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

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

Abstract

Being popular makes it easier for dictators to govern. A growing body of scholarship therefore focuses on the factors that influence authoritarian popularity. However, it is possible that the perception of popularity itself affects incumbent approval in autocracies. We use framing experiments embedded in four surveys in Russia to examine this phenomenon. These experiments reveal that manipulating information—and thereby perceptions—about Russian President Vladimir Putin's popularity can significantly affect respondents' support for him. Additional analyses, which rely on a novel combination of framing and list experiments, indicate that these changes in support are not due to preference falsification, but are in fact genuine. This study has implications for research on support for authoritarian leaders and defection cascades in nondemocratic regimes.

*We commend Israel Marques II for facilitating our work with the POADSRR surveys. We also thank Elizabeth Wood and five anonymous reviewers for their thorough comments during the revision process; and Henry Hale, Junyan Jiang and David Szakonyi for their comments on earlier drafts. We would also like to thank audiences at the Milwaukee Area Political Science Seminar (MAPSS), ASEEES 2021, SPSA 2022, and the Comparative Political Economy and Behavior group at UCL. This material is based, in part, upon work supported by the National Science Foundation under award #2049448

†Assistant Professor, Department of Political Science, Trinity College Dublin, buckleno@tcd.ie, ORCID: 0000-0002-4641-6664

‡Associate Professor, Department of Comparative Politics, University of Bergen, kyle.marquardt@uib.no, ORCID: 0000-0003-2043-5880

§Associate Professor, Department of Political Science, University of Wisconsin–Milwaukee, reutero@uwm.edu, ORCID: 0000-0002-2576-3018

¶Associate Professor, University College London, Department of Political Science, k.tertytchnaya@ucl.ac.uk, ORCID: 0000-0002-6216-5893

For dictators, being popular is better than being unpopular. Evidence of regime popularity, such as favorable opinion polls or election victories, can prevent voter and elite defections as well as bolster regime control (Hale and Colton 2017; Tertytchnaya 2020; Reuter and Szakonyi 2019). A growing literature has therefore explored the factors that make authoritarian leaders popular, focusing primarily on the role of ideology (Colton and Hale 2009), performance evaluations (Magaloni 2006; Treisman 2011), and information manipulation in the form of propaganda or censorship (Guriev and Treisman 2019, 2020*a*).

An under-examined question is the extent to which the *perception* of an autocrat's popularity can itself influence their popularity (e.g., Greene and Robertson 2019). Individuals may be more likely to express support for leaders when presented with evidence suggesting that support for the authorities is widespread. Similarly, individuals may be less likely to profess support when such evidence suggests that support for the regime is low or in decline. Such dynamics may reflect sincere preference change or *insincere* change, where respondents' publicly-expressed views and preferences *do not* align with their privately-held beliefs and opinions.[1]

We examine these issues with a framing experiment that presents respondents with information about Russian president Vladimir Putin's standing in opinion polls in the period 2020–2021. Our experiment takes advantage of a unique circumstance: while a majority of Russians expressed support for Putin in surveys during this period, this support had 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 four survey waves in the period 2020–2021—three nationally-representative (two face-to-face and one online) and one subnationally-representative (online)—we find that inducing respondents to consider Putin's ratings as relatively low leads to lower levels of support for him. However, showing respondents a frame that prompts them to consider Putin's approval as relatively high does not influence their support for him.

We furthermore examine whether sincere preference updating or preference falsification drives these changes in support, taking advantage of the large sample size in the subnationally-representative survey to pair our framing experiment with a list experiment. As in the direct questions, we find that the "low popularity" frame reduces estimated support for Putin. These results suggest that some Russians become *genuinely* less supportive of Putin when presented with information that suggests he is unpopular. This evidence of sincere preference change implies that the popularity of autocratic leaders can be endogenous: perceptions of regime support can influence actual support.

[1] Public views are views expressed to strangers, including responses to survey questions (see also Hale and Colton 2017).

The Popularity of Autocrats

Most contemporary autocrats rely on their popularity to ensure social control (Guriev and Treisman 2019). Autocrats can draw popularity from some of the same sources as democratic leaders: citizens may support the leader's programmatic positions or character traits (Colton and Hale 2009; Hale and Colton 2017) or they may believe the autocrat is performing well in office (Magaloni 2006; Treisman 2011). Contemporary authoritarian regimes also try to actively *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 and Treisman 2019, 2020*a*).

Less attention has been paid to how perceptions of a regime's popularity can itself influence support for that regime. Simpser (2013) argues that perceptions of incumbent popularity can persuade potential challengers that it is not worth challenging the regime. In the case of Russia, Greene and Robertson (2019) have suggested that Putin's popularity is, in part, founded on social pressures to conform with the dominant view. Similarly, Hale (2021) shows that the need to conform with a socially acceptable view could account for rally-round-the-flag effects.

This type of conformist behavior may reflect sincere support for the autocrat. As Bicchieri (2005) shows, people may choose to follow the preferences of others because they feel that others' choices are based on information that dominates their own. For example, opinion polls indicating majority support for an incumbent may lead citizens to infer that the leader is competent and trustworthy. Such updating may reflect a conscious consideration if individuals explicitly reason that the leader is more worthy of support simply because others support him.

New information may also lead to sincere preference changes by communicating the dominant, socially desirable view in society (Lohmann 1994; Hale and Colton 2017, p.324). A long line of research shows that many individuals derive pleasure from conforming with the views held those around them Durkheim (1951). By being in harmony with a meaningful reference group—here, the rest of society—individuals can derive positive utility (Edwards 1957; Hale 2021, p.2).

In the political realm, evidence that the ruling regime is popular may encourage some individuals to adopt and report more favorable assessments of the incumbent. However, a similar mechanism could lead to the opposite result: information that regime support is in decline or that opposition to the authorities is becoming socially desirable could lead individuals to (genuinely) adopt less favorable assessments of the regime. In both cases, the updating reflects true preference change.

However, a desire to conform with the majority may also encourage individuals to misreport their true views of the regime—to engage in preference falsification. Individuals could report public views that contrast with their private beliefs because they strive for social approval (Tourangeau and Yan 2007). 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 and Clark 1983; Blair, Coppock and Moor 2020).

Thus, changes in the perception of regime popularity may lead to changes in rates of preference falsification. Reputational cascade models also 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 (Kuran 1991). For example, opinion polls suggesting that opposition to the regime is growing could encourage individuals who previously only *privately* disapproved of the authorities' performance to reveal their sincere preferences now that publicly expressing opposition is seen as common. The opposite could also be true: as politicians become discredited, individuals who privately support them may publicly express opposition (e.g., Kuran 1991; Hale 2021).

The distinction between preference falsification and sincere conformism is more stark in theory than it is in practice. Individuals' publicly expressed beliefs are a balance between social pressures (e.g., the expectation to express certain views about a regime) and personal considerations (e.g., experiences). For many individuals, preference updating is likely to reflect a mix of both sincere and insincere updating. However, the distinction between sincere and insincere opinion change is important because these phenomena have different implications for regime stability.

Autocratic Popularity in Russia

Most observers agree that President Vladimir Putin's popularity is fundamental to the stability of Russia's authoritarian regime (Hale 2014; Greene and 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 largely sincere (Frye et al. 2017; Greene and Robertson 2019).

Although Putin's approval ratings have historically been quite high (above 80% for almost four years following the annexation of Crimea in 2014), they declined dramatically in early 2018 following an unpopular pension reform. Putin's popularity hovered just above 60% through the end of 2021.

Research Design

There is already suggestive evidence that perceptions of Putin's popularity affect support for him. The Levada Center, Russia's most respected polling agency, routinely includes popularity 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, 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."[2]

Unfortunately, such responses cannot form the basis for reliable inferences about how perceptions of regime approval drive Putin's popularity. Respondents who sincerely adhere to social norms about supporting Putin are likely to rationalize their support by identifying concrete reasons that they support Putin. Moreover, respondents may be loathe to admit that they are so easily swayed by the opinion of those around them.

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 identified two instances in which Levada posed this question: March 2015, when respondents were asked about perceptions of Putin's support levels[3] and July 2018, when respondents were asked to estimate Putin's popularity in society.[4] In both cases, support for Putin was strongly associated with believing that Putin was popular. However, respondents may have drawn conclusions about Putin's general popularity based on their own support, or the two factors may be co-determined by unobserved factors.

To exogenously manipulate respondents' beliefs about Putin's popularity, we employ a framing experiment that attempts to shift respondents' perceptions about the popularity of the regime. To our knowledge, this is the first effort 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 levels of support for Putin were objectively high in 2020–21 (over 60%), but still much lower than in recent memory (over 80% following the annexation of Crimea). This makes it possible to frame Putin's poll numbers in both positive and negative light without deceiving respondents. Figure 1 shows the phrasing of the survey experiment.

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).[5] The positive frame notes that this quantity represents a strong and stable majority, while the negative frame notes that *only* that many Russians support Putin and that his approval rating is lower than it has been in recent years.

As previously noted, respondents who update in response to these experimental frames 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

[2] https://www.levada.ru/2016/03/21/vladimir-putin-otnoshenie-i-doverie-2/

[3] http://sophist.hse.ru/dbp/S=2054/Q=14/

[4] http://sophist.hse.ru/dbp/timeser/?S=2122&Q=44

[5] In our November 2020 pilot, 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 response scales were slightly different. Given the broad similarity in results between the pilot and the other three surveys, these differences are unlikely to be consequential.

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, two thirds 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 two thirds 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

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.

In principle, list experiments allow respondents to reveal support for a political figure in aggregate without doing so individually (Imai 2011; Blair and Imai 2012; Glynn 2013; Blair, Coppock and Moor 2020). Respondents are exposed to either a control or treatment list and asked to report the *number* of items pertaining to them. In our application, respondents see either a control list of international political figures or a treatment list with the same figures *and* "the President of Russia" (Figure 2). The lists are identical, save that the treatment list includes the sensitive item (Putin) in addition to the items on the control list. The average difference between control and treatment responses should therefore reflect the overall prevalence of support for Putin. However, since respondents only report a number, not specific items, respondents in the treatment group do not reveal if they support Putin specifically.

In this application, the list experiment should enable us to estimate the degree to which the framing experiment results are due to changes in levels of preference falsification. 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. However, if the frames affect estimates of Putin's support from the direct question—but not the list experiment—it is evidence that the frames are changing levels of preference falsification.

In practice, design effects can limit the validity of list experiments. In the Russian

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**

context, Frye et al. (2023) argue that lists of political figures such as that which we use here could result in artificially deflated estimates of support for Putin. While these concerns imply that we should be cautious in using the results to make claims about Putin's general popularity, they are of minimal relevance to our particular application. Even if design effects affect overall estimated support for Putin, they should be constant across framing experiment conditions. As a result, the treatment effect estimates from our framing experiment should not be subject to list design effects.

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) surveys are nationally and subnationally 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 surveys were pilots for the RES nationally- and POADSRR subnationally-representative surveys.[6] Since the changes between the pilots and pre-registered surveys were minimal we report the results together.[7] All surveys included the framing experiment, while the POADSRR surveys also included the framing $\times$ list experiment. Since the nationally-representative POADSRR survey was severely underpowered for this framework, we only report framing and list results from the subnationally-representative survey.

Using multiple survey firms and modes helps ensure that results are not driven by a specific firm or mode, alleviating concerns about experimenter demand effects.[8]

[6] Survey details can be found in Appendix A.

[7] Pre-registration available at `osf.io/8fj2q/?view_only=cfaf91f9e03043ac9b17d1863728efb8`.

[8] Experimenter demand effects likely vary across mode. For example, online experiments minimize experimenter-participant interaction and thereby (perceived) social pressure from the experimenter.

Models

To estimate the direct effect of the negative and positive frames on support for President Putin, we dichotomize the 4-point 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:[9]

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

To estimate framing effects in the list experiment, we use standard linear regression.[10] 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:

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

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 control list parameters, which are not of substantive interest.

Results

Table 1 reports the results from these analyses, which are remarkably consistent across survey waves.[11] Columns 1–4 show the direct effect of the two experimental frames on support for Putin, while Column 5 estimates framing effects 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 list experiment control list (β_1–β_3).

[9] We use dichotomized outcomes so that the results are comparable to those in the framing $\times$ list experiment. We also analyze the data using ordered probit models and investigate the effects of the framing experiment on the outcome distribution in Appendix C.

[10] We implement a pre-registered algorithm to clean the list experiment data (Appendix A.3.1). List results are robust to maximum-likelihood estimators which Imai (2011) and Blair and Imai (2012) propose (Appendix A.4.2).

[11] Appendix B provides balance checks on experimental treatments and Appendix A.4.1 shows list experiment diagnostics. Results are robust to clustering standard errors by region and including demographics (Appendix C.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	16,329	14,582
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.

In all survey waves, the positive frame shows no statistically significant effect. In contrast, the negative frame shows a consistently significant and substantively strong effect across direct responses: a 6–11 percentage point decrease in estimated support. Respondents who received information that Putin's popularity was subpar were significantly less likely to report support for Putin than those in the control condition. These treatment effects 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 suggests that results from the framing experiment are attributable to sincere changes in preferences, and not to changes in the extent of preference falsification.[12] When respondents are exposed to negative information about Putin's popularity, a substantial proportion sincerely revise their support for him downward.

The larger impact of the negative frame may be due to the fact that it provides more new information to respondents. If most respondents already believe that Putin's popularity is high and stable—in line with the positive frame—the effect of this frame would be biased toward zero.[13] It is also possible that respondents pay more attention to negative news (Trussler and Soroka 2014).[14]

Appendix A.3.5 presents results from analyses of heterogenous treatment effects. The main demographic trait that appears to moderate treatment effects is age: the positive frame increases support for Putin among older respondents, while the negative frame appears to have a weaker effect among this group relative to other age categories.

Conclusion

Autocrats in the 21st century are attuned to their image.[15] In place of overt repression, they manipulate the informational environment to convince the masses that they are popular (Guriev and Treisman 2020*b*, 2019). Here we examine one reason why this manipulation may be particularly important: perceptions of incumbent popularity might themselves inflate incumbents' approval levels. 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

[12]Appendix A.3.4 provides additional analyses of preference falsification across experimental conditions.

[13]Survey evidence suggests that this explanation is plausible: in December 2021 around 42% of Russians believed that the Russian president enjoys the support of a majority of citizens.

[14]The asymmetry of framing effects is not consistent with experimenter demand effects. If both the positive and negative frames help respondents infer the purpose of an experiment and thereby encourage them to adjust their behavior, then both should lead to attitudinal updating. Moreover, Russia's authoritarian nature should make such compliance with the positive frame more likely than with the negative frame.

[15]The Kremlin's obsession with monitoring and promoting its own opinion ratings has even been termed 'ratingocracy' (Hale 2010).

of public opinion in Russia. We find that a frame revealing relatively low support for Putin makes respondents less likely to report support for him. A combined framing and list experiment indicates that the results from the framing experiment are, in fact, due to sincere updating of preferences.[16] These findings demonstrate that perceptions of Putin's popularity can influence his actual level of support.

These results imply that shaping perceptions—through propaganda, indoctrination, schools, and the media—is an important element of authoritarian popularity and thus stability. While conformist impulses likely shape support for politicians in democracies as well, this phenomenon is of particular importance in autocratic settings, where incumbents have an outsized ability to shape both their own popularity and perceptions of their popularity. Many contemporary autocrats have high approval ratings when compared to their democratic counterparts; our research demonstrates how this popularity can be self-sustaining, even in the absence of significant preference falsification.

At the same time, endogenous popularity can be fragile. Indeed, our results show that relatively mild negative information can reduce support for an autocrat by 6–11 percentage points. This fragility has important implications for regime stability. When social consensus breaks down, regimes can dissolve rapidly. Such cascades are likely to be even more abrupt when consensus rests on perceptions, as opposed to being manufactured through intimidation, normative congruence, or ideology. Individuals who support the authorities because they think that the authorities are popular may be quick to withdraw support when they think that others around them have begun to do the same.

[16] While our research design cannot determine the precise psychological mechanism that underlines updating, this finding provides a necessary basis for such research in the future.

The authors affirm that this article adheres to the APSA's Principles and Guidance on Human Subject Research. The authors declare the human subjects research in this article was reviewed and approved by the Institutional Review Board at the University of Wisconsin-Milwaukee and The George Washington University Office of Human Research. Certificate numbers are provided in the appendix. The authors declare no ethical issues or conflicts of interest in this research. Research documentation and data that support the findings of this study are openly available in the APSR Dataverse at `https://doi.org/10.7910/DVN/NVXQOG`.

References

Bicchieri, Cristina. 2005. *The Grammar of Society: The Nature and Dynamics of Social Norms.* Cambridge University Press.

Blair, Graeme, Alexander Coppock and 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.

Blair, Graeme and Kosuke Imai. 2012. "Statistical analysis of list experiments." *Political Analysis* 20(1):47–77.

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

Durkheim, Emile. 1951. "Suicide: a study in sociology [1897]." *Translated by JA Spaulding and G. Simpson (Glencoe, Illinois: The Free Press, 1951)* .

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

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

Frye, Timothy, Scott Gehlbach, Kyle L. Marquardt and Ora John Reuter. 2023. "Is Putin's Popularity Still Real? A Cautionary Note on Using List Experiments to Measure Popularity in Authoritarian Regimes." *Post-Soviet Affairs* Early View:1–10.

Glynn, Adam N. 2013. "What can we learn with statistical truth serum? Design and analysis of the list experiment." *Public Opinion Quarterly* 77(S1):159–172.

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

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

Guriev, Sergei and Daniel Treisman. 2020*a*. "The Popularity of Authoritarian Leaders: A Cross-National Investigation." *World Politics* 72(4):601–638.

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

Hale, Henry E. 2010. "Eurasian polities as hybrid regimes: The case of Putin's Russia." *Journal of Eurasian studies* 1(1):33–41.

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. and 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.

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

Kuran, Timur. 1991. "Now Out of Never: The Element of Surprise in the East European Revolution of 1989." *World Politics* 44(1):7–48.

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 and 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: Cambridge University Press.

Reuter, Ora John and David Szakonyi. 2019. "Elite defection under autocracy: Evidence from Russia." *American Political Science Review* 113(2):552–568.

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 and 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.

Trussler, Marc and Stuart Soroka. 2014. "Consumer demand for cynical and negative news frames." *The International Journal of Press/Politics* 19(3):360–379.

Figure 1

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, two thirds 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 two thirds 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

Figure 2

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**

Endogenous Popularity Supplementary Appendix

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

Contents

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

A Details on survey data

A.1 Survey design and implementation

We placed the framing experiment in two face-to-face surveys. In November 2020, we included the framing experiment in a monthly Omnibus Survey carried out by the Levada Center. This nationally representative survey probabilistically sampled 1607 adults in 140 settlements in Russia. Interviews were conducted face-to-face. Informed consent was achieved when interviewers read the consent text at the beginning of the interview and requested an interview.

We also included the framing experiment in the first round of the 2021 RES, which was conducted August-September 2021. This nationally representative survey probabilistically sampled 2,677 adults in Russia and interviews were conducted face-to-face. Informed consent was achieved when interviewers read the consent text at the beginning of the interview and requested an interview.

We included the framing $\times$ list experiment (and associated list experiment diagnostic questions) in both the nationally- and regionally-representative POADSRR surveys. The POADSRR surveys were administered by the Faculty of Social Sciences of HSE University, using a sampling frame from a well-respected Russian firm that requested anonymity due to the potential political sensitivity of the questions.

The nationally-representative survey sampled approximately 1,500 respondents and the regionally-representative survey sampled approximately 16,250 respondents. Both sampled respondents in 60 regions using quotas, with a maximum for each region and quotas set for age, gender and education.

The recruitment company randomly selected respondents from their frame and emailed them a personalized link. Respondents who follow the link were directed to an HSE University server, where they are presented with informed consent text. Respondents who affirmed their consent are allowed to participate. Respondents who completed the survey received compensation between 50 and 100 Russian rubles (roughly 0.65 to 1.30 USD).

A.2 Human Subjects Research

The surveys for this article were approved by the Institutional Review Board at University of Wisconsin-Milwaukee [Approval Certificates 22.012 and 21.130] and the George Washington University Office of Human Research [IRB no. NCR213582]

A.3 The POADSRR subnationally-representative survey

Given the large sample size of the POADSRR subnationally-representative survey ($N = 16,342$), 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.[1]

A.3.1 List experiment cleaning algorithm

Analyses of the POADSRR nationally-representative (pilot) survey indicated that a substantial proportion of respondents in the online setting nonsensically inflate their

[1] Preregistration: `osf.io/8fj2q/?view_only=cfaf91f9e03043ac9b17d1863728efb8`

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.[2] 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 subnationally-representative survey. Specifically, we clean the dataset such that respondents in the control group can only report ± 1 the number of figures they directly report supporting in the control list, while respondents in the treatment group can only report only one fewer figure 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 respondents who report two more figures in the treatment list than they do directly are doing so in error, not because they support the president). On the other hand, this approach might underestimate support because it removes 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 cleaned and the full dataset for the sake of robustness. 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.3.2 Cleaning algorithm 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.

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 an 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

[2]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*

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

Figure A.1: 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.

A.2, A.3 and A.4 report the predictors of being a list falsifier, both by framing effects and with heterogenous treatment effects (description of covariates in Figure A.1). 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.

A.3.3 Analyses of direct and indirect treatment effects

In the appendix our baseline analyses are the same as in the text. We estimate the direct effects of the framing experiment using Equation 1 in the text, and their indirect effects with the list experiment using Equation 2 in the text. To briefly reiterate, we use a linear probability model to regress dichotomized directly-reported support for Putin 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 using the list experiment, 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)

Table A.2: Demographic and experimental (framing and list) correlates of probability of being a list falsifier

Constant	0.08*** (0.01)	0.11*** (0.01)	0.12*** (0.01)
Positive Frame	−0.01 (0.01)		
Negative Frame	0.003 (0.01)		
List Treatment	0.05*** (0.01)		
Positive Frame × List Treatment	0.02* (0.01)		
Negative Frame × List Treatment	−0.01 (0.01)		
Anonymous elections		−0.02*** (0.01)	−0.002 (0.01)
Rural		0.005 (0.01)	0.01 (0.01)
Political interest		−0.02*** (0.01)	
UR supporter		0.07*** (0.01)	
TV			−0.01* (0.005)
Age<45		0.02*** (0.01)	0.03*** (0.01)
Age>64		0.02 (0.02)	0.02 (0.02)
Male		−0.0003 (0.01)	−0.01** (0.01)
Higher education		−0.04*** (0.005)	−0.05*** (0.005)
Observations	16,334	16,341	16,341
R^2	0.01	0.02	0.01

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

Analyses use linear probability model with dichotomous indicator of being a list falsifier as the outcome.

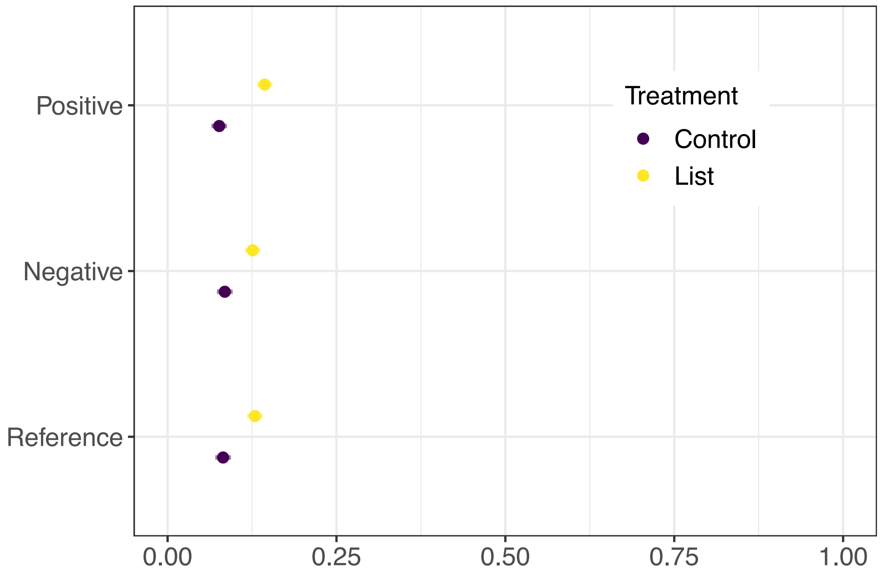
Table A.3: Heterogeneous framing effects on probability of being a list falsifier

Constant	0.14^{***} (0.02)	0.15^{***} (0.02)
Anonymous elections	−0.001 (0.02)	0.02 (0.02)
Rural	0.01 (0.02)	0.02 (0.02)
Political interest	-0.02^{*} (0.01)	
UR supporter	0.11^{***} (0.02)	
TV		−0.005 (0.01)
Age<45	0.01 (0.01)	0.01 (0.01)
Age>64	0.05 (0.04)	0.04 (0.04)
Male	−0.002 (0.01)	−0.02 (0.01)
Higher education	-0.07^{***} (0.01)	-0.07^{***} (0.01)
Positive Frame	0.03 (0.02)	0.03 (0.03)
Negative Frame	−0.02 (0.02)	−0.03 (0.03)
Positive Frame interactions		
Anonymous elections	−0.03 (0.02)	−0.03 (0.02)
Rural	−0.01 (0.02)	−0.01 (0.02)
Political interest	−0.01 (0.02)	
UR supporter	−0.02 (0.02)	
TV		−0.02 (0.02)
Age<45	0.01 (0.02)	0.01 (0.02)
Age>64	0.003 (0.06)	0.01 (0.06)
Male	−0.0001 (0.02)	0.0002 (0.02)
Higher education	−0.01 (0.02)	−0.01 (0.02)
Negative Frame interactions		
Anonymous elections	−0.03 (0.02)	−0.02 (0.02)
Rural	−0.01 (0.02)	−0.01 (0.02)
Political interest	0.01 (0.02)	
UR supporter	0.02 (0.02)	
TV		0.02 (0.02)
Age<45	0.003 (0.02)	0.01 (0.02)
Age>64	−0.004 (0.06)	−0.0001 (0.06)
Male	0.02 (0.02)	0.01 (0.02)
Higher education	0.02 (0.02)	0.02 (0.02)
Observations	8,180	8,180
R^2	0.03	0.01

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

Analyses use linear probability model, and are restricted to list treatment condition for ease of interpretation. A dichotomous indicator of being a list falsifier is the outcome.

Figure A.2: Estimated probability of being list falsifier by framing condition



Note: Analyses show predicted probabilities from linear probability model. All values held constant at zero except for specified indicator or indicators (in case of interactions). Horizontal lines represent 95% confidence intervals about predicted probabilities. Full model specification in Table A.2, column 1.

indicators of the framing treatments, and 3) the interaction of the experimental treatments:

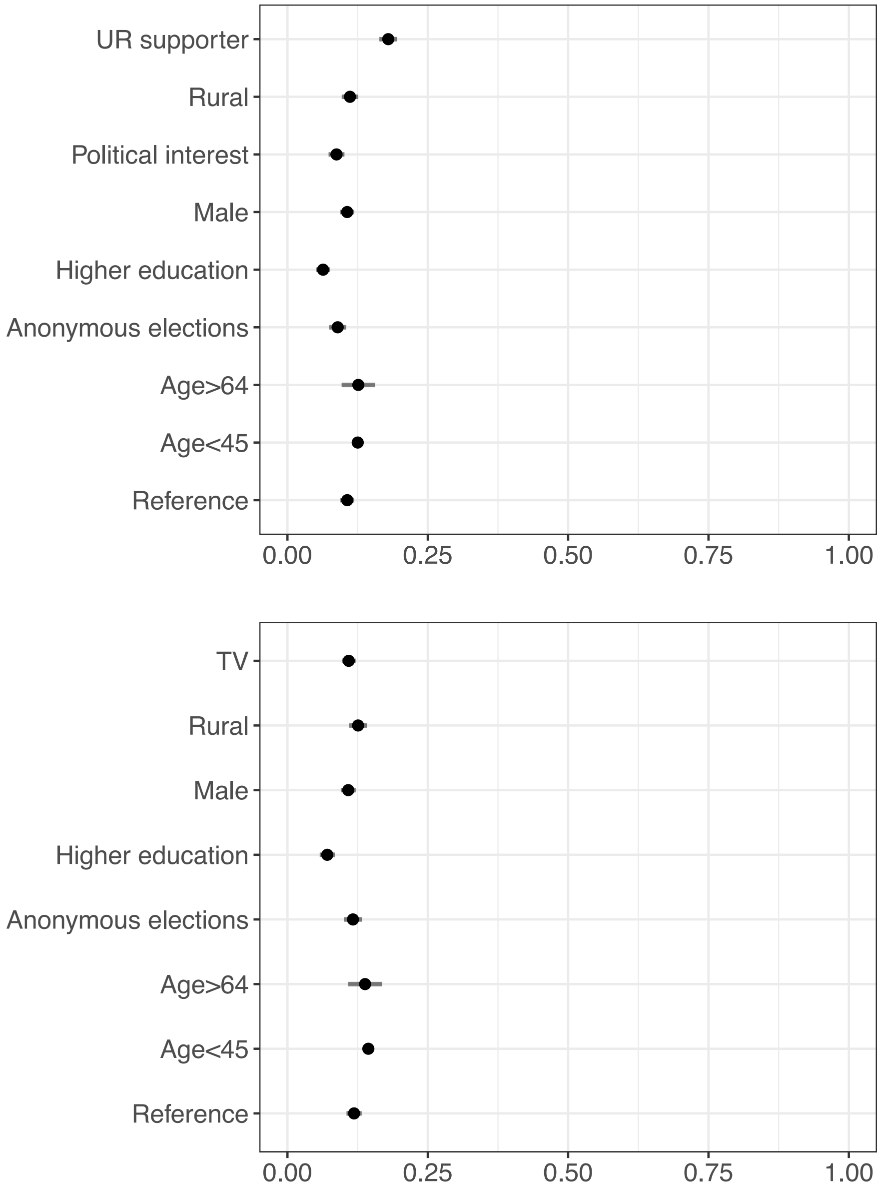
$$y_i = \beta_1 + \beta_2 Negative_i + \beta_3 Positive_i + \alpha_1 List_i + \alpha_2 List_i \times Negative_i + \alpha_3 List_i \times Positive_i + \epsilon_i \tag{A.2}$$

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.

Table A.4 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.

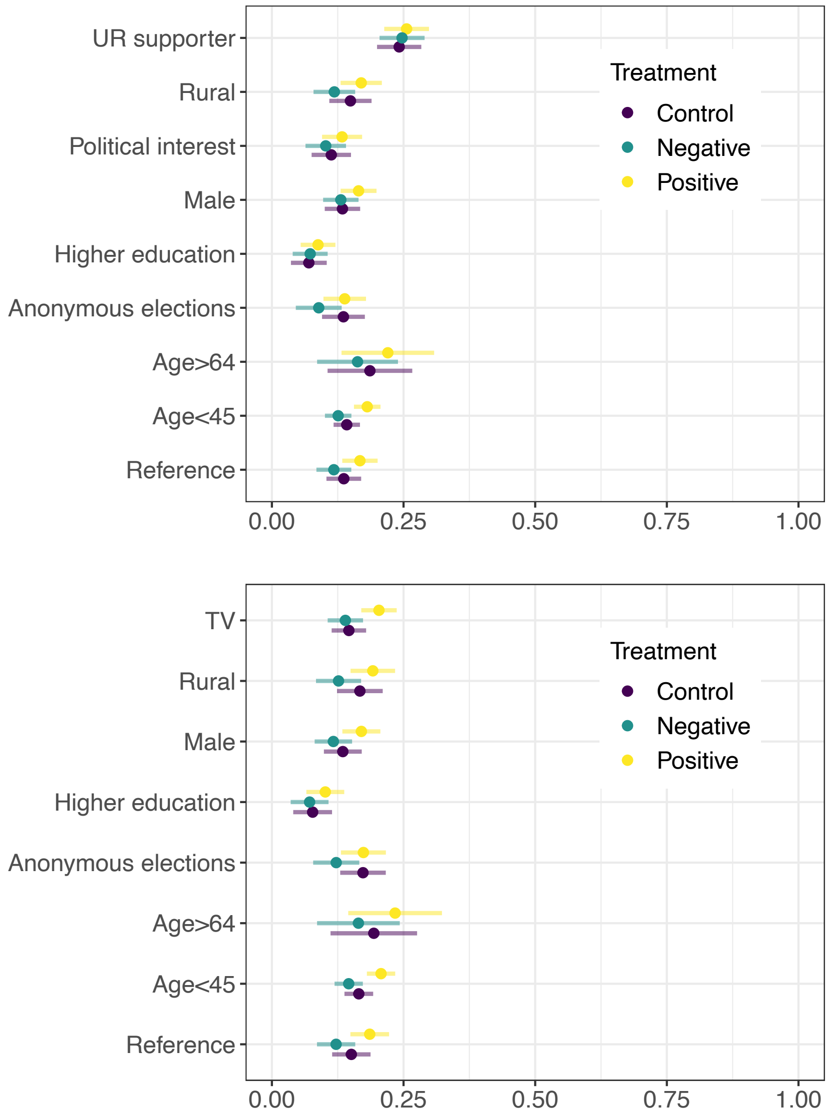
The statistically significant effect of the negative frame in direct experiment is evidence that the frame makes respondents less likely to report support for Putin; the fact that the effect is similar (significant and negative) in both sets of list experiment data is strong evidence that this result is not due to preference falsification. It is also worth noting that the magnitude of the negative frame's effect is similar in the full list data, indicating that the result is not a relic of the data cleaning. The constant (control) condition in the full list indicates substantial preference falsification in support for Putin in that the estimate

Figure A.3: Estimated probability of being list falsifier by demographic correlates



Note: Analyses show predicted probabilities from linear probability model. All values held constant at zero except for specified indicator or indicators (in case of interactions). Horizontal lines represent 95% confidence intervals about predicted probabilities. Full model specifications in Table A.2, columns 2 (top cell) and 3 (bottom cell).

Figure A.4: Estimated probability of being list falsifier, by framing condition and with heterogeneous treatment effects



Note: Analyses show predicted probabilities from linear probability model. All values held constant at zero except for specified indicator or indicators (in case of interactions). Horizontal lines represent 95% confidence intervals about predicted probabilities. Full model specifications in Table A.3, columns 1 (top cell) and 2 (bottom cell).

Table A.4: Estimated support for President across experimental conditions

	Direct (LPM)	List (OLS)	
		Cleaned	Full
		Support for President	
Constant	0.56*** (0.01)	0.56*** (0.03)	0.72*** (0.03)
Positive Frame	−0.002 (0.01)	−0.05 (0.04)	−0.004 (0.04)
Negative Frame	−0.11*** (0.01)	−0.12*** (0.04)	−0.13*** (0.04)
		Control items	
Constant		1.00*** (0.02)	1.05*** (0.02)
Positive Frame		0.02 (0.03)	0.01 (0.03)
Negative Frame		0.01 (0.03)	0.01 (0.03)
Observations	16,329	14,582	16,329
R^2	0.01	0.06	0.08
Note:			*p<0.1; **p<0.05; ***p<0.01

of support is substantially higher in these data; however, this result is likely due to list falsifiers.

A.3.4 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 $\alpha_1 + \alpha_2$ (from Equation A.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 sample size; in the cleaned dataset we only use data from respondents who are not list falsifiers).

Table A.5 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.

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.6 reports these quantities. Focusing on the cleaned data, there is evidence—

Table A.5: Estimated levels of preference falsification and design effects in support for president, across experimental conditions

	Full	Cleaned
Control	-0.15 (-0.17, -0.12)	-0.01 (-0.04, 0.02)
Positive	-0.15 (-0.18, -0.13)	0.04 (0.01, 0.07)
Negative	-0.12 (-0.15, -0.10)	-0.01 (-0.04, 0.02)

Note: 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 the President is higher in list experiment than direct estimates. Refer to the first paragraph of A.3.4 for a description of the estimation strategy, which uses the estimates from Table A.4 to simulate probabilities of support for the President in different experimental frames.

Table A.6: Δ_{PF} in support for the president across framing treatments

	Full	Cleaned
Positive	0.01 (-0.03, 0.04)	-0.05 (-0.09, -0.01)
Negative	-0.02 (-0.06, 0.01)	0.00 (-0.04, 0.04)

Note: Point estimates represent average estimated difference in preference falsification in support for president between control and framing condition, with associated 95% confidence intervals. Positive values indicate that estimated preference falsification is higher in control condition. Refer to the penultimate paragraph of A.3.4 for a description of the estimation strategy, which uses the estimates from Table A.5.

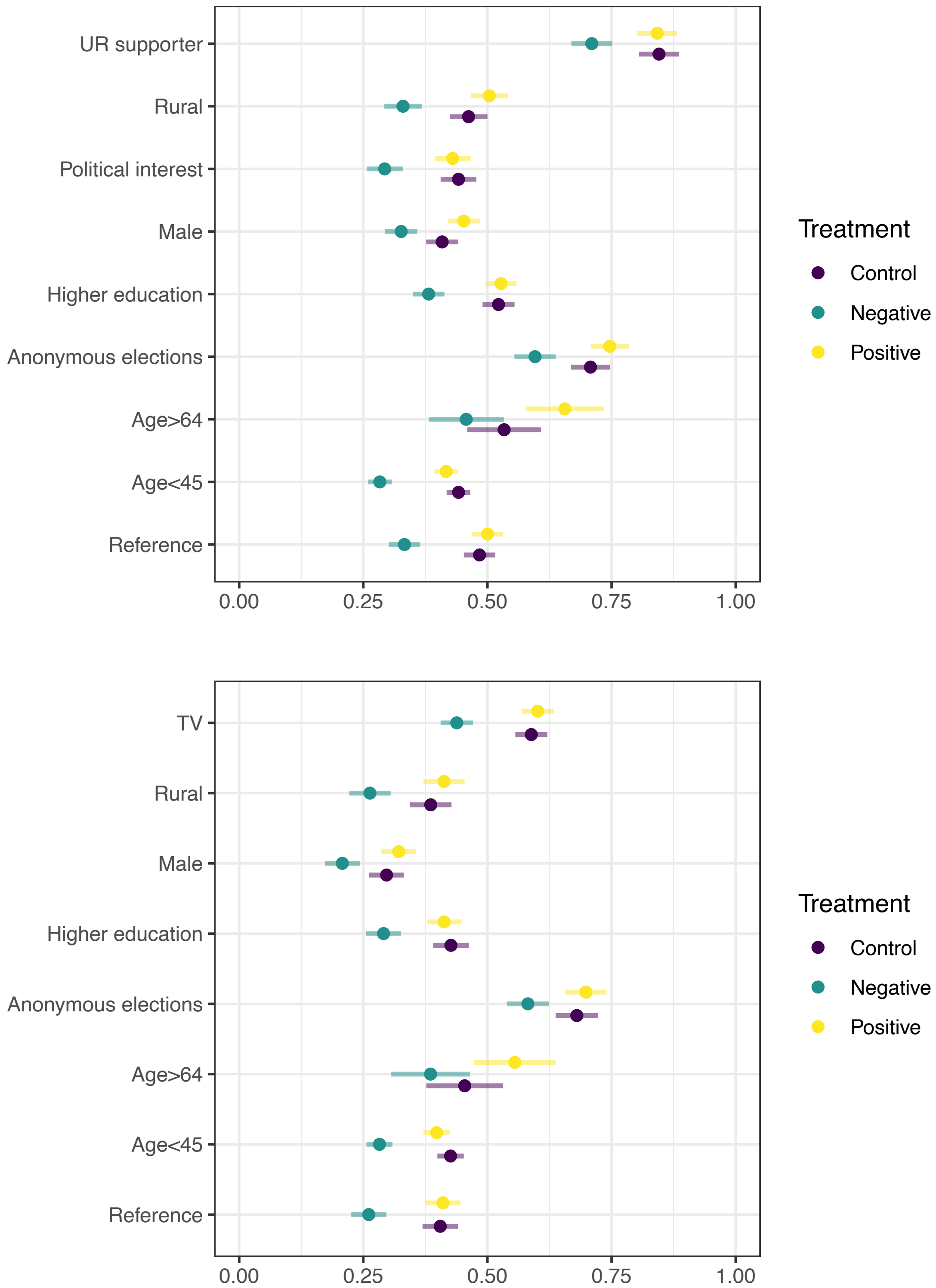
albeit small in magnitude—that the positive frame reduces preference falsification, though this may be a relic of the cleaning procedure.

A.3.5 Heterogeneous effects

We also analyze heterogeneous treatment effects using potential correlates of preference falsification (Figure A.1) using simple OLS analyses, interacted with the framing conditions in the direct analysis and both the framing and list treatments in the list analyses.

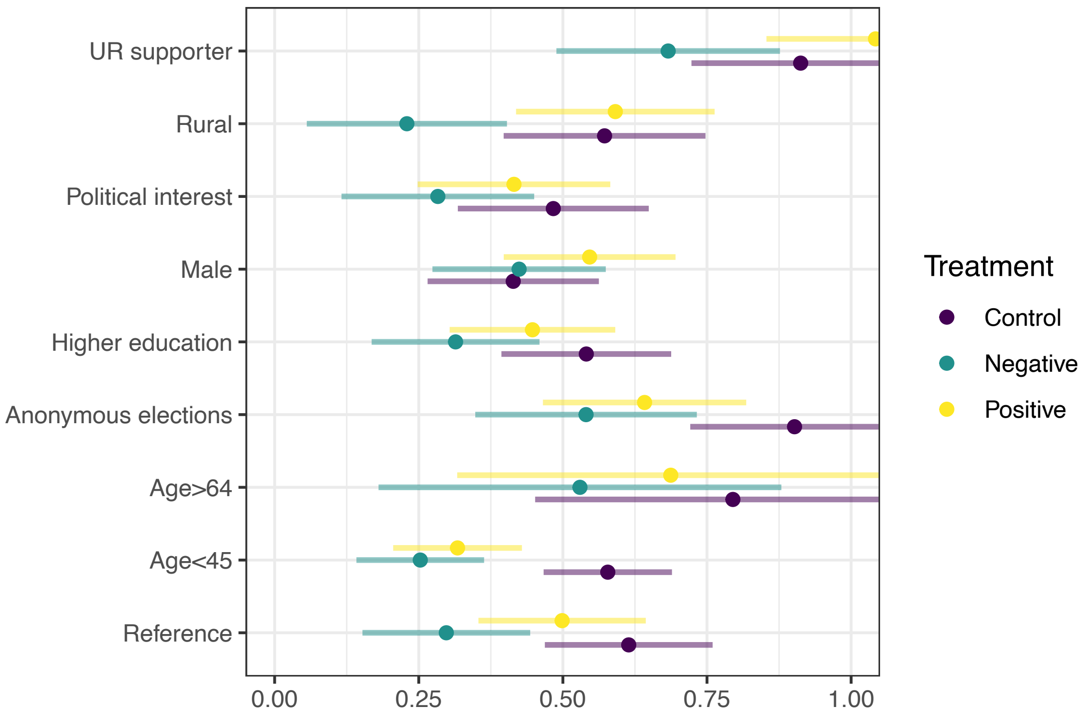
In the direct question, Figure A.5, there is minimal evidence of heterogeneous treatment effects: the negative and positive frames largely affect all subgroups equally. There is perhaps more evidence of heterogenous treatment effects in the list experiment (only cleaned data reported), although, as these results are accompanied by the substantial uncertainty associated with list experiment designs, we refrain from drawing substantive conclusions from these analyses.

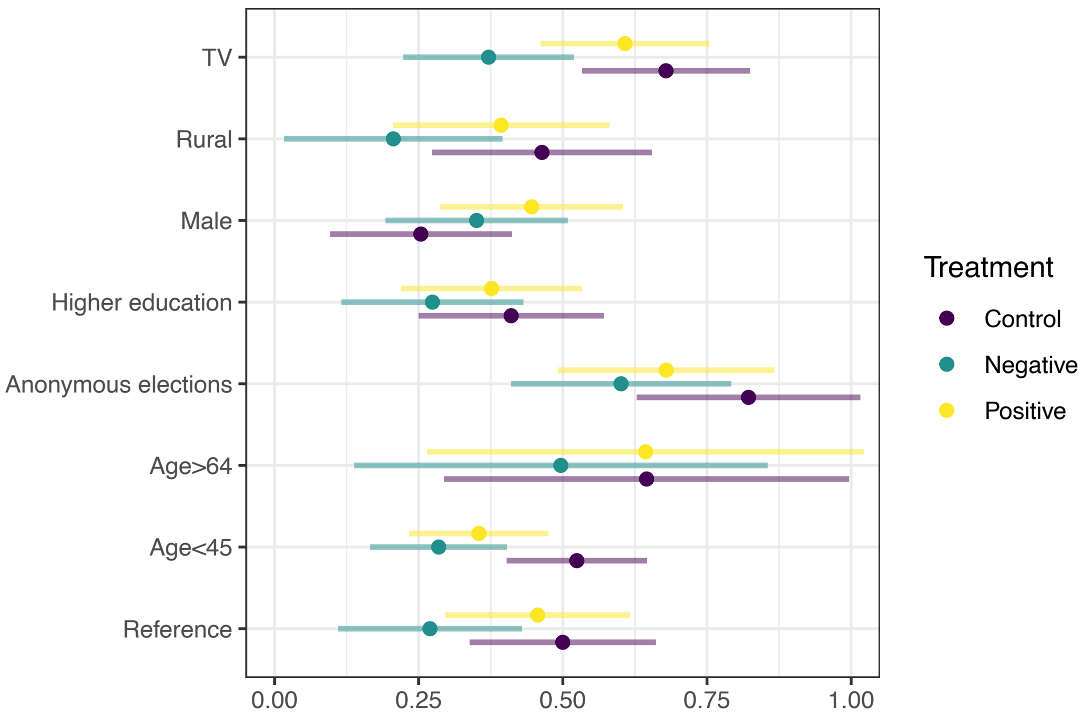
Figure A.5: Heterogenous treatment effects on directly-estimated support for the Russian president



Note: Predicted probabilities from linear probability model interacting covariates with framing experiment conditions. All values held constant at zero except for specified indicator or indicators (in case of interactions). Horizontal lines represent 95% confidence intervals. Full model specifications in Supplementary Table S.1, columns 1 (top cell) and 2 (bottom cell).

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





Note: Predicted probabilities from linear regression interacting covariates with framing experiment conditions × list experiment treatment condition. All values held constant at zero except for specified indicator or indicators (in case of interactions). Horizontal lines represent 95% confidence intervals. Full model specifications in Supplementary Table S.2, columns 1 (top cell) and 2 (bottom cell).

Table A.7: Blair and Imai (2012) design effect test Bonferroni-corrected p-values

	Full Dataset	Cleaned dataset			
	All Treatments	All Treatments	Control	Negative	Positive
P-value	**0.00**	0.57	0.08	0.85	1.00
N	16,329	14,582	4,852	4,860	4,870

Note: We reject the null hypothesis of no design effects for p-values below $\alpha = .05$ (highlighted in bold).

A.4 Additional list experiment analyses

A.4.1 Additional diagnostics

We conduct two sets of diagnostics of our list experiment in addition to those discussed in relation to the cleaning algorithm. First, we analyze our list experiments using the Blair and Imai (2012) test for design effects (Table A.7). While we can reject the null hypothesis of no design effects for the full dataset (first column), after applying the pre-registered cleaning algorithm the tests (both of the overall experiment and specific framing conditions) do not provide strong evidence to reject the null hypothesis.

Second, we graphically analyze the relationship between the control items and both the sensitive item (support for Putin) and the list responses. Figure A.7 illustrates the relationship between direct support for control list items (i.e., the sum of heads of government whom a respondent reports supporting in direct questions) and directly-stated support for Putin on the four-point response scale, divided by framing experiment condition. The graphic indicates that there is a positive but substantively not very strong correlation between the number of control list figures for whom a respondent directly voices support and support for Putin. The figure further indicates that this relationship is similar across framing experiment conditions, though the intercept for the negative framing condition is lower than the control and positive frame since it reduces overall support for Putin.

Figure A.8 illustrates the relationship between the number of control items respondents reported supporting directly and their list responses. Lines represent linear regression estimates of this relationship; yellow represents respondents in the list treatment condition and purple those in the list control condition. Quadrants represent different framing experiment conditions, with the upper left representing the overall relationship across all framing conditions.

In the absence of design effects or ceiling/floor effects, we would expect the yellow and purple lines to a) show a strong positive correlation between list and direct responses and b) run parallel to each other. A strong positive correlation would indicate that, across list treatment conditions, the number of control list figures a respondent reports supporting directly correlates with the number of respondents they report supporting on the list. Parallel lines indicate that the proportion of respondents who report an additional item in the treatment condition (i.e., the proportion who supports Putin) is consistent regardless of control list items. If there are floor effects, we would expect a relatively high proportion of respondents in the treatment condition to report supporting the sensitive figure, resulting in a more negative slope in the purple line. If the addition of the sensitive item to the list in the treatment condition changes evaluation of the control list items,

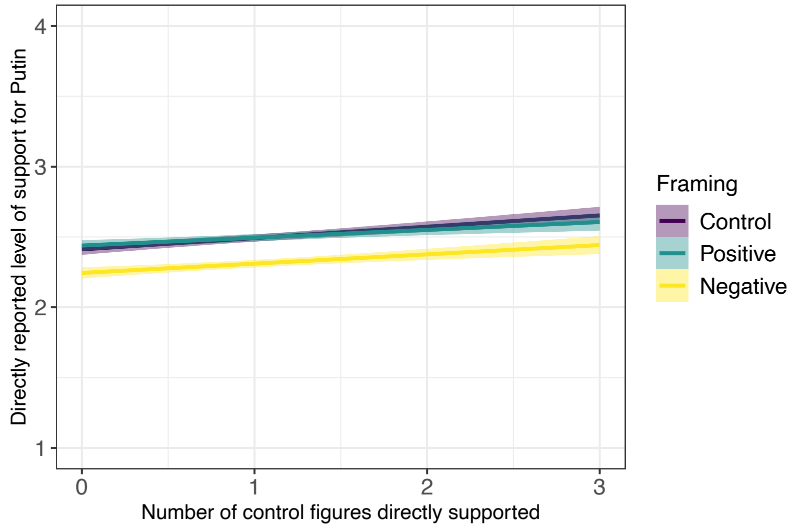
Table A.8: Relationship between number of control items supported and support for the President in ordinal-scale direct question and list experiment

	Direct	List experiment			
	All	All	Control	Negative	Positive
Constant	2.53***	1.46***	1.45***	1.47***	1.47***
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Control Items	0.08***	0.88***	0.88***	0.88***	0.89***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Positive Frame	−0.01				
	(0.02)				
Positive Frame × Control Items	−0.02				
	(0.02)				
Negative Frame	−0.19***				
	(0.02)				
Negative Frame × Control Items	−0.01				
	(0.02)				
List Treatment		0.48***	0.53***	0.42***	0.50***
		(0.01)	(0.02)	(0.02)	(0.02)
List Treatment × Control Items		0.04***	0.07***	0.05**	0.02
		(0.01)	(0.02)	(0.02)	(0.02)
Observations	14,577	14,582	4,852	4,860	4,870
R^2	0.01	0.68	0.69	0.67	0.69

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

All analyses use linear regression. Control items centered at zero.

Figure A.7: Relationship between direct responses to control list items and 4-pt support for Putin (cleaned data)



Note: Shaded areas represent 95% confidence intervals about linear regression estimates. Model specification in Table A.8, column 1.

the slope of the treatment condition should be different from the control condition.

Across framing experiment conditions, the yellow and purple lines run roughly parallel to each other and show a strong positive correlation between the list responses and the control direct responses. These analyses therefore provide no evidence of design effect issues in the list experiment. Note also that the main difference across framing experiment conditions is that the distance between the yellow and purple lines is the least in the negative framing condition, illustrating that fewer respondents support Putin in that condition.

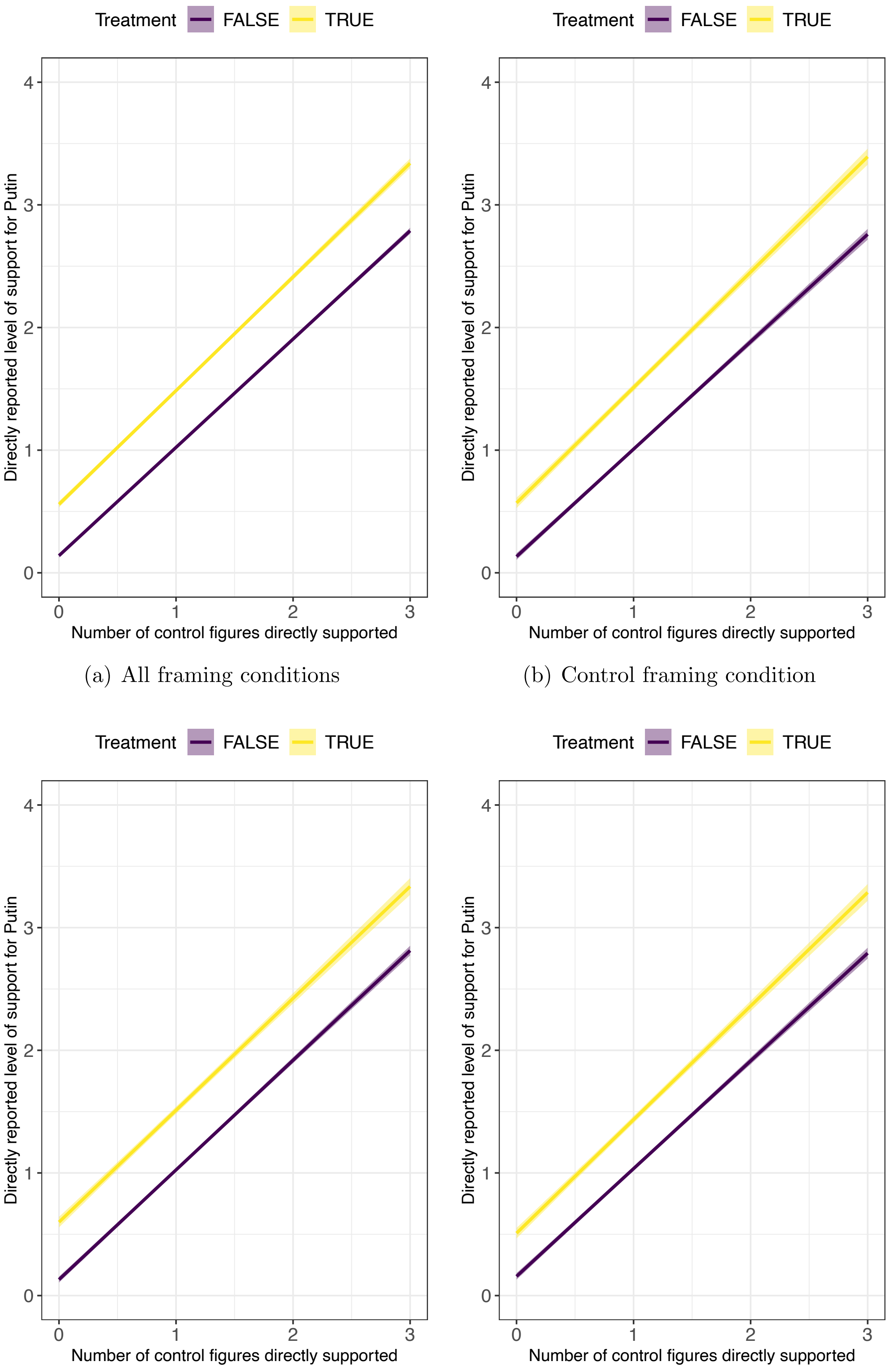
A.4.2 Maximum likelihood models

We also analyze framing effects in our list experiment using three maximum likelihood (ML) algorithms from Imai (2011) and Blair and Imai (2012). First, we use their standard ML algorithm, which can increase statistical efficiency. Second, we use the algorithm that corrects for floor effects, a plausible concern in our context: if a lack of support for Putin is sensitive, then respondents in the treatment condition who do not support any figures in the list may feel compelled to report supporting at least one figure. Third, we use the algorithm that corrects for overdispersion given that there are a large number of zeroes in the lists. Table A.9 presents predicted probabilities of support from these analyses by framing condition, while A.10 presents the coefficient estimates. The results in the main text are robust to the use of these algorithms, though the effect of the negative frame is slightly attenuated (an eight percentage point difference between the control and negative treatment, compared to 12 percentage points in the linear regression reported in the text).

B Balance tests

Figure B.1 shows the p-values for the estimated coefficients on four demographic variables in each of our four framing experiments and three treatment arms (Tables B.1-4 reports

Figure A.8: Relationship between direct responses to control list items and list responses (cleaned data)



(a) All framing conditions

(b) Control framing condition

(c) Positive framing condition

(d) Negative framing condition

Note: Shaded areas represent 95% confidence intervals about linear regression estimates. Model specifications in Table A.8, columns 2-5.

Table A.9: Predicted prevalence of support for Putin across experimental conditions, using Imai (2011) and Blair and Imai (2012) MLE algorithms

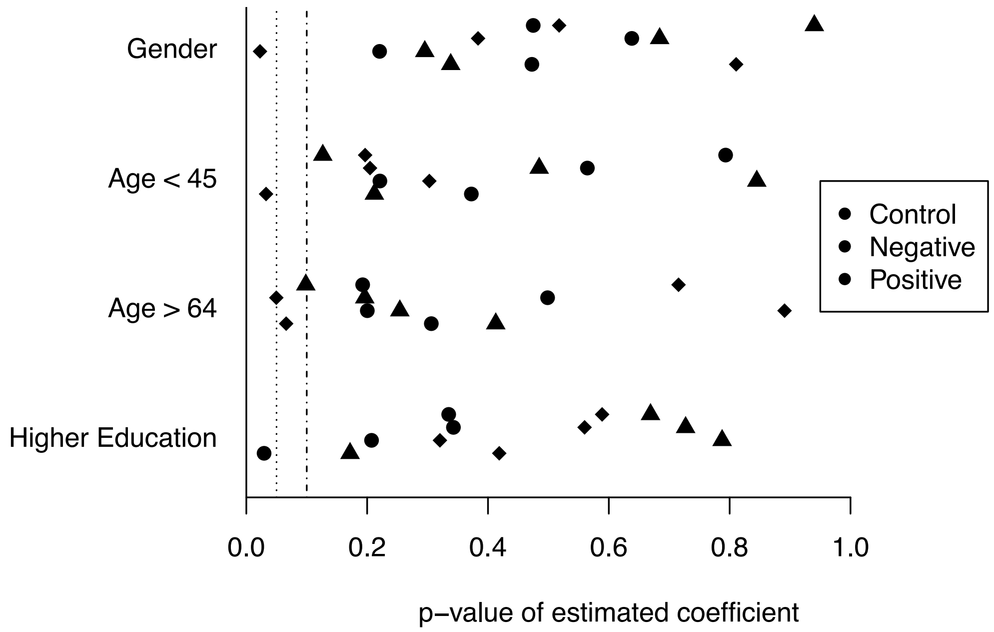
	Standard	Floor	Overdispersed
Control	0.48 (0.02)	0.56 (0.01)	0.51 (0.01)
Positive frame	0.46 (0.02)	0.52 (0.01)	0.50 (0.03)
Negative frame	0.40 (0.02)	0.47 (0.01)	0.43 (0.03)

Note: Predicted prevalence based on parameter estimates from Table A.10.

Table A.10: Parameter estimates of support for Putin across experimental conditions, using Imai (2011) and Blair and Imai (2012) MLE algorithms

	Standard	Floor	Overdispersed
Sensitive item			
Control	-0.08 (0.08)	0.23 (0.10)	0.05 (0.09)
Positive frame	-0.10 (0.12)	-0.17 (0.14)	-0.07 (0.12)
Negative frame	-0.32 (0.11)	-0.35 (0.14)	-0.34 (0.13)
Control items			
Control	-0.64 (0.02)	-0.65 (0.02)	-0.66 (0.02)
Positive frame	0.02 (0.03)	0.02 (0.03)	0.01 (0.03)
Negative frame	-0.01 (0.03)	-0.01 (0.03)	-0.01 (0.03)
Floor			
Control		-0.51 (0.12)	
Positive frame		-0.51 (0.12)	
Negative frame		-0.51 (0.12)	
Dispersion parameter			-1.73 (0.05)
LogLik	-20,205	-20,123	-19,853
N	14,582	14,582	14,582

Figure B.1: Balance tests by framing condition



Note: Points represent p-values of coefficient estimates from Tables B.1, B.2, B.3 and B.4.

the model estimates shown in this figure). Each point represents the p-value from an OLS regression of a treatment arm indicator (control group, negative frame, or positive frame) on a set of four binary respondent demographic characteristics. Only 4 out of 48 p-values (8.33%) are significant at the 5% level, which is very close to random chance. Based on these balance tests, we have no reason to believe that any of our randomizations in the four framing experiments we conducted are systematically flawed.

C Additional analyses and robustness

C.1 Table 1 robustness

We estimate framing treatment effects using separate t-tests of the two frames relative to the control. Table C.1 presents the results, which are in line with those reported in the text. In Table C.2, we replicate the results from Table 1, but with clustering of the standard errors by Russian subnational unit (region). The results are robust to clustering standard errors by region. In Table C.3, we replicate the results from Table 1, but with the addition of four demographic control variables: gender, an indicator variable for age under 45, an indicator variable for age over 64, and an indicator variable for having higher education. The addition of these variables does not substantively affect our results.

Table B.1: Balance tests for Levada framing experiment by frame

	Control	Negative	Positive
Constant	0.36*** (0.02)	0.31*** (0.02)	0.33*** (0.02)
Male	−0.02 (0.02)	0.002 (0.02)	0.02 (0.02)
Higher education	−0.02 (0.03)	0.01 (0.03)	0.01 (0.03)
Age under 45	−0.01 (0.03)	0.04 (0.03)	−0.03 (0.03)
Age over 64	−0.05 (0.04)	0.06* (0.04)	−0.01 (0.04)
Observations	1,607	1,607	1,607
R^2	0.002	0.002	0.001

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

Linear probability models where outcome is a dichotomous indicator for a given experimental frame.

Table B.2: Balance tests for POADSRR nationally representative framing experiment by frame

	Control	Negative	Positive
Constant	0.30*** (0.03)	0.34*** (0.03)	0.36*** (0.03)
Male	0.01 (0.02)	0.01 (0.03)	−0.02 (0.02)
Higher education	0.02 (0.03)	−0.01 (0.03)	−0.02 (0.03)
Age under 45	0.02 (0.03)	0.02 (0.03)	−0.03 (0.03)
Age over 64	−0.03 (0.05)	−0.06 (0.05)	0.09** (0.05)
Observations	1,504	1,504	1,504
R^2	0.001	0.002	0.005

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

Linear probability models where outcome is a dichotomous indicator for a given experimental frame.

Table B.3: Balance tests for RES framing experiment by frame

	Control	Negative	Positive
Constant	0.35*** (0.03)	0.34*** (0.03)	0.31*** (0.03)
Male	−0.03 (0.03)	−0.03 (0.03)	0.06** (0.03)
Higher education	−0.04 (0.03)	0.01 (0.03)	0.03 (0.03)
Age under 45	0.04 (0.03)	−0.01 (0.03)	−0.03 (0.03)
Age over 64	−0.05 (0.04)	0.05 (0.04)	0.01 (0.04)
Observations	1,324	1,324	1,324
R^2	0.01	0.003	0.01

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

Linear probability models where outcome is a dichotomous indicator for a given experimental frame.

Table B.4: Balance tests for POADSRR regionally representative experiments, by frame and list treatment

	Framing experiment			List
	Control	Negative	Positive	Treatment
Constant	0.33*** (0.01)	0.32*** (0.01)	0.35*** (0.01)	0.50*** (0.01)
Male	0.01 (0.01)	−0.01 (0.01)	0.002 (0.01)	−0.004 (0.01)
Higher education	−0.02** (0.01)	0.01 (0.01)	0.01 (0.01)	0.004 (0.01)
Age under 45	0.01 (0.01)	0.01 (0.01)	−0.02** (0.01)	0.01 (0.01)
Age over 64	0.02 (0.02)	0.02 (0.02)	−0.04* (0.02)	−0.01 (0.02)
Observations	16,341	16,341	16,341	16,333
R^2	0.0004	0.0003	0.0004	0.0001

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

Linear probability models where outcome is a dichotomous indicator for a given experimental condition.

Table C.1: Estimated effect of framing treatments on prevalence of support for Putin across survey waves

	Levada National Nov 2020 Direct	POADSRR National Jun 2021 Direct	RES National Sep 2021 Direct	POADSRR Regional Aug 2021 Direct	POADSRR Regional Aug 2021 List
Positive frame	(-0.04, 0.08)	(-0.07, 0.06)	(-0.05, 0.08)	(-0.02, 0.02)	(-0.04, 0.09)
Negative frame	(0.02, 0.13)	(-0.01, 0.12)	(0.01, 0.14)	(0.09, 0.13)	(0.04, 0.17)
Observations	1,554	1,503	1,277	16,342	7,092

Note: Quantities represent 95% confidence intervals from t-tests estimating the effect of framing conditions relative to the control. Effects in list experiments estimated only using list treatment condition.

Table C.2: Framing effects on support for President Putin, clustered standard errors

	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***
	(0.03)	(0.02)	(0.03)	(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)	(0.05)
	Control list				
Constant					1.00***
					(0.02)
Positive					0.02
					(0.03)
Negative					0.01
					(0.02)
Observations	1,554	1,503	1,277	16,342	14,577
Num clusters	50	82	62	60	60
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. Standard errors clustered CR2 by region.

Table C.3: Framing effects on support for President Putin, with demographic controls

	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	0.62***	0.59***	0.69***	0.60***	0.56***
	(0.03)	(0.03)	(0.03)	(0.01)	(0.03)
Positive	−0.03	0.001	−0.02	−0.003	−0.05
	(0.03)	(0.03)	(0.03)	(0.01)	(0.04)
Negative	−0.08***	−0.05*	−0.08**	−0.11***	−0.12***
	(0.03)	(0.03)	(0.03)	(0.01)	(0.04)
			Control list		
Constant					1.08***
					(0.03)
Positive					0.02
					(0.03)
Negative					0.02
					(0.03)
			Demographic controls		
Male	-0.02	-0.10***	-0.10***	-0.09***	0.07***
	(0.02)	(0.03)	(0.03)	(0.01)	(0.02)
Age under 45	-0.05*	-0.07**	-0.01	-0.03***	-0.13***
	(0.03)	(0.03)	(0.03)	(0.01)	(0.02)
Age over 64	0.20***	0.14***	0.16***	0.13***	0.16***
	(0.04)	(0.05)	(0.04)	(0.02)	(0.05)
Higher education	0.06*	0.003	0.005	0.03**	-0.05***
	(0.03)	(0.03)	(0.03)	(0.01)	(0.02)
Observations	1,554	1,503	1,272	16,329	14,581
R^2	0.04	0.02	0.03	0.02	0.07

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.

C.2 Ordered probit analyses of framing experiment

In Table C.4, we replicate the analysis presented in Columns 1–4 in Table 4 using ordinal probit rather than the linear probability model. Our results are largely unchanged, though the negative coefficient in the RES survey loses statistical significance.

Table C.4: Ordered probit analyses of framing experiment

	Levada National Nov 2020	POADSRR National Jun 2021	RES National Sep 2021	POADSRR Regional Aug 2021
Positive	0.001 (0.07)	−0.02 (0.07)	0.03 (0.07)	−0.01 (0.02)
Negative	-0.13^{**} (0.07)	-0.11^{*} (0.07)	−0.07 (0.07)	-0.21^{***} (0.02)
		Thresholds		
1|2	-0.91^{***} (0.05)	-0.74^{***} (0.05)	-1.11^{***} (0.06)	-0.96^{***} (0.02)
2|3	-0.29^{***} (0.05)	−0.05 (0.05)	-0.36^{***} (0.06)	-0.13^{***} (0.02)
3|4	0.66^{***} (0.05)	0.99^{***} (0.06)	1.02^{***} (0.06)	1.11^{***} (0.02)
AIC	4,219	4,051	3,189	42,161
Observations	1,554	1,503	1,277	14,577

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

C.3 Changes in outcome distribution across experimental conditions

We also note another important consistency across survey waves: treatment effects are largely concentrated in the bottom three categories (Table C.5). 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 C.5: 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	0.23	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

Note: POADSRR data from subnationally-representative survey.