100
1	262	*2022*	524	86
2	80	301	*1159*	101
3	82	145	135	*357*

Treatment list

	0	1	2	3	4
0	*1196*	*1057*	124	51	**403**
1	211	*973*	*1348*	99	**389**
2	69	216	*692*	*484*	161
3	63	112	95	*147*	*289*

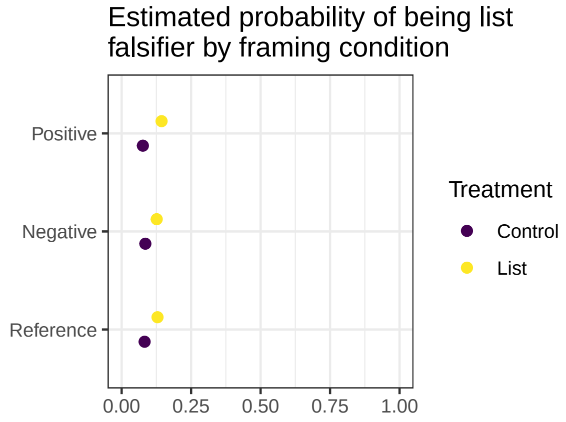
Rows represent number of figures supported in direct questions; columns the number of figures supported in list.

In principle, these results could be due to floor effects, a grave concern in list experiments: respondents who support none of the control list figures and do not support the president might still feel compelled to report "1" on the treatment list so as not to reveal their lack of support for the president. However, there is no literature of which we are aware that suggests that such respondents would drastically over-compensate by reporting more than 1.

In this context, this overcompensation creates a huge inferential problem because it inflates the number of respondents at the ceiling of the treatment list and thus the estimated difference between the control and treatment lists. As a result, it almost certainly results in an overestimate of support for the sensitive figure. We therefore remove these respondents (as well as other respondents whose list responses diverge substantially from their direct responses) from the dataset.

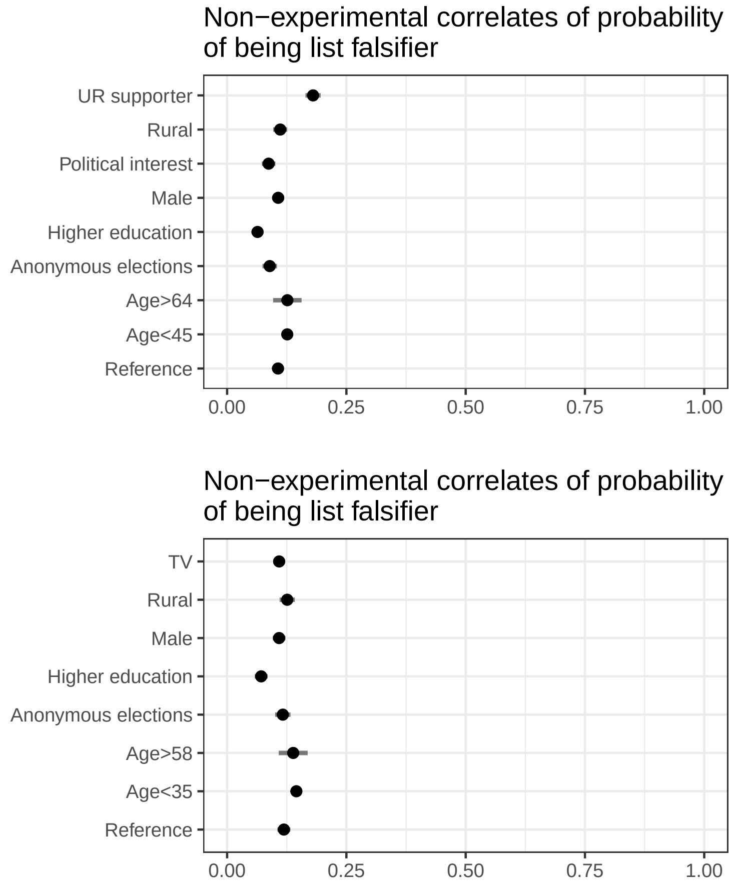
To further investigate these results, we also create a dichotomous indicator for list-falsifiers (i.e. those respondents whom we remove from the "cleaned" dataset). Figures A.1, A.2 and A.3 report the predictors of being a list falsifier, both by framing effects and with heterogenous treatment effects (description of covariates in Figure A.4). Note that the top cell shows little evidence that framing affects the probability of being a list falsifier. Results from analyses of demographic correlates indicate that United Russia (UR—the party of the Russian President) supporters are the most likely to be list falsifiers, while those with higher education are the least.

Figure A.1: Probability of being list falsifier by experimental condition



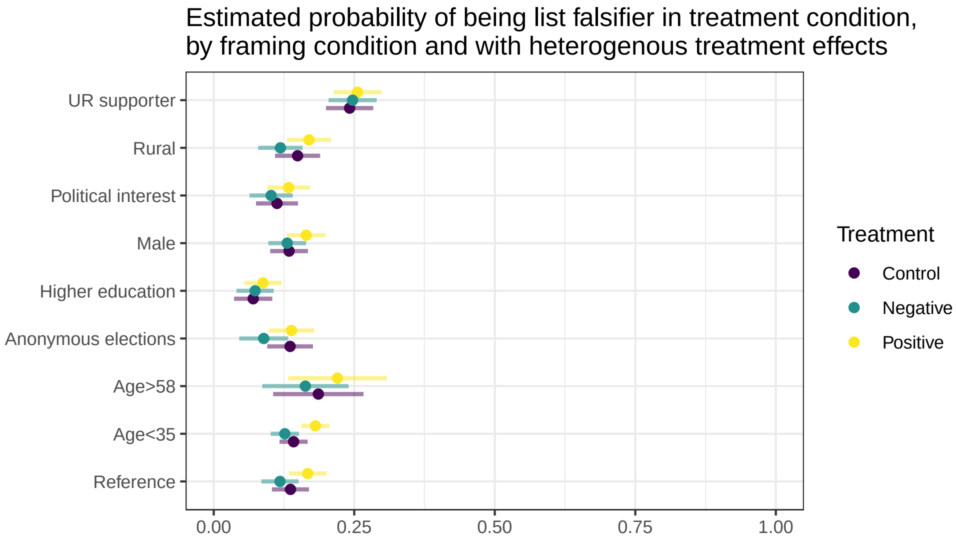
Analyses use predicted probabilities from linear probability model. Horizontal lines represent 95% confidence intervals about predicted probabilities.

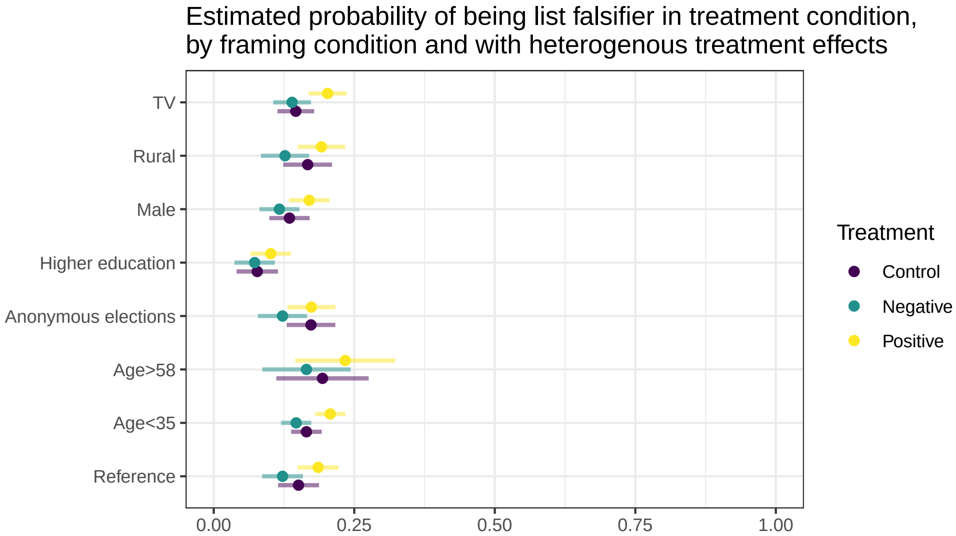
Figure A.2: Probability of being list falsifier by demographic correlates



Analyses use predicted probabilities from linear probability model. Horizontal lines represent 95% confidence intervals about predicted probabilities.

Figure A.3: Probability of being list falsifier, heterogeneous treatment effects





Analyses use predicted probabilities from linear probability model. Horizontal lines represent 95% confidence intervals about predicted probabilities.

A.2 Analyses of direct and indirect treatment effects

In this appendix our baseline analyses are the same as in the text. First, to estimate the direct effects of the framing experiment we dichotomize the 4-pt Likert scale support for Putin (President of Russia) question, coding the top two categories as 1 ("support") and the bottom two categories as 0 ("do not support"). We use a linear probability model to regress this outcome on dichotomous indicators for the Negative and Positive frame, leaving the control condition as the reference category:

$$y_i = \alpha_1 + \alpha_2 Negative_i + \alpha_3 Positive_i + \epsilon_i \tag{A.1}$$

To estimate indirect support for the president, we analyze the list experiment data. Specifically, we use a standard ordinary least squares analysis to regress the number of political figures (0-3/4) a respondent reports supporting on 1) an indicator for the list experiment treatment, 2) indicators of the framing treatments, and 3) the interaction of the experimental treatments:

$$y_i = \beta_1 + \beta_2 Negative_i + \beta_3 Positive_i + \alpha_1 List_i + \alpha_2 List_i \times Negative_i + \alpha_3 List_i \times Positive_i + \epsilon_i \tag{A.2}$$

Here, the quantities of interest are denoted by α. α_1 represents estimated proportion of the population which supports for Putin in the list experiment in the control framing condition, and α_2 and α_3 the equivalent proportions in the negative and positive framing conditions. β represents coefficients pertaining to the control list, which serve mainly to check for design issues in the experimental framework: the framing experiment should not influence the number of political figures a respondent supports in the control list.

We conducted balance checks on all experimental conditions (framing and list) using standard demographic covariates (the natural logarithm of age and indicators for male respondents and those with higher education), and found little evidence of imbalance.

Note that we analyze both the full and the cleaned POADSRR datasets—as opposed to just the cleaned dataset—for the sake of robustness, though evidence of systematic trends in those who engage in preference falsification means that the cleaned dataset should take precedence in the case of discrepancies.

Table A.2 presents results regarding both direct and indirect support for Russian President Putin. In all columns, the first three rows represent coefficient estimates for α; the remaining three rows β estimates (for the list experiments). The first column shows results for the direct responses to the framing experiment, the second and third results from the framing $\times$ list experiment (cleaned and full dataset, respectively). In all experiments, we can reject the null hypothesis of no effect of the negative frame; we cannot reject the null for the positive frame.

The very similar results between the direct question and the cleaned list provide strong evidence that there is little preference falsification in the results, indicating that the negative frame induces preference falsification. It is also worth noting that the magnitude of the negative frame's effect is similar in the full list data, further supporting this argument. The constant (control) condition in the full list indicates substantial preference falsification in support for Putin in that the estimate of support is substantially higher in these data; however, this result is likely due to list falsifiers.

Table A.2: Estimated support for President across experimental conditions

	Direct (LPM)	List (OLS)	
		Cleaned	Full
	Support for President		
Constant	0.56***	0.56***	0.72***
	(0.01)	(0.03)	(0.03)
Positive Frame	−0.002	−0.05	−0.01
	(0.01)	(0.04)	(0.04)
Negative Frame	−0.11***	−0.12***	−0.13***
	(0.01)	(0.04)	(0.04)
	Control items		
Constant		1.00***	1.05***
		(0.02)	(0.02)
Positive Frame		0.02	0.01
		(0.03)	(0.03)
Negative Frame		0.01	0.01
		(0.03)	(0.03)
Observations	16,324	14,577	16,324
R^2	0.01	0.06	0.08
Note:			*p<0.1; **p<0.05; ***p<0.01

A.3 Estimating preference falsification

To estimate preference falsification, we compare results from the direct and list experiments. Doing so requires several steps. First, we take a random draw from the distribution of α to estimate the probability that a respondent in both the list treatment condition and a given framing condition would support the President. For example, the probability that a respondent in the negative framing condition would support the President is distributed normally with a mean of $\alpha_1 + \alpha_2$ and a standard deviation $\sqrt{\sigma^2_{\alpha_1} + \sigma^2_{\alpha_2} + 2 \times Cov(\alpha_1, \alpha_2)}$, restricted to values between 0 and 1. We then take a draw from a Bernoulli distribution using this probability to estimate whether or not a respondent supported the president. Finally, we estimate the difference in means between these estimates and the indicators of support we used in the direct experiment. (Note: We only use data from respondents in the list treatment condition to avoid inflating the N; in the cleaned dataset we only data from respondents who are not list falsifiers).

Table A.3 provides the results from theses for both the full dataset and and the cleaned dataset. Results from both datasets are inconsistent, due to the influence of list falsifiers in the experiment. In the cleaned dataset, it is worth noting that the president is estimated to be *less* popular in the list than in the direct positive frame.

Table A.3: Estimated political desirability bias and design effects in support for president, across experimental conditions

	Full	Cleaned
Control	-0.15(-0.18, -0.13)	-0.01 (-0.04, 0.01)
Positive	-0.15 (-0.18, -0.13)	0.04 (0.01, 0.07)
Negative	-0.11 (-0.14, -0.08)	-0.01 (-0.04, 0.01)

Point estimates represent average estimated difference in support for president between direct and list experiments, with associated 95% confidence intervals. Negative values indicate that estimated support for President is higher in list experiment than direct estimates.

Finally, we also estimate the effect of framing on preference falsification. For example, this quantity for the Control vs. Negative framing conditions is $\Delta_{PF} = PF - PF^- = (Direct_{Control} - Indirect_{Control}) - (Direct_{Negative} - Indirect_{Negative})$. To estimate uncertainty about these estimates, we use the formula for a t-test with unequal sizes and similar variances.

Table A.4 reports these quantities. Focusing on the cleaned data, there is evidence—albeit small in magnitude—that the positive frame reduces preference falsification.

Table A.4: Δ_{PF} in support for the president across framing treatments

	Full	Cleaned
Positive	0.00 (-0.04, 0.04)	-0.05 (-0.09, -0.01)
Negative	-0.04 (-0.08, -0.01)	-0.00 (-0.04, 0.04)

Point estimates represent average estimated difference in DB in support for president between control and framing condition, with associated 95% confidence intervals. Positive values indicate that estimated DB is higher in control condition.

A.4 Heterogeneous effects

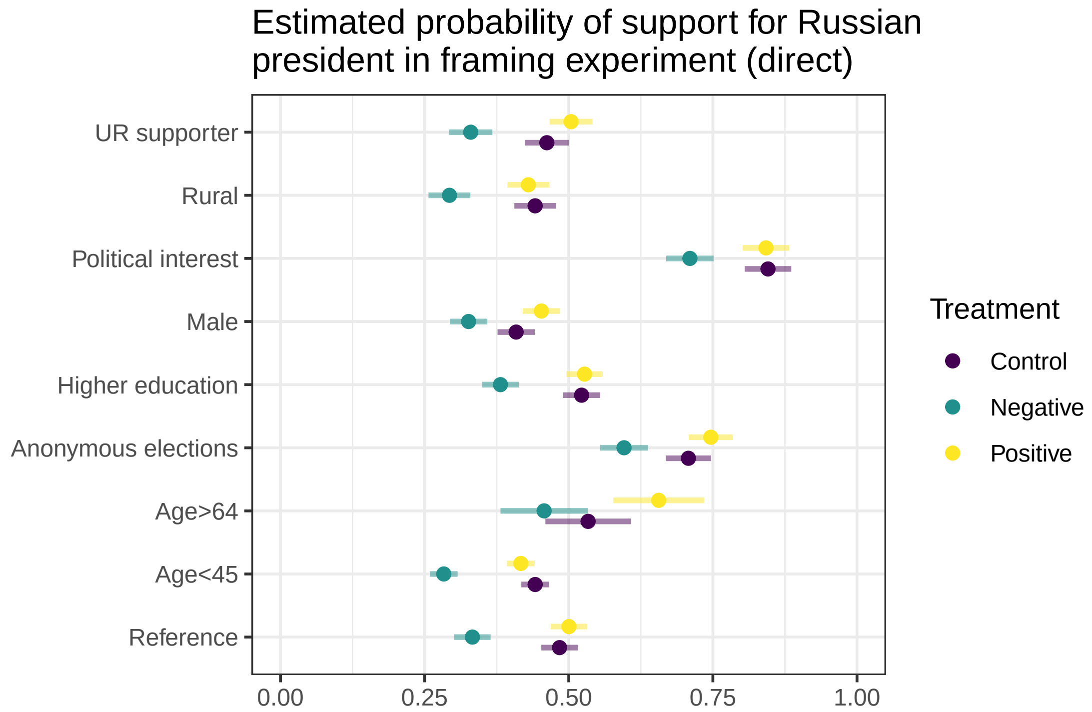
We also analyze heterogeneous treatment effects using potential correlates of preference falsification (Figure A.4) using simple OLS analyses, interacted with the framing conditions in the direct analysis and both the framing and list treatments in the list analyses.

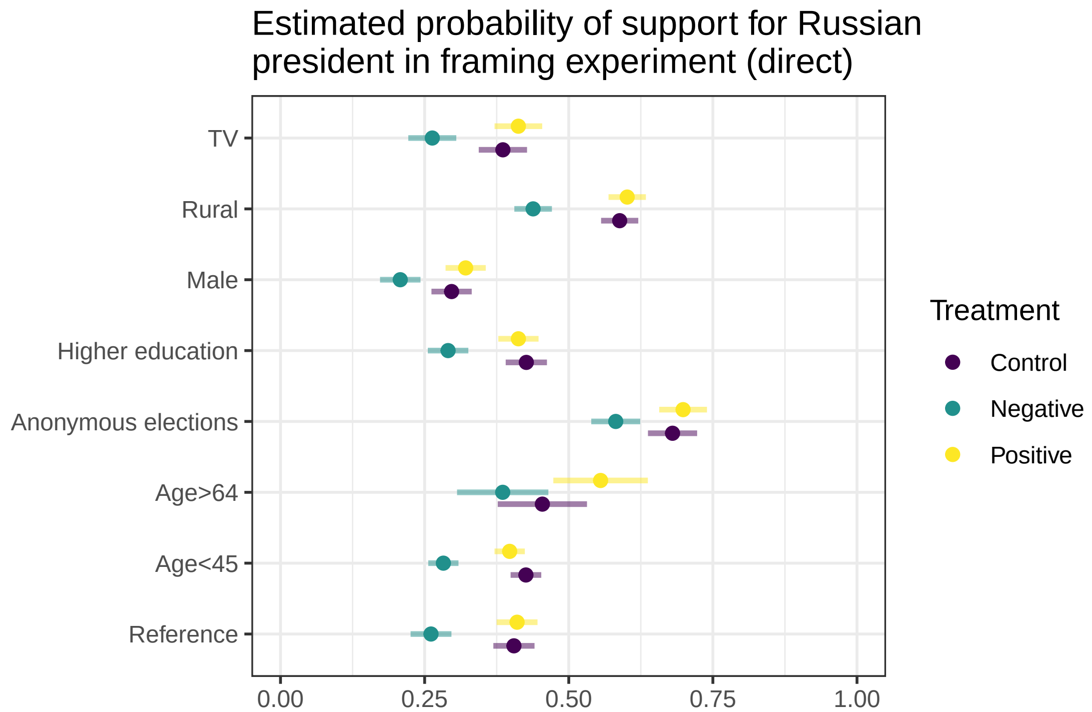
Figure A.4: POADSRR covariates

Age	Two dichotomous indicators for respondents below the age of 45 ("Young") and above the age of 65 ("Old") age quantiles.
Male	Indicator for male respondents.
Higher education	Respondents with higher education. Proxy for political information
Rural	Respondents living in localities with less than 100k respondents.
Anon elections	Indicator for respondents who believe elections in Russia are anonymous (top three categories on seven-point scale). Proxy for perceptions of anonymity.
Pol interest	Indicator for respondents who report being interested in politics (top three categories on seven-point scale). Proxy for political information.
UR supporter	Indicator for respondents who report UR as being the party closest to them from list. Proxy for pro-regime partisanship.
TV watcher	Indicator for respondents who report watching TV at least 2-3 times a week for news. Proxy for both political information and pro-regime partisanship.

In the direct question, A.5, there is minimal evidence of heterogeneous treatment effects: the negative and positive frames largely affect all subgroups equally. There is perhaps more evidence of heterogenous treatment effects in the list experiment (only cleaned data reported), though these results are accompanied by massive uncertainty.

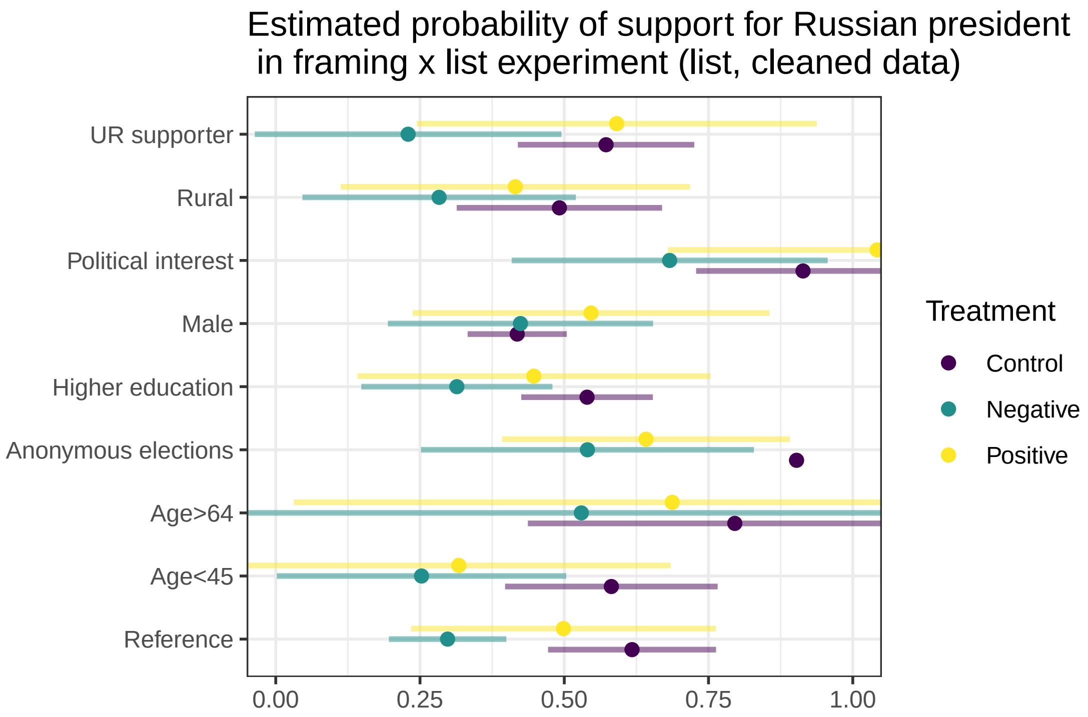
Figure A.5: Heterogenous treatment effects on directly-estimated support for the Russian president



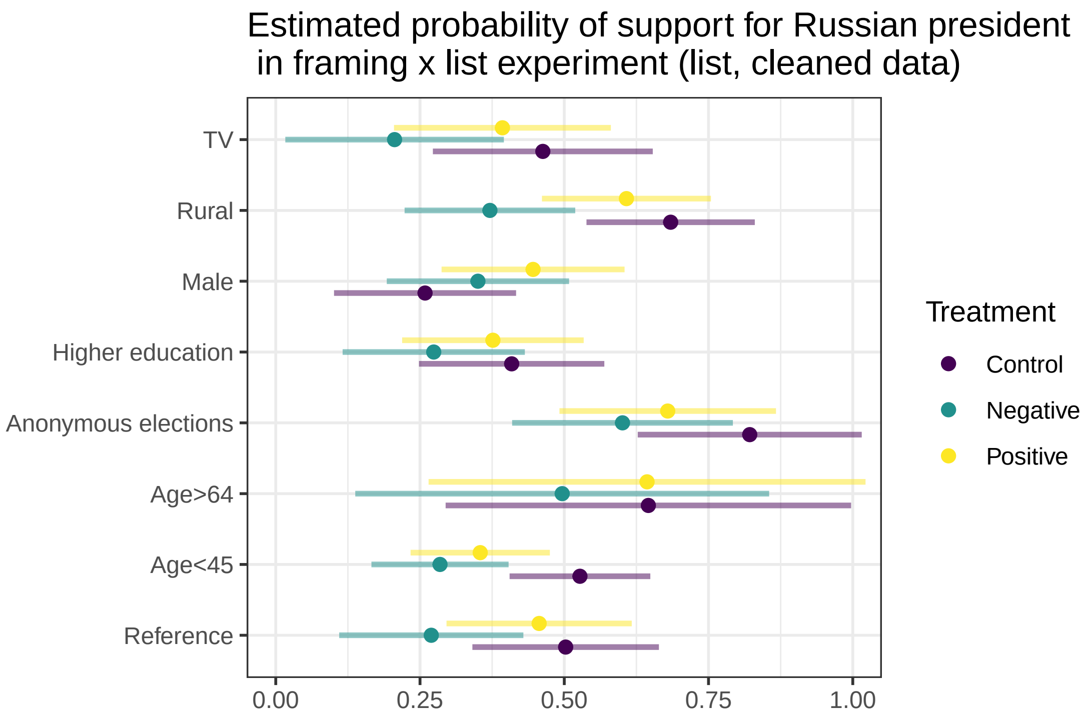


Predicted probabilities from linear probability model interacting covariates with framing experiment conditions. Horizontal lines represent 95% confidence intervals.

Figure A.6: Heterogenous treatment effects on estimated support for the Russian president (cleaned data)



(a) List analysis (cleaned dataset)



(b) List analysis (cleaned dataset)

Predicted probabilities from linear regression interacting covariates with framing experiment conditions × list experiment treatment condition. Horizontal lines represent 95% confidence intervals.

B Ordered probit analyses of framing experiment

Table B.5: Ordered probit analyses of framing experiment

	Levada National Nov 2020	POADSRR National Jun 2021	RES National Sep 2021	POADSRR Regional Aug 2021
Positive frame	0.001	−0.02	0.03	0.002
	(0.07)	(0.07)	(0.07)	(0.02)
Negative frame	-0.13^{**}	-0.11^{*}	−0.07	-0.21^{***}
	(0.07)	(0.07)	(0.07)	(0.02)
Thresholds				
1|2	-0.91^{***}	-0.74^{***}	-1.11^{***}	-0.93^{***}
	(0.05)	(0.05)	(0.06)	(0.02)
2|3	-0.29^{***}	−0.05	-0.36^{***}	-0.10^{***}
	(0.05)	(0.05)	(0.06)	(0.02)
3|4	0.66^{***}	0.99^{***}	1.02^{***}	1.13^{***}
	(0.05)	(0.06)	(0.06)	(0.02)
Observations	1,554	1,503	1,277	14,577

Note: *p$<$0.1; **p$<$0.05; ***p$<$0.01