

Countering Authoritarian Censorship via Grassroots Communication Campaigns: Evidence from an Experimental Mega-Study in Wartime Russia

Julia Minson

`julia_minson@hks.harvard.edu`

Harvard University

Aaron Erlich

McGill University

Jordan Gans-Morse

Northwestern University

Christopher Higgins

Harvard University

Ruth Elizabeth Appel

Stanford University

Christiane Baumann

Harvard University

Hui Bai

Stanford University <https://orcid.org/0000-0003-2671-5955>

Sarah Chen

Harvard University

Sierra Davis Thomander

Stanford University

Charles Dorison

Georgetown University

Eran Halperin

The Hebrew University of Jerusalem <https://orcid.org/0000-0002-3379-2935>

John Jost

New York University <https://orcid.org/0000-0002-2844-4645>

Fabian Falch

Mail2RU

Irene Lee

Harvard University

Jennifer Lerner

Harvard University <https://orcid.org/0000-0002-1255-4196>

Alexey Makarin

MIT Sloan School of Management

Molly Moore

Harvard University <https://orcid.org/0000-0001-5868-8272>

Nimrod Nir

The Hebrew University of Jerusalem

Scott Radnitz

University of Washington

Jane Risen

University of Chicago

Maurice Schweitzer

University of Pennsylvania

Anton Shirikov

University of Kansas

Georgiy Syunyaev

Vanderbilt University <https://orcid.org/0000-0002-4391-6313>

Madalina Vlasceanu

Stanford University <https://orcid.org/0000-0003-2138-1968>

Ke Wang

University of Chicago

Nimrod Zeldin

The Hebrew University of Jerusalem

Article

Keywords:

Posted Date: May 12th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8633606/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

**Countering Authoritarian Censorship via Grassroots Communication Campaigns:
Evidence from an Experimental Mega-Study in Wartime Russia**

Julia Minson^{1*}, Aaron Erlich^{2*}, Jordan Gans-Morse^{3*}, Christopher Higgins^{1*}, Ruth Elisabeth Appel⁴, Christiane Baumann¹, Hui Bai⁵, Sarah Chen¹, Sierra Davis Thomander⁶, Charles Dorison⁷, Eran Halperin⁸, John Jost⁹, Fabian Koenig¹⁰, Irene Lee¹, Jennifer Lerner¹, Alexey Makarin¹¹, Molly Moore¹, Nimrod Nir⁸, Scott Radnitz¹², Jane Risen¹³, Maurice Schweitzer¹⁴, Anton Shirikov¹⁵, Georgiy Syunyaev¹⁶, Madalina Vlasceanu¹⁷, Ke Wang¹, and Nimrod Zeldin¹⁸

Affiliations:

*: Core Research Team

¹: Harvard Kennedy School, Harvard University, Cambridge, USA

²: Department of Political Science, McGill University, Montréal, Canada

³: Department of Political Science, Northwestern University, Evanston, USA

⁴: Department of Communication, Stanford University, Stanford, USA

⁵: Political Belief Lab, Stanford University, Stanford, USA

⁶: Department of Political Science, Stanford University, Stanford, USA

⁷: McDonough School of Business, Georgetown University, Washington D.C., USA

⁸: Department of Psychology, The Hebrew University of Jerusalem, Jerusalem, Israel

⁹: Department of Psychology, New York University, New York City, USA

¹⁰: Boston Consulting Group, Boston, USA

¹¹: Sloan School of Management, Massachusetts Institute of Technology, Cambridge, USA

¹²: The Henry M. Jackson School of International Studies, University of Washington, Seattle, USA

¹³: Booth School of Business, The University of Chicago, Chicago, USA

¹⁴: Wharton School, University of Pennsylvania, Philadelphia, USA

¹⁵: Department of Political Science, The University of Kansas, Lawrence, USA

¹⁶: Department of Political Science, Vanderbilt University, Nashville, USA

¹⁷: Department of Environmental Social Sciences, Stanford University, Stanford, USA

¹⁸: Department of Geography, The Hebrew University of Jerusalem, Jerusalem, Israel

Abstract

In response to Russia's 2022 full-scale invasion of Ukraine, citizen-led communication campaigns leveraged modern information technology to transmit millions of messages providing uncensored information to Russians via emails, texts, social media apps, and website commentary. Yet it remains unclear to what extent such initiatives can penetrate authoritarian censorship and propaganda, and which types of messaging are most likely to succeed. To investigate, we partnered with Mail2Ru, a prominent citizen-led grassroots initiative, to conduct two pre-registered field experiments on the effectiveness of citizen-to-citizen email communication. For Study 1, a large-scale "mega-study," we invited behavioral science experts to design theoretically informed persuasive messages encouraging Russians to watch an uncensored video about the invasion. Approximately 260,000 email recipients were randomly assigned to one of eleven treatments or two control conditions. Study 1 showed that information technology's scalability enables even poorly designed messages to engage large audiences, but also that most expert-designed messages failed to outperform controls. Our top-performing intervention, however, nearly doubled engagement with the video. Study 2 successfully replicated the benefits of our top-performing intervention six months later, confirming its robustness. Study 3 probed mechanisms underlying our field experiment results, producing initial evidence that in the challenging context of censorship and repression, strategies emphasizing receptiveness and respectful engagement with opposing perspectives might be most effective. Together, these findings offer evidence-based guidance for citizen-led efforts to counter authoritarian propaganda.

Research Transparency Statement

The authors have no conflicts of interest or funding sources to report. There was no use of artificial intelligence, and the research was approved by the IRB of the first author's university. Our pre-registrations, materials, and code can be found [here](#).

Introduction

Russia's full-scale invasion of Ukraine in February of 2022 plunged Europe into its deadliest conflict since World War II. In response, numerous grassroots citizen-to-citizen communication campaigns emerged across the globe, employing modern information technology to counter Kremlin propaganda by delivering uncensored content behind the front lines. For example, individuals worldwide used the online reviews of Russian businesses to urge anti-war protests (Stokel-Walker, 2023), the 'Papa pover' ("Papa, believe me") project implored Ukrainians to call relatives in Russia (Rebane, 2022), and the Polish project Squad303 allowed anyone with a cell phone to send anti-war text messages to Russian phone numbers (Harwell, 2022).

Swaying conflict outcomes through unsolicited persuasive communication is not new. During World War 2, the Allied Forces distributed two *billion* leaflets over Japan alone (Szasz, 2009). Nevertheless, the global response to Russia's invasion underscores the extent to which technology now allows direct citizen-to-citizen contact at massive scale. All told, these initiatives have transmitted hundreds of millions of emails, text messages, phone calls, and pieces of online commentary aimed at providing uncensored information to Russians.

Large scale notwithstanding, it remains unclear to what extent such initiatives can penetrate authoritarian censorship and propaganda, or which types of messaging are most likely to succeed. Indeed, these questions have relevance far beyond Russia's invasion of Ukraine. Scholars argue that even in authoritarian regimes, regime stability and leaders' survival depend on sustaining a considerable degree of popular support (Gandhi, 2008; Greene & Robertson, 2019; Svolik, 2012). Consequently, censorship and propaganda serve as essential tools for authoritarian regimes seeking to bolster public approval (Rosenfeld & Wallace, 2024). Recognizing that popular support becomes all the more important when wars require citizen

sacrifice, authoritarian regimes often ramp up propaganda efforts during times of war. Since 2022, for example, Russia has employed censorship to limit public awareness of the invasion's scale and consequences, even criminalizing references to "war" and requiring citizens and journalists to instead use the term "special military operation" (Tebor, 2022). Likewise, starting with its initial invasion of Ukraine in 2014, the Russian government has utilized propaganda to frame its war as defensive, citing, among other justifications, NATO expansion and threats to ethnic Russians in Ukraine (Erlich & Garner, 2021). Given authoritarian regimes' rising global influence, conflict outcomes and authoritarians' grip on power will critically depend on battles for hearts and minds. Accordingly, the citizen-to-citizen communication campaigns that emerged in response to Russia's 2022 invasion are likely to become an enduring phenomenon worthy of rigorous study.

Of course, such campaigns face immense challenges. Even without censorship, propaganda, and "rally around the flag" effects common during times of war, generating engagement with information that contradicts deeply entrenched beliefs is difficult. An extensive literature in communications, psychology, and political science has documented the phenomenon of selective exposure (or "congeniality bias"), defined as individuals' tendency to consume information that supports prior beliefs and avoid information that contradicts them (Frey, 1986; Hart et al., 2009; Sears & Freedman, 1967). Moreover, the difficulty of reaching citizens in an authoritarian regime under wartime conditions requires the use of communication strategies (such as unsolicited messages from unfamiliar senders) that are widely recognized as suboptimal (Kardas et al., 2022; McGinnies & Ward, 1980; Wittenberg et al., 2021). In the case of Russia's war in Ukraine, these challenges only increased over time. Russian censors have become more adept at blocking incoming campaigns, millions of the most war-averse Russians have emigrated, and

those who have stayed may have learned about the high costs of resistance (Dean & Porter, 2024; Tertychnaya & Tiratsoo, 2024). These issues are particularly relevant for our studies – our first field experiment was conducted in the summer of 2023, nearly a year and a half into the war, and our second experiment approximately six months later. Interventions in other authoritarian contexts also are likely to confront similar dynamics.

Nevertheless, the behavioral sciences are well-positioned to shed light on the questions of whether and through what mechanisms such campaigns can be effective. For example, in recent years scholars have developed numerous theoretically grounded interventions for reducing selective exposure and information avoidance (Dorison, Minson & Rogers, 2019; Golman, Hagmann, & Loewenstein, 2017; Kirgios et al., forthcoming; Moore, Dorison & Minson, 2023). However, most relevant experiments have been conducted primarily in the lab rather than in the field, disproportionately in North America or western Europe, and often with a small number of treatment arms. None, to the best of our knowledge, have been conducted in the challenging context of a society at war. More broadly, prominent communication studies scholars such as O’Keefe (2018) have emphasized the need for experimental evaluations of persuasive messaging, especially in light of evidence that commonly used non-experimental measures of expected or perceived persuasiveness often correlate weakly with experimental findings.

To fill these gaps, we report the results of two large-scale, pre-registered field experiments on the effectiveness of citizen-to-citizen email communication in wartime Russia and a third study designed to probe the mechanisms underlying our field experiment results. As discussed in [Appendix 6](#), we undertook extensive measures to ensure that our research minimized risk to participants and adhered to principles of ethical research involving human subjects. The Harvard University Institutional Review Board approved both field experiments.

The field experiments were carried out in partnership with [Mail2Ru](#), a citizen-led, non-profit initiative that created a semi-automated platform through which people from around the globe can send emails with uncensored information about the war to Russian citizens. For Study 1, we implemented a “mega-study,” a study that tests numerous treatments simultaneously while employing a common sample, set of outcome measures, and control conditions (Milkman et al., 2021; Voelkel et al., 2024). Specifically, we invited over 100 experts from several academic fields to design theoretically informed persuasive messages encouraging Russians to watch an uncensored video about the invasion. In July and August 2023, we then randomly assigned approximately 260,000 Russian email recipients to eleven of the most promising expert messages and two control groups, one of which featured messages written by laypeople, the other of which contained only the link to the target video.

Study 1 offered compelling evidence of citizen-to-citizen communication campaigns’ potential for penetrating authoritarian censorship and propaganda. Even when messaging is poorly designed and click rates are low, information technology’s scalability facilitates engagement with large audiences. For example, since the 2022 invasion, our partner organization Mail2Ru has sent hundreds of millions of emails. Even if Mail2Ru had employed the content of our control group message (which, as discussed below, was the equivalent of spam), our estimated baseline click rate indicates that these efforts would have exposed approximately three-quarters of a million Russians to the target video.

Study 1 also showed both the promise and limitations of persuasive messaging informed by behavioral science theories. On the one hand, even facing an extraordinarily challenging test – persuasion via unsolicited messages to recipients in a wartime authoritarian regime – our top-performing intervention approximately doubled engagement with the video. This successful

intervention, a message best characterized as “expressing respectful curiosity,” drew on theories emphasizing the importance of conveying respect, warmth, and receptiveness to others’ perspectives (Collins et al., 2022; Lerner & Tetlock, 1999; Minson & Chen, 2022; Shapiro, 2017). On the other hand, *all* other interventions failed to outperform either of the controls.

Given that only one intervention outperformed the control treatment, we sought to confirm this finding’s robustness with a replication study. Study 2, conducted in February 2024, randomly assigned approximately 148,000 email recipients to the successful intervention from Study 1 or to a control group. This replication study also facilitated analysis of whether interventions remain feasible and effective as the war goes on and attitudes become further entrenched. Relative to Study 1, conducted half a year earlier, we faced greater technical challenges in getting emails delivered and a lower baseline click rate in the control group, possibly reflecting Russians’ growing caution in the face of escalating repression and censorship. But the successful intervention from Study 1 again approximately doubled engagement with the uncensored video.

Finally, we implemented a third study to investigate the mechanisms that set apart our most successful treatment message. Combining layperson ratings and algorithmic text analysis to evaluate the psychological strategies embedded in each intervention, we identify a cluster of mechanisms – seeking perspective, communicating with warmth, and conversational receptiveness – that are positively associated with treatment effects. These findings offer suggestive evidence that in the challenging context of censorship and repression, strategies emphasizing receptiveness and respectful engagement with opposing perspectives (Hagmann, Minson, & Tinsley, 2024; Minson & Chen, 2022; Tulan, Dorison, & Minson, 2024; Yeomans, et al. 2022) might be more effective than direct appeals based on facts or argumentation.

Study 1

Study 1 is a mega-study field experiment examining the effectiveness of eleven treatments designed by expert academics in a citizen-to-citizen email communication in wartime Russia. Our primary, pre-registered outcome of interest was Russian email recipients' engagement with a video containing accurate, uncensored information, measured as the rate at which recipients clicked on the video link after taking into account undelivered (i.e., "bounced") emails.¹ Of the 260,088 emails sent in Study 1, we confirmed that 233,132 were delivered. As discussed in the Method section, our message delivery protocol was designed to ensure approximately equal sample sizes across experimental arms in the case of technical problems or censorship.

Click rates are a widely used metric in behavioral science and communications research that captures an unambiguous first step to deeper engagement with any content: All other measures are only informative contingent on clicking (Chatterjee et al., 2003; Manchanda et al., 2006; Rogers et al., 2017). Moreover, although data limitations (detailed in the Method section) prevent precise analysis of viewing duration, aggregate-level data confirm that recipients who clicked the link watched, on average, at least a third of the 90-second video, indicating that clicks led to meaningful engagement. Our focus on click rates was additionally motivated by ethical concerns related to Russia's criminalization of speech that defames the military. Whereas collecting data on self-reported attitudes would have potentially endangered participants and raised concerns about truthful reporting, our behavioral indicator minimizes such risks. Overall, our focus on initial engagement is especially fitting for research in authoritarian contexts, in which even brief exposure to alternative narratives constitutes a meaningful departure from regimes' efforts to

¹ As secondary measures, we also collected data on comments and likes posted by viewers of the video, but these were too few to be analyzable.

maintain a propaganda-sustaining monopoly over information (Rosenfeld & Wallace, 2024, p. 273).

Our first result concerns citizen-to-citizen communication campaigns' potential for penetrating authoritarian censorship and propaganda. We find that even with poorly designed messaging, citizen-to-citizen campaigns have the capacity to reach large audiences. Concretely, 0.24% of recipients assigned to the control group that received emails containing only the link to the target video and no text (i.e., the equivalent of spam) at least partially viewed the video. Though small in absolute terms, this is a meaningful figure in the context of highly scalable modern communication campaigns. The small non-profit Mail2Ru has sent over 300,000,000 emails since the start of the 2022 invasion, suggesting that a message of similar effectiveness would have exposed at least 720,000 Russians to information they otherwise would not have seen. Given the financial and technological resources that both governmental and non-governmental organizations regularly deploy toward combating misinformation, this click rate suggests that similar campaigns have the potential to wield substantial influence.

Our second result addresses the question of whether the type of messaging employed affects campaigns' effectiveness. As detailed in the Method section, our mega-study solicited, selected, and tested eleven messages designed by scholars with expertise related either to persuasion or to regional studies. Table 1 provides an overview of the theoretical rationale for each message.

[Appendix 1](#) includes the full text of each message sent and the theoretical justifications submitted by each expert or team of experts for their proposed intervention's design.

Table 1: Overview of Treatment Messages

<p>Treatment A: "Expressing respectful curiosity"</p> <p>This treatment identifies the sender as belonging to an international group of students and faculty seeking to learn from Russian citizens' firsthand perspectives rather than relying on the media. Drawing on theories of respect (Shapiro, 2017), receptiveness to others' viewpoints (Minson et al., 2020), and the power of gratitude (Grant & Gino, 2010), the message emphasizes the unique value of the recipient's perspective. The treatment employs multiple behavioral science principles, including descriptive norms ("thousands of people have watched it") and accountability to encourage systematic, unbiased responses (Lerner & Tetlock, 1999).</p>
<p>Treatment B: "Invoking a common identity"</p> <p>This treatment identifies the sender as someone with Russian heritage, highlighting the sender's in-group status, and establishes connection with specific aspects of Russian identity – strategic thinking exemplified in chess victories and heroism exemplified in the triumphs in World War II. The treatment applies group-affirmation theory (Badea & Sherman, 2019) to frame engagement with new information as an expression of traditionally valued Russian traits: intelligence, strength, and strategic thinking.</p>
<p>Treatment C: "Leveraging the 'chosen one' trope"</p> <p>This treatment portrays Russians as uniquely capable of critically evaluating information and approaching diverse perspectives objectively, stating, for example: "Russians like you are strong people. Rather than relying on what others say you should think, you are dedicated to figuring things out for yourself." It leverages the "chosen one" trope from popular culture to validate the recipient's self-image as being "one of a few" qualified to evaluate the video.</p>
<p>Treatment D: "Warning about deception"</p> <p>Employing the style of online conspiracy content to capture attention, this treatment warns that corrupt forces are concealing information about the "special military operation" to advance their own interests. The message incorporates stylistic elements common to conspiracy theorists (Fong et al., 2021) and clickbait techniques (Mormol, 2019). The message also is intentionally short, motivated by the idea that fewer words increase the likelihood that readers will read to the end.</p>
<p>Treatment E: "Seeking to correct biases"</p> <p>This message presents itself as part of a Western academic study examining potential biases in media coverage, acknowledging that "Western journalists and media can cover world events in a biased way." Drawing on research on news credibility assessment (Pennycook et al., 2021), the treatment positions the recipient's feedback as valuable input that could help improve Western media coverage. The message deliberately frames the task as measuring group views rather than changing individual opinions, aiming to reduce resistance to engagement.</p>
<p>Treatment F: "Using brevity"</p> <p>This briefest treatment consisted of a single sentence making a simple request framed as a favor. The approach draws on research about moral reframing (Feinberg & Willer, 2019) and the importance of benevolence in Russian society (Stepanova, 2022).</p>
<p>Treatment G: "Shared emotions"</p> <p>This treatment shares the author's vulnerability by expressing feelings of helplessness and confusion about the war and emphasizing solidarity with Russians who feel similarly conflicted. The message carefully avoids challenging beliefs or values (Taber & Lodge, 2006; Nyhan & Reifler, 2010), instead focusing on shared concerns about Russian soldiers' lives and connecting to cherished historical memories of World War II (Sharafutdinova, 2020). By acknowledging common feelings of helplessness (Zhelnina, 2020; Alyukov, 2022), the message positions viewing the video as a way to find relief through accessing reliable information.</p>
<p>Treatment H: "Highlighting responsibility"</p> <p>This message combines decreased psychological distance (Kossowska et al., 2023) with appeals to personal responsibility, asking "How will you feel about your actions today 20 years from now?" It leverages dynamic social norms (Sparkman & Walton, 2017) and identity social norms (Howe, Carr, & Walton, 2021) while emphasizing collective action efficacy by referencing past Russian achievements. The treatment links viewing the video to making a decision that will make the reader proud in the future.</p>

Treatment I: "Seeking nuanced perspectives"

Positioning itself as a response to perceived Western media bias, this message invites recipients to "challenge the one-sided narrative" by sharing their perspectives. The treatment activates Russian identity while presenting comment-leaving as a concrete way to address perceived threats to Russia's image. It frames engagement as taking action to ensure Russian voices are heard in global dialogue, appealing to desires for balanced representation and agency in international discourse.

Treatment J: "Stoking conspiracy intuitions"

Drawing on research regarding the effects of "conspiracy intuitions" (Roberts & Risen, 2022), this treatment taps into the common feeling that official narratives may be incomplete or deliberately misleading. Without naming a specific conspiracy theory, the message instead points to the recipient's possible sense that "powerful groups are hiding information from ordinary people" and presents the video as an opportunity to learn potentially hidden information.

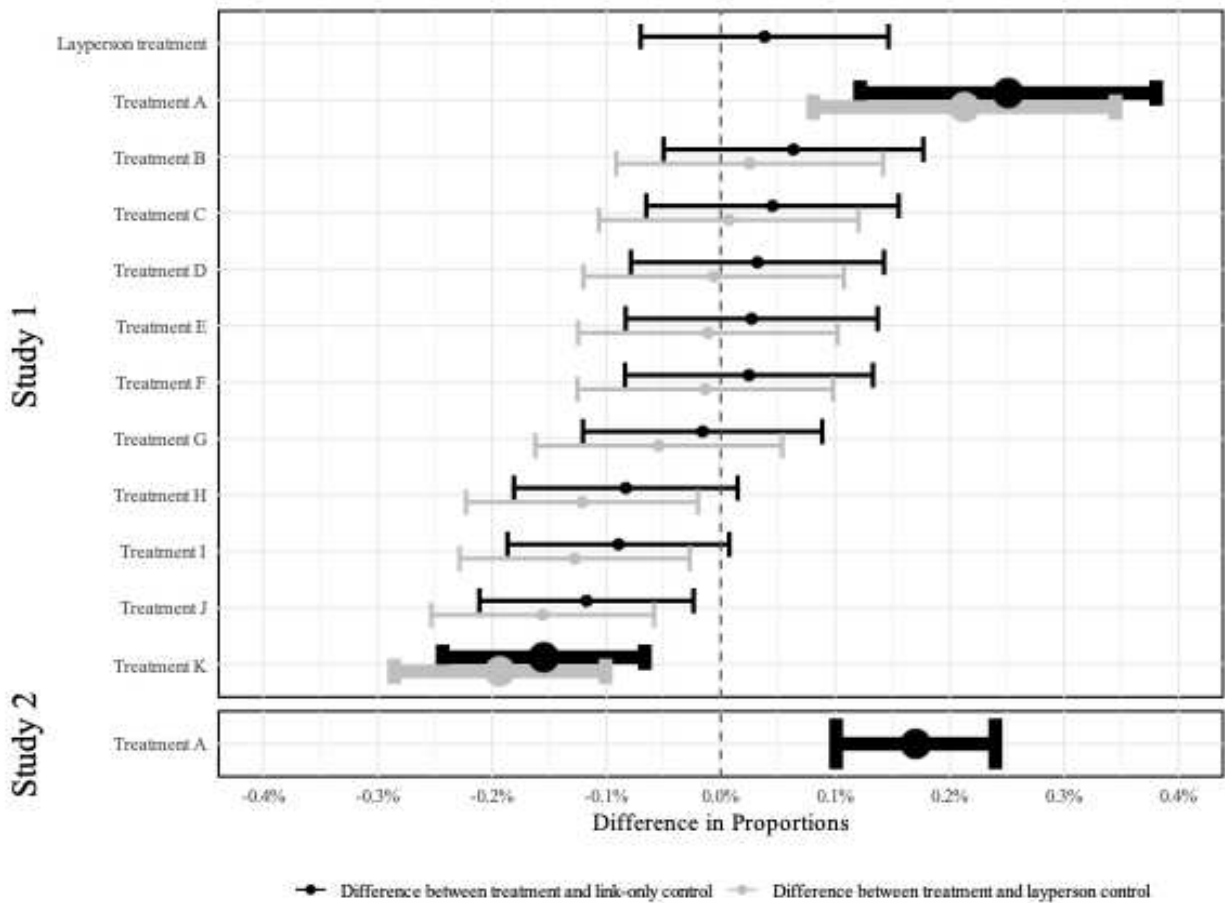
Treatment K: "Leveraging a literary metaphor"

Using a literary metaphor from respected Russian author Victor Pelevin, this treatment compares watching the video to picking up a traveler in a truck driver's cab. The message draws on theories about curiosity (Loewenstein, 1994), familiarity-based trust (Gulati, 1995; Renn & Levine, 1991), and persuasive analogies (Roehm & Sternthal, 2001) while leveraging Pelevin's non-political cultural authority to encourage engagement.

We compared these interventions to two control conditions. The first consisted of the above-mentioned "link only" message which contained no text, only a hyperlink to the target video. The second consisted of a set of messages written by laypeople. We constructed this condition by soliciting 100 messages from paid participants on the Prolific Academic platform, randomly selecting ten out of 100, and randomizing each participant in the layperson control group to receive one of the ten. In sum, approximately 20,000 email recipients were randomly assigned to each of the thirteen experimental arms (eleven treatment interventions, two control groups), a number chosen to maintain sufficiently high levels of statistical power (see [Appendix 4](#)).

The upper panel of Figure 1 presents our primary results, showing differences in click rates for each of the expert messages relative to the link-only control (effect sizes denoted in black) and the layperson control (effect sizes denoted in grey). The top of the figure also displays the difference between the link-only and layperson control conditions, which were statistically and substantively indistinguishable. Treatments with effect sizes greater than zero and confidence intervals that do not intersect the vertical dotted line overperformed the control

Figure 1: Differences in click rates (effect sizes) by treatment for Studies 1 and 2



Note: The figure above presents differences in click rates (effect sizes) for each treatment relative to the link-only control (black) and layperson control (grey). In total, 233,132 and 72,159 emails were delivered for Studies 1 and 2, respectively. Error bars represent 95% confidence intervals. The top panel presents treatments from Study 1. In Study 1, bolded lines represent results where the effect is statistically significant relative to both controls (link-only and layperson), and after adjusting for multiple comparisons using the Benjamini-Hochberg adjustment (Benjamini & Hochberg, 1995). The bottom panel compares the single treatment from Study 2 to the link-only control.

conditions at a .05 significance level; treatments with effect sizes less than zero and non-intersecting confidence intervals underperformed. We statistically adjust our hypothesis tests to account for multiple comparisons by employing the Benjamini-Hochberg correction (Benjamini & Hochberg, 1995). Interventions that remain statistically significant at a .05 level relative to both control conditions when accounting for multiple comparisons are bolded.

Figure 2: Text for Treatment A

Hello! My name is XXX. Please forgive me for writing out of the blue. I saw online that you live in Russia, and I thought you might be able to help me and my colleagues understand the Ukraine-Russian conflict better.

I'm part of a group of university students and faculty from different countries around the world who want to learn more about the conflict by hearing from actual citizens rather than from major media. You are in a better position than us to know the first-hand realities of what is going on.

We're curious to know your opinion about the news clip below, especially about how it ends; it was suggested by a researcher originally from Russia. Apparently, thousands of people have watched it, so you may have already seen it. We're eager to hear your opinion so we can learn what is really going on.

*Click here to watch the news clip [LINK TO VIDEO]
(free on YouTube; watch as many times as you like)*

=> To help us learn your views, please add a comment on YouTube. My colleagues and I will check there to learn what you (and others in Russia) think about the clip.

Note: The figure above shows the message for the top-performing intervention in Study 1.

Both the promise and limitations of experts' understanding of persuasive messaging strategies are apparent in our results. On the promising side, the top-performing intervention – Treatment A – significantly outperformed both control groups, producing a click rate of 0.50%, more than doubling the 0.24% click rate of the link-only condition (CI [0.12%-0.38%], $p < .001$, using a difference in proportions test and adjusting for multiple comparisons) and nearly twice the 0.28% click rate of the layperson condition. This intervention, the text of which is displayed in Figure 2, drew on theories emphasizing the importance of conveying receptiveness to others' viewpoints (Minson & Chen, 2022), respect for the recipient's time and expertise (Shapiro, 2017), gratitude (Grant & Gino, 2010), and accountability (Lerner & Tetlock, 1999). It also sought to raise the salience of descriptive norms by emphasizing the video's widespread viewership. (See [Appendix 1](#) for additional details.) On the less promising side, despite being designed by experts and selected by the research team due to their perceived promise, *all* other interventions failed to outperform (with several underperforming) the control conditions.

Study 2

To further examine the robustness of our findings about Treatment A's effectiveness, we conducted a second field experiment. As reported above, the effect for Treatment A was statistically robust: it remained statistically significant even when adjusting for the mega-study's multiple interventions. Nevertheless, it remained an open question whether it was replicable. Considering that Russian censorship and propaganda were becoming more extreme over time, we wished to evaluate Treatment A's effectiveness under changing conditions. Pilot testing confirmed that our partner Mail2Ru's capacity to deliver large numbers of emails to Russian recipients was declining as Russian censorship became more effective. Accordingly, to maintain sufficiently high levels of statistical power, we decided to re-test only Treatment A and, given the two control conditions' indistinguishable effects in Study 1, we employed only the link-only condition in the second experiment, resulting in a two-arm design comparing Treatment A to a single control group.

Launched approximately six months after our first experiment, Study 2 consisted of sending 148,141 messages, of which 72,159 were confirmed to be delivered. The baseline link-only click rate for Study 2 was 0.13%, which was lower than in Study 1, possibly reflecting greater wariness among recipients. But, as shown above in the lower panel of Figure 1, Treatment A again outperformed the control condition at a statistically significant level (.17%, CI [.10%-.24%], $p < .001$). Indeed, as in Study 1, the Treatment A click rate of 0.30% was more than double that of the baseline link-only condition.

Overall, the robust, replicable results for Treatment A demonstrate that carefully designed interventions can meaningfully boost engagement, even when conveyed via unsolicited messages in the context of a wartime authoritarian regime. Yet the failure of ten out of eleven expert-

designed messages to outperform control messages also underscores the difficulty of developing effective interventions that perform well under extremely challenging conditions.

Study 3

Relying on the methodology of field experiments, Studies 1 and 2 produced results with high levels of ecological validity. However, the nature of these experiments restricted the types of data we could collect from participants, constraining what can be inferred about the psychological processes driving our results. It would have been infeasible, for example, to ask Russian email recipients to answer additional Likert-style survey questions about our treatment messages and what they found more or less compelling about each. As such, despite having strong evidence that one intervention (Treatment A) was markedly more effective than other interventions, our understanding of why remained incomplete.

As a complement to the two field experiments, we conducted Study 3 to probe for differences in psychological reactions to our treatment messages and examined whether these differences offered insights into why Treatment A outperformed the others. We recruited 553 participants living in the United States from the Prolific Academic platform. Participants read two randomly selected treatment messages and rated the extent to which each message relied on a variety of strategies that prior scholars have identified as psychological mechanisms behind effective influence attempts. We took this approach to identify whether lay people perceived particular persuasion strategies to be particularly prominent in the individual expert messages and whether these strategies correlated with Study 1 message effectiveness.

We identified potential strategies by drawing on the theoretical justifications that experts submitted along with their proposed interventions for Study 1, as well as based on our review of the broader research literature. We ultimately settled on the following eleven strategies: seeking

perspective, honest disclosure, communicating with warmth, conversational receptiveness, stoking curiosity, strategic affirmation, invoking social norms, highlighting accountability, highlighting similarity, brevity, and demonstrating cultural knowledge. The definitions of each potential mechanism and citations to relevant research studies for each can be found in Table 2.

An exploratory factor analysis revealed that participant ratings of these eleven potential persuasion mechanisms could be reduced to two interpretable factors, which we labelled “Highlighting Connectedness” and “Perceived Receptiveness.” The factor loadings for each persuasion mechanism on each factor are reported in Table 2. The central items for Highlighting Connectedness, the factor that explained relatively more overall variance across the eleven mechanisms, were demonstrating cultural knowledge, highlighting similarity, highlighting accountability, invoking social norms, and stoking curiosity. For Perceived Receptiveness, the central items were receptiveness (unsurprisingly), seeking perspective, and communicating with warmth. Strategic affirmation and honest disclosure loaded similarly on both factors; brevity did not load onto either.

As shown in Figures 3 and 4, despite accounting for relatively less overall variance, it is the Perceived Receptiveness factor that differentiates Treatment A from the other messages. Figure 3 presents the means of layperson ratings on the 0 to 4 scale for each treatment and each persuasion mechanism, with column-wise color coding. Blue indicates that subjects perceived the message to more extensively embody a given mechanism relative to other messages; red indicates the opposite. Treatment A stands out as scoring highly on the mechanisms most central to Perceived Receptiveness: It scores highest among all messages on conversational receptiveness and seeking perspective, and second highest for communicating with warmth. In contrast, Treatment A is far less differentiated on the potential psychological mechanisms central

Table 2: Potential Psychological Mechanisms

Psychological Mechanism	Relevant Studies	Factor Loading on Highlighting Connectedness	Factor Loading on Perceived Receptiveness
Seeking perspective means asking the recipient for their opinion, perspective, or insight about a topic. This can include showing that the writer wanted to know what the reader thinks, feels, or believes about a particular topic.	- Heltzel & Laurin (2021) - Eyal et al. (2018) - Goldstein et al. (2014)	-.07	.76
Honest disclosure means being transparent, clear, honest, and accurate about the intentions of the message.	- Cooper et al. (2023) - Shalvi et al. (2025)	.38	.31
Communicating with warmth refers to using a friendly communication style, rather than being cold and tough.	- Fiske (2018) - Jeong et al. (2019) - Eisenbruch et al. (2022)	.23	.54
Conversational receptiveness refers to the use of language to express willingness to thoughtfully engage with opposing views	- Yeomans et al. (2018) - Yeomans et al. (2020)	.01	.83
Stoking curiosity means leading the recipient to believe that they do not fully understand a situation or have all the relevant information to make them curious to learn more.	- Loewenstein (1994) - Wojtowicz & Loewenstein (2020)	.37	.08
Offering strategic affirmation means elevating the status of the message recipient by showing respect and engaging in flattery. Examples include affirming status or affirming the value of the recipient.	- Vonk (2002) - Anderson et al. (2015)	.43	.29
Invoking social norms refers to emphasizing how a desired behavior is a social norm. Social norms are common ways of behaving in a particular situation that are often widely accepted as appropriate and correct.	- Goldstein & Cialdini (2011) - Cialdini & Goldstein (2004)	.55	.23
Highlighting accountability means reminding the recipient that other people are attending to their actions and that their actions have important consequences.	- Lerner & Tetlock (1999) - Lerner et al. (1998)	.55	.07
Highlighting similarity means pointing out what the writer and the message recipient have in common. This could be common background, common experiences, common beliefs, or common goals.	- Cialdini & Goldstein (2004) - Simons et al. (1970)	.75	.02
Communicating with brevity refers to using a concise (vs. wordy) written style.	- Rogers & Lasky-Fink (2023) - Shulman et al. (2024)	-.03	.11
Demonstrating cultural knowledge means showing the message recipient that the writer is familiar with the Russian culture, language, and history.	- Agrawal (2015) - Huang & Shen (2016)	.86	-.11

Figure 3: Mean Ratings of Perceived Psychological Mechanisms by Treatment
 Color scaled within each column: blue = high, red = low (relative to other messages)

	Effect Size (pp)	Seeking Perspective	Honest Disclosure	Communicating with Warmth	Conversational Receptiveness	Stoking Curiosity	Strategic Affirmation	Invoking Social Norms	Highlighting Accountability	Highlighting Similarity	Brevity	Demonstrating Cultural Knowledge
Treatment A	0.21	3.44	2.89	2.86	2.74	2.43	2.39	2.11	1.71	1.46	1.44	1.33
Treatment B	0.03	2.07	2.52	2.46	2.17	2.52	2.70	2.14	2.05	3.07	1.60	3.42
Treatment C	0.01	2.71	2.01	2.05	2.22	1.94	3.25	1.90	2.02	1.72	1.51	1.96
Treatment D	-0.01	0.99	2.51	1.06	1.28	2.78	1.08	1.52	2.12	1.09	2.46	2.05
Treatment E	-0.01	3.14	2.92	2.23	2.49	2.17	1.83	1.90	1.77	1.37	2.25	1.52
Treatment F	-0.01	0.95	1.37	1.75	1.22	1.47	1.06	1.41	0.65	0.55	2.98	0.41
Treatment G	-0.05	2.10	2.97	3.08	2.49	2.59	2.44	2.30	2.02	2.84	1.53	2.66
Treatment H	-0.12	1.94	2.27	1.86	1.75	2.58	2.38	2.38	3.17	2.13	1.55	2.50
Treatment I	-0.13	2.91	2.43	2.17	2.35	2.38	2.25	1.95	2.57	1.81	2.05	1.79
Treatment J	-0.16	2.32	1.92	2.12	2.25	3.09	1.85	1.74	1.43	1.89	2.26	1.17
Treatment K	-0.19	2.41	1.96	2.15	1.98	2.69	1.57	1.87	1.53	1.94	2.26	1.83

Note: The leftmost column shows treatment effects relative to the link-only control, measured in percentage points, from the Study 1 field experiment. All other columns display mean layperson ratings from Study 3 assessing the extent to which a given message relied on a given persuasion mechanism, on a scale from 0 (“not at all”) to 4 (“very”). Color coding is scaled within each column (i.e., relative to other messages), with darker blue indicating greater perceived use of a given mechanism and darker red indicating lesser perceived use.

to Highlighting Connectedness. For example, it scores third lowest among all messages for demonstrating cultural knowledge, the highest loading item on the Highlighting Connectedness factor.

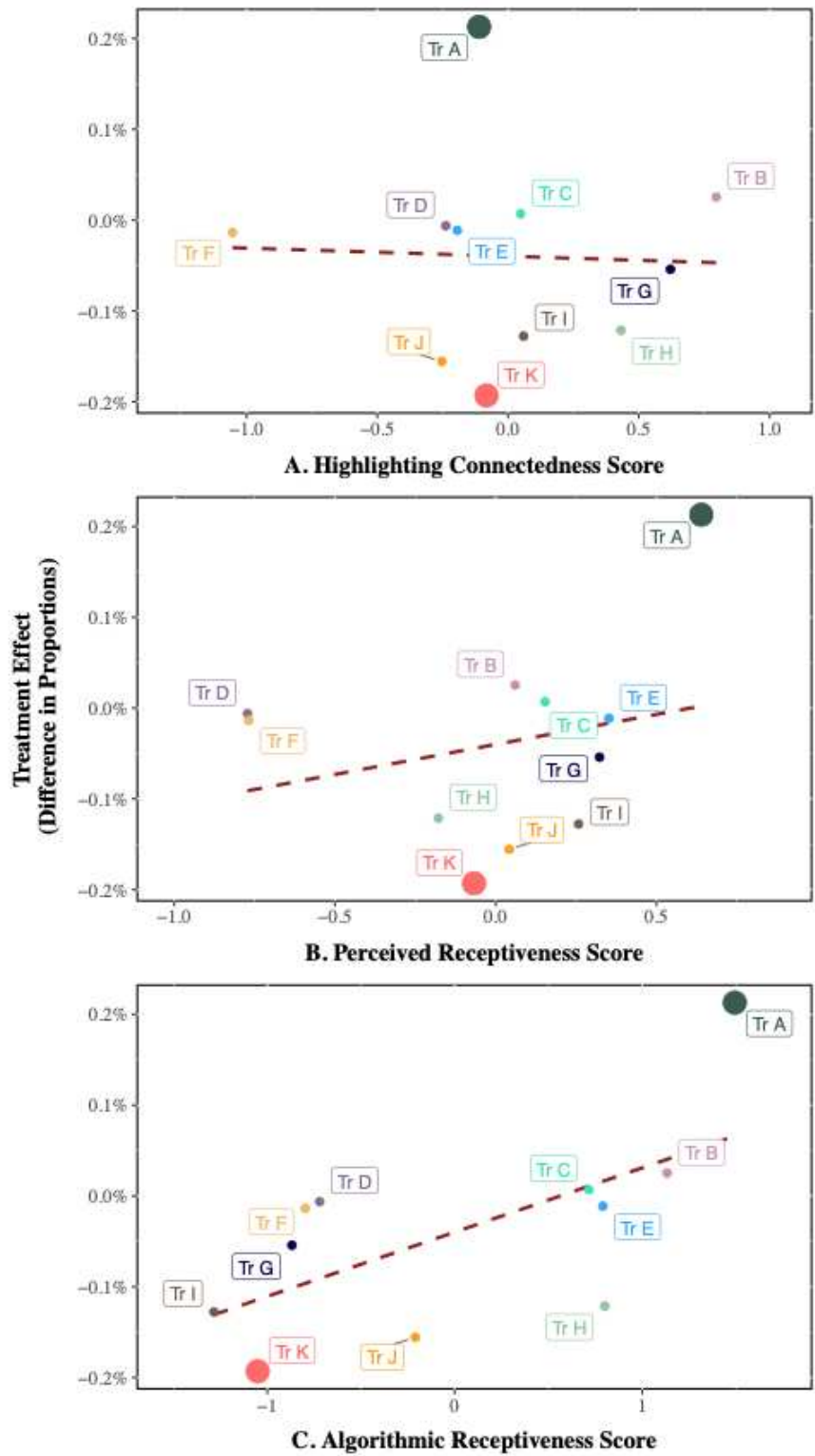
Figure 4 extends our analysis by showing that Treatment A’s greater Perceived Receptiveness, relative to other messages, is positively associated with its higher effectiveness. Panels A and B plot the factor scores (*x*-axis) for Highlighting Connectedness and Perceived Receptiveness, respectively, against the treatment effects of the Study 1 field experiment (*y*-axis). The nearly flat OLS regression line in Panel A indicates no meaningful relationship between Highlighting Connectedness scores and effect sizes ($r = -.04, p = .91$), though, given the limited number of data points ($n = 11$) on which these analyses draw, our results should be seen as

exploratory. By contrast, the upward sloping regression line in Panel B indicates that, on average, messages with higher Perceived Receptiveness scores achieved larger effect sizes ($r = .26, p = .44$).

The results above suggest that receptiveness may be a key factor explaining Treatment A's relative effectiveness. However, the Perceived Receptiveness factor scores rely on subjective layperson ratings. As a complementary approach, we therefore employed an established algorithmic measure of conversational receptiveness (Minson et al., 2024; Yeomans et al., 2018, 2022). This validated algorithmic approach measures the level of conversational receptiveness by the presence or absence of specific words and phrases (called "features") that the recipients of a communication experience as signaling engagement with their point of view. Use of positive features (e.g., acknowledgement, hedging) increases the algorithmic receptiveness score of a piece of text, whereas use of negative features (e.g., negations, reasonings) decreases it. We term the scores outputted by this algorithm "algorithmic receptiveness."

As depicted in Panel C of Figure 4, Treatment A produced a higher algorithmic receptiveness score than any of the other interventions, as indicated by its position on the x -axis. This pattern mirrors Treatment A's differentiation in Figure 3 on the mechanisms that load most strongly onto the Perceived Receptiveness factor. Moreover, just as Panel B of Figure 4 established a positive relationship between Perceived Receptiveness and effect sizes, Panel C shows that algorithmic receptiveness scores are meaningfully predictive of the treatment effects from Study 1 despite our extremely limited sample size of eleven interventions ($r = .63, p = .036$).

Figure 4: Treatment Effects by Factor Scores & Algorithmic Receptiveness Scores



Note: Figures show Study 1 treatment effects relative to the link-only control, measured in percentage points, plotted against each message's Highlighting Connectedness, Perceived Receptiveness, and Algorithmic Receptiveness scores. Bolded dots indicate the statistical significance of treatment effects after adjusting for multiple comparisons (Benjamini & Hochberg, 1995). Trend lines computed from a linear model regressing treatment effects on scores.

Though exploratory, the results are striking, particularly in the convergence of findings from both subjective layperson ratings and objective algorithmic ratings (Figure 4, Panels B and C, respectively). Together, these findings suggest that the principles of conversational receptiveness and related psychological mechanisms may offer a promising starting point for the design of messaging in challenging conditions, a point we return to in the Discussion section.

Discussion

Russia's 2022 invasion of Ukraine demonstrated the extent to which technology is changing information warfare. Citizens from anywhere in the world can establish direct contact with each other, even under conditions of significant repression and censorship. As Russia, China, and other non-democratic regimes seek to reshape the global order, the extent to which citizen-led grassroots campaigns can effectively counter authoritarian censorship and propaganda will remain a pressing question. These campaigns also present an extreme test for behavioral science theories of persuasion. Generating engagement with information that contradicts deeply held beliefs is difficult even under ideal conditions. Generating such engagement under wartime conditions in authoritarian regimes, often via suboptimal modes of communication such as unsolicited text messages from unfamiliar senders, constitutes an extraordinary challenge. Moreover, although the behavioral sciences have in recent years produced numerous theoretically grounded interventions for reducing selective exposure and information avoidance, empirical evaluations of these interventions under challenging circumstances have remained relatively scarce.

By contrast, our mega-study examining the effectiveness of eleven treatments designed by expert academics was implemented via two large-scale pre-registered field experiments on the effectiveness of citizen-to-citizen email communication in wartime Russia. Overall, our findings

offer both sobering insights and grounds for optimism regarding the potential for citizen-led communication campaigns to counter authoritarian censorship and propaganda, and each of our three main results has important implications for the study of information dissemination.

First, the scalability of citizen-to-citizen campaigns' information technology makes these campaigns valuable, exposing large audiences to uncensored information. However, the fall in click rates and increased difficulty in delivering emails between our first and second experiments suggests that campaigns' effectiveness may decline over time, possibly due to evolving censorship strategies or the hardening of citizen attitudes. Although more research is needed, it may be the case that citizen-to-citizen campaigns are more valuable in early stages of a conflict.

Second, the success of our top-performing intervention provides compelling evidence that, even under very challenging conditions, persuasion is possible – and that behavioral science can inform more effective messaging strategies. This treatment, which drew on theories highlighting the importance of conveying curiosity, respect, warmth, and receptiveness to others' viewpoints (Collins et al., 2022; Lerner & Tetlock, 1999; Minson & Chen, 2022; Shapiro, 2017), approximately doubled click rates relative to the control conditions. Yet, ten of the eleven treatments in our study failed to outperform controls, even though they too were based on relevant theories and crafted by expert academics. Indeed, one treatment appears to have backfired. In line with O'Keefe (2018), these findings draw attention to how essential it is for influence campaigns to employ the experimental method to rigorously evaluate messages' effectiveness, rather than relying on intuition, even the intuition of behavioral science or regional experts.

Third, our findings suggest that in the challenging and highly sensitive context of wartime censorship and repression, thoughtful engagement with opposing perspectives may be more

effective than efforts to directly persuade through facts and argumentation. This finding aligns with earlier work on conversational receptiveness (Minson & Chen, 2022; Yeomans, et al. 2022). Building on our exploratory analyses, future research confirming and expanding on these striking patterns would likely be highly productive.

Our studies are not, of course, without limitations. Given our focus on a context in which empirical research is almost nonexistent, we employed many multi-pronged interventions. An important goal of future research should be to examine more comprehensively the effects of individual psychological mechanisms by testing individual intervention components.

Additionally, as discussed in the Method section, our reliance on YouTube for hosting the videos limited access to various types of data that could be useful. Although we justifiably chose to use YouTube because many Russians at the time of our studies used and trusted the platform, future research approaches that allow for collection of additional data likely would prove fruitful.

Overall, our studies demonstrate that even in an extremely challenging setting, carefully designed interventions can meaningfully increase engagement with uncensored information. Moreover, our results show that rigorous evaluation of citizen-led grassroots communication campaigns aimed at countering authoritarian censorship is feasible – and that more of such experimental research is urgently needed.

Method

Our two field experiments were made possible through a collaboration with the non-profit Mail2Ru, a Norwegian-based, semi-automated email platform founded in March 2022 to direct Russians toward uncensored information. Mail2Ru leverages an ever-growing database of over 350,000,000 Russian email addresses scraped from publicly available sources to enable global volunteers to send messages to Russian recipients. A typical message from Mail2Ru contains text

in Russian urging the recipient to learn about the war and providing a set of hyperlinks to reliable Western news sources (e.g., the Russian-language service of the BBC). For the purposes of our research, we used the Mail2Ru infrastructure to randomly assign recipients to messages authored by researchers and then tracked engagement.

Open science and research ethics. We pre-registered our procedure, sample size, power analyses, and analysis plan before data collection. Our pre-registrations, materials, and code can be found [here](#). Deviations from pre-registrations, largely necessitated by technical considerations, are described in the Supplementary Materials ([Appendix 5](#)). This research was reviewed by Harvard University IRB, with great additional care taken to avoid endangering participants, given the research context (see [Appendix 6](#)). We additionally emphasize that Mail2Ru's core mission is sending emails with uncensored information into Russia, and email recipients in our studies would have received messages with or without our involvement. Our experiments affected only the content and distribution procedure for some of these emails.

Given the authoritarian political environment in contemporary Russia, it would have been unsafe and impractical to involve local researchers in this project. Our research team did, however, include three Russian social scientists who are based at US universities. Likewise, because of Russia's political environment, and considering the nature of our research, it was not possible nor advisable to seek approval from local Russian institutions. This approach was approved in advance by the Harvard University IRB. More broadly, approval from western-style institutional review boards is not a widely employed practice for social science in Russia.

Expert-designed messages. We contacted over 100 academics in diverse disciplines and described the planned experiment (Hameiri & Moore-Berg, 2022) (see [Appendix 2](#)). We reached out to scholars in psychology, conflict management and negotiations, political science, and

regional studies. Given the sensitivity of the project, we did not publish an open call for submissions, instead relying on “word of mouth” transmission among colleagues.

We solicited non-deceptive messages under 300 words long in English, intended to persuade a Russian reader to view an uncensored video about the invasion. We described the measures that would be collected and explained that messages would be selected by the core project team blind to authorship based on anticipated effectiveness and theoretical rigor. We did not share the video with the message authors, so as to encourage them to produce messages that would be broadly effective, rather than focused on a specific video. The members of the core project team ranked the submissions independently and then, through discussion, made final selections. We intended to evaluate 10 interventions, but 11 were ultimately included due to ties in our selection procedure ([Appendix 2](#)). The chosen messages represented a variety of disciplinary fields and theoretical approaches, with several tapping deep knowledge of Russian culture and history.

Control messages. For Study 1, we employed two control conditions. To construct the first, we collected 100 messages from laypeople on Prolific using analogous instructions to those sent to the academics. After eliminating three messages that did not conform to our pre-registered criteria, we randomly selected 10. We then randomized which of these 10 messages was seen by each email recipient assigned to our layperson control condition ([Appendix 3](#)). Comparison to this condition enabled us to test which expert messages would outperform seemingly sensible lay messages. Our second control condition (the “link-only” control) featured a message with no text at all, containing only a hyperlink to the video. We considered this condition to be a minimal bar for any persuasive message to outperform since most computer-savvy email recipients should be wary of clicking on an unexplained link from an unknown sender.

Message construction. We used a professional translator to translate all messages from English to Russian. We then had the Russian versions of the messages professionally edited by a former Russian journalist familiar with psychological research to ensure that the translation matched both the literal meaning and the psychological intention of the English originals.

All emails were sent with the same subject header in Russian: “Important information about ongoing events.” All messages contained a link to the same video. The video was 92 seconds long and described the destruction of the city of Bakhmut. In the video (available [here](#)), a camera slowly pans over an arial view of Bakhmut, before and after the 8 months-long battle for the city. The video is set to instrumental music and features minimal text, presented in black boxes on the screen. We chose this video after reviewing many others for several reasons: First, while clearly portraying the costs of war, the video contains no images of graphic violence that an unsuspecting email recipient might find disturbing. Second, because the video does not feature specific journalists or protagonists, we were less concerned about results being driven by a particular set of people appearing to be especially trustworthy or sympathetic. Finally, because this video was produced by a Russian opposition news agency (Meduza.io), there is no English text that might have served to distract Russian-speaking viewers or heighten distrust.

Dependent variable. For reasons described in Study 1, our main pre-registered dependent variable was the rate at which participants clicked on the video link, after accounting for undelivered (i.e., “bounced”) emails. As also noted in Study 1, to the extent possible, we collected data on viewer duration to confirm that clicks represent substantive engagement with the video. However, YouTube analytics do not provide individual-level view durations associated with each click and, more broadly, we sought to employ data aggregated at the level of experimental arms to preserve email recipients’ safety and anonymity. Although these data

limitations prevent us from conducting rigorous comparisons of viewer duration across treatment arms, YouTube’s aggregated statistics provide enough information to confirm that for ten of our eleven treatments – including the top-performing Treatment A – participants who clicked viewed at least thirty seconds of the video on average, indicating a meaningful level of engagement. Finally, as potential secondary measures, we collected data on comments and likes posted by viewers of the video, but these were far too few to be analyzable.

Experimental design. We hosted the video on a private YouTube channel and created unique links associated with each treatment in order to track the clicks associated with each experimental condition. Study 1 featured 13 total conditions: 11 comprised of messages written by academics, 1 control condition comprised of 10 messages randomly drawn from the pool written by lay people (for each email recipient in this condition, 1 of these 10 messages was randomly assigned), and 1 “link-only” condition. Study 2 was comprised of two arms: the top-performing expert message from Study 1 and the link-only control condition. To avoid deception while also varying sender names, we assigned email sender names based on Russian translations of first names drawn from members of the research team.²

Sending procedure: We conducted Study 1 between July 16 and August 26, 2023. We sent each of the 11 expert messages to approximately 20,000 email addresses ([Appendix 4](#) presents power analyses and shows our studies were sufficiently powered). The lay messages were also sent to 20,000 addresses total, with the ten individual messages sent to 2,000 addresses each.

² To consider the effects of gender, we employed three male and three female sender names. Accordingly, Study 1 ultimately featured 26 experimental arms – 13 conditions crossed with two genders of email senders – with each arm assigned to a unique YouTube link. For all but one treatment message, click rates were higher for emails sent from accounts with female names, but these differences were not statistically significant. At the start of the project, we had also intended to evaluate the effect of recipient gender by inferring gender from email addresses. However, after observing lower-than-expected delivery rates and a large proportion of email addresses with no interpretable gender markers, we collapsed our data across recipient genders (see [Appendix 5](#)).

Finally, the link-only control was sent to an additional 20,000 addresses. In total, 260,088 emails were sent, and 233,132 were confirmed to be delivered.

Study 2 was conducted February 16-18, 2024. Increased concerns about filters and censorship required a different distribution approach, and our pre-analysis plan called for sending a minimum of 20,000 messages for each of the two experimental arms or to continue sending until blocked by Russian censors. In all, 148,141 emails were sent, and 72,159 were determined to have been delivered.

Analysis of mechanisms (Study 3): We recruited 535 participants living in the United States from Prolific Academic ($M_{\text{age}} = 41$, 51% Female). We conducted Study 3 in June of 2025. We briefly described the earlier field studies to the participants and asked them to rate the extent to which a given expert message relied on one of the persuasion strategies described in Table 2. To avoid participant fatigue, each participant evaluated the persuasion strategies present in two of the expert messages used in the field study. Participants rated the extent to which any given expert message leveraged each persuasion strategy using a scale from 0 (“Not at All”) to 4 (“Very”). The eleven different potential persuasion strategies were presented in a random order. After evaluating two randomly selected messages, the participants reported their gender and age.

We conducted an exploratory factor analysis on these ratings to more formally assess the extent to which the eleven ratings could be reduced to interpretable dimensions. A scree plot revealed that three potential factors achieved higher eigenvalues than would be expected from simulated or resampled data. The first factor attained an eigenvalue of almost four; the second factor attained an eigenvalue of almost one; the third factor attained an eigenvalue of almost 0.5.

We began by conducting an exploratory factor analysis with a three-factor solution. We used oblique rotation because we expected the psychological mechanisms identified by scholars to

correlate with each other. However, an analysis of the three-factor solution revealed that the third factor was not readily interpretable, including a mix of brevity, honest disclosure, and strategic affirmation (negatively coded). Given our goal of interpretability, we chose to proceed with two factors, which were readily interpretable (as described in the results section above). We did not consider a one-factor solution given our goal of differentiating distinct potential psychological mechanisms.

References

- Agrawal, N. I. D. H. I. (2015). Culture and persuasion. *Handbook of culture and consumer behavior*, 121-134.
- Aidis, R., Estrin, S., & Mickiewicz, T. (2008). Institutions and entrepreneurship development in Russia: A comparative perspective. *Journal of Business Venturing*, 23(6), 656-672.
- Alyukov, M. (2022). Propaganda, authoritarianism and Russia's invasion of Ukraine. *Nature Human Behaviour*, 6(6), 763-765.
- Anderson, C., Hildreth, J. A. D., & Howland, L. (2015). Is the desire for status a fundamental human motive? A review of the empirical literature. *Psychological bulletin*, 141(3), 574.
- Badea, C., & Sherman, D. K. (2019). Self-Affirmation and Prejudice Reduction: When and Why? *Current Directions in Psychological Science*, 28(1), 40–46.
<https://doi.org/10.1177/0963721418807705>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), 289-300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Chatterjee, P., Hoffman, D. L., & Novak, T. P. (2003). Modeling the clickstream: Implications for web-based advertising efforts. *Marketing Science*, 22(4), 520–541.
<https://doi.org/10.1287/mksc.22.4.520.24906>
- Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annu. Rev. Psychol.*, 55(1), 591-621.
- Collins, H. K., Dorison, C. A., Gino, F., & Minson, J. A. (2022). Underestimating counterparts' learning goals impairs conflictual conversations. *Psychological Science*, 33(10), 1732-1752.
<https://doi.org/10.1177/09567976221085494>

- Cooper, B., Cohen, T. R., Huppert, E., Levine, E. E., & Fleeson, W. (2023). Honest behavior: Truth-seeking, belief-speaking, and fostering understanding of the truth in others. *Academy of Management Annals*, *17*(2), 655-683.
- Dean, M. C., & Porter, B. (2024). Sentiment Analysis of Russian-Language Social Media Posts Discussing the 2022 Russian Invasion of Ukraine. *Armed Forces & Society*, 0095327X241235987.
- Eisenbruch, A. B., & Krasnow, M. M. (2022). Why warmth matters more than competence: A new evolutionary approach. *Perspectives on Psychological Science*, *17*(6), 1604-1623.
- Erlich, A., & Garner, C. (2021). Subgroup differences in implicit associations and explicit attitudes during wartime. *International Studies Quarterly*, *65*(2), 528-541.
- Eyal, T., Steffel, M., & Epley, N. (2018). Perspective mistaking: Accurately understanding the mind of another requires getting perspective, not taking perspective. *Journal of personality and social psychology*, *114*(4), 547.
- Feinberg, M., & Willer, R. (2019). Moral reframing: A technique for effective and persuasive communication across political divides. *Social and Personality Psychology Compass*, *13*(12), e12501.
- Fiske, S. T. (2018). Stereotype content: Warmth and competence endure. *Current directions in psychological science*, *27*(2), 67-73.
- Fong, A., Roozenbeek, J., Goldwert, D., Rathje, S., & van der Linden, S. (2021). The language of conspiracy: A psychological analysis of speech used by conspiracy theorists and their followers on Twitter. *Group Processes & Intergroup Relations*, *24*(4), 606–623.
<https://doi.org/10.1177/1368430220987596>

- Frey, D. (1986). Recent research on selective exposure to information. In Berkowitz L. (Ed.), *Advances in Experimental Social Psychology* (19, pp. 41–80). New York, NY: Academic Press.
- Gandhi, J. (2008). *Political Institutions Under Dictatorship*. New York, NY: Cambridge University Press.
- Goldenberg, A., Cohen-Chen, S., Goyer, J. P., Dweck, C. S., Gross, J. J., & Halperin, E. (2018). Testing the impact and durability of a group malleability intervention in the context of the Israeli–Palestinian conflict. *Proceedings of the national academy of sciences*, 115(4), 696–701.
- Goldstein, N. J., & Cialdini, R. B. (2011). Using social norms as a lever of social influence. In *The science of social influence* (pp. 167-191). Psychology Press.
- Goldstein, N. J., Vezich, I. S., & Shapiro, J. R. (2014). Perceived perspective taking: when others walk in our shoes. *Journal of personality and social psychology*, 106(6), 941.
- Grant, A. M., & Gino, F. (2010). A little thanks goes a long way: Explaining why gratitude expressions motivate prosocial behavior. *Journal of Personality and Social Psychology*, 98(6), 946–955. <https://doi.org/10.1037/a0017935>
- Greene, S. A., and Robertson, G. B. (2019). *Putin v. the People: The Perilous Politics of a Divided Russia*. New Haven, CT: Yale University Press.
- Gulati, R. (1995). Social structure and alliance formation patterns: A longitudinal analysis. *Administrative science quarterly*, 619-652.
- Hameiri, B., & Moore-Berg, S. L. (2022). Intervention Tournaments: An Overview of Concept, Design, and Implementation. *Perspectives on Psychological Science*, 17(6), 1525-1540. <https://doi.org/10.1177/17456916211058090>

- Hart, W., Albarracín, D., Eagly, A. H., Brechan, I., Lindberg, M. J., & Merrill, L. (2009). Feeling validated versus being correct: A meta-analysis of selective exposure to information. *Psychological Bulletin*, 135(4), 555–588. <https://doi.org/10.1037/a0015701>
- Harwell, D. (2022, March 17). Computer programmers are taking aim at Russia’s propaganda wall. *The Washington Post*. Retrieved from <https://www.washingtonpost.com/technology/2022/03/17/countering-russian-propaganda-efforts-ukraine/>
- Hasan-Aslih, S., Shuman, E., Pliskin, R., van Zomeren, M., Saguy, T., & Halperin, E. (2020). With or without you: The paradoxical role of identification in predicting joint and ingroup collective action in intergroup conflict. *European Journal of Social Psychology*, 50(6), 1334-1343.
- Heltzel, G., & Laurin, K. (2021). Seek and ye shall be fine: Attitudes toward political-perspective seekers. *Psychological Science*, 32(11), 1782-1800.
- Howe, L. C., Carr, P. B., & Walton, G. M. (2021). Normative appeals motivate people to contribute to collective action problems more when they invite people to work together toward a common goal. *Journal of Personality and Social Psychology*, 121(2), 215.
- Huang, Y., & Shen, F. (2016). Effects of cultural tailoring on persuasion in cancer communication: A meta-analysis. *Journal of Communication*, 66(4), 694-715.
- Jeong, M., Minson, J., Yeomans, M., & Gino, F. (2019). Communicating with warmth in distributive negotiations is surprisingly counterproductive. *Management Science*, 65(12), 5813-5837.
- Kardas, M., Schroeder, J., & O’Brien, E. (2022). Keep talking: (Mis)understanding the hedonic trajectory of conversation. *Journal of Personality and Social Psychology*, 123, 717–740.

- Kirgios, E.L., Athey, S., Duckworth, A.L., Karlan, D., Luca, M., Milkman, K.L., & Offer-Westort, M. Does Q&A Boost Engagement? Health Messaging Experiments in the United States and Ghana. *Management Science*. Forthcoming.
- Kossowska, M., Szwed, P., Szumowska, E., Perek-Białas, J., & Czernatowicz-Kukuczka, A. (2023). The role of fear, closeness, and norms in shaping help towards war refugees. *Scientific Reports*, *13*(1), 1465.
- Lerner, J. S., Goldberg, J. H., & Tetlock, P. E. (1998). Sober second thought: The effects of accountability, anger, and authoritarianism on attributions of responsibility. *Personality and Social Psychology Bulletin*, *24*(6), 563-574.
- Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, *125*(2), 255.
- Lizzio-Wilson, M., Thomas, E. F., Louis, W. R., Wilcockson, B., Amiot, C. E., Moghaddam, F. M., & McGarty, C. (2021). How collective-action failure shapes group heterogeneity and engagement in conventional and radical action over time. *Psychological Science*, *32*(4), 519-535.
- Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin*, *116*(1), 75–98. <https://doi.org/10.1037/0033-2909.116.1.75>
- Manchanda, P., Dubé, J. P., Goh, K. Y., & Chintagunta, P. K. (2006). The effect of banner advertising on internet purchasing. *Journal of Marketing Research*, *43*(1), 98–108. <https://doi.org/10.1509/jmkr.43.1.98>
- McDonald, R. I., Chai, H. Y., & Newell, B. R. (2015). Personal experience and the ‘psychological distance’ of climate change: An integrative review. *Journal of environmental psychology*, *44*, 109-118.

- McGinnies, E., & Ward, C. D. (1980). Better liked than right: Trustworthiness and expertise as factors in credibility. *Personality and Social Psychology Bulletin*, 6(3), 467-472.
- Milkman, K. L., Gromet, D., Ho, H., Kay, J. S., Lee, T. W., Pandiloski, P., Park, Y., Rai, A., Bazerman, M., Beshears, J., Bonacorsi, L., Camerer, C., Chang, E., Chapman, G., Cialdini, R., Dai, H., Eskreis-Winkler, L., Fishbach, A., Gross, J. J., ... Duckworth, A. L. (2021). Megastudies improve the impact of applied behavioural science. *Nature*, 600(7889), 478–483. <https://doi.org/10.1038/s41586-021-04128-4>
- Minson, J. A., Chen, F. S., & Tinsley, C. H. (2020). Why Won't You Listen to Me? Measuring Receptiveness to Opposing Views. *Management Science*, 66(7), Article 7. <https://doi.org/10.1287/mnsc.2019.3362>
- Minson, J. A., & Chen, F. S. (2022). Receptiveness to opposing views: Conceptualization and integrative review. *Personality and Social Psychology Review*, 26(2), 93-111. <https://doi.org/10.1177/10888683211061037>
- Moore, M., Dorison, C. A., & Minson, J. A. (2023). The contingent reputational benefits of selective exposure to partisan information. *Journal of Experimental Psychology: General*, 152(12), 3490–3525. <https://doi.org/10.1037/xge0001463>
- Mormol, P. (2019). 'i urge you to see this...'. clickbait as one of the dominant features of contemporary online headlines. *Social Communication*, 5(2), 1-10.
- Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, 32(2), 303-330.
- O'Keefe, D. J. (2018). Message pretesting using assessments of expected or perceived persuasiveness: Evidence about diagnosticity of relative actual persuasiveness. *Journal of Communication*, 68(1), 120-142.

- Ott, B. L. (2017). The age of Twitter: Donald J. Trump and the politics of debasement. *Critical studies in media communication*, 34(1), 59-68.
- Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., & Rand, D. G. (2020). Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. *Psychological science*, 31(7), 770-780.
- Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*, 592(7855), 590-595. <https://doi.org/10.1038/s41586-021-03344-2>
- Rebane, T. (2022, April 2). Cold calling in the Iron Curtain: How Europe is engaging with Russia and Ukraine. *CNN*. Retrieved from <https://www.cnn.com/2022/04/02/europe/cold-calling-ukraine-russia-iron-curtain-intl-hnk-dst/index.html>
- Renn, O., & Levine, D. (1991). *Credibility and trust in risk communication* (pp. 175-217). Springer Netherlands.
- Roberts, R., & Risen, J. L. (2022). Introducing conspiracy intuitions to better understand conspiracy beliefs. *Current Opinion in Psychology*, 47, 101395. <https://doi.org/10.1016/j.copsy.2022.101395>
- Roehm, M. L., & Sternthal, B. (2001). The moderating effect of knowledge and resources on the persuasive impact of analogies. *Journal of Consumer Research*, 28(2), 257-272.
- Rogers, T., & Lasky-Fink, J. (2023). *Writing for Busy Readers: Communicate More Effectively in the Real World*. Penguin.
- Sears, D. O., & Freedman, J. L. (1967). Selective Exposure to Information: A Critical Review. *Public Opinion Quarterly*, 31(2), 194. <https://doi.org/10.1086/267513>

- Shalvi, S., Levine, E. E., Thielmann, I., Jayawickreme, E., Van Rooij, B., Teodorescu, K., ... & Ritov, I. (2025). The science of honesty: A review and research agenda. *Advances in experimental social psychology*, 72, 241-327.
- Shapiro, D. (2017). *Negotiating the nonnegotiable: How to resolve your most emotionally charged conflicts*. Penguin.
- Sharafutdinova, G. (2020). *The red Mirror: Putin's leadership and Russia's insecure identity*. Oxford University Press.
- Shulman, H. C., Markowitz, D. M., & Rogers, T. (2024). Reading dies in complexity: Online news consumers prefer simple writing. *Science Advances*, 10(23), eadn2555.
- Shuman, E., Goldenberg, A., Saguy, T., Halperin, E., & van Zomeren, M. (2023). When are social protests effective?. *Trends in Cognitive Sciences*, 28(3), 252-263.
- Simons, H. W., Berkowitz, N. N., & Moyer, R. J. (1970). Similarity, credibility, and attitude change: A review and a theory. *Psychological bulletin*, 73(1), 1.
- Sparkman, G., & Walton, G. M. (2017). Dynamic norms promote sustainable behavior, even if it is counternormative. *Psychological science*, 28(11), 1663-1674.
- Stepanova, E. A. (2022). “Everything good against everything bad”: traditional values in the search for new Russian national idea. *Zeitschrift für Religion, Gesellschaft und Politik*, 7(1), 97-118.
- Stokel-Walker, C. (2023, February 23). Researchers show how activists use online reviews within Russia to protest the Ukraine invasion. *Fast Company*. Retrieved from <https://www.fastcompany.com/90853673/researchers-show-how-activists-use-online-reviews-within-russia-to-protest-the-ukraine-invasion>

- Svolik, M.W. (2012). *The Politics of Authoritarian Rule*. New York, NY: Cambridge University Press.
- Szasz, F. M. (2009). “Pamphlets away”: The Allied propaganda campaign over Japan during the last months of World War II. *Journal of Popular Culture*, 42(3).
- Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American journal of political science*, 50(3), 755-769.
- Tebor, C. (2022, March 9). Russia increases censorship with new law: 15 years in jail for calling Ukraine invasion a “war.” *USA Today*. Retrieved from <https://www.usatoday.com/story/news/world/ukraine/2022/03/08/russia-free-speech-press-criminalization-misinformation/9433112002/>
- Tertychnaya, K., & Tiratsoo, M. (2024). Legal Repression in Russia. In *Oxford Research Encyclopedia of Politics*.
- Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin.
- Voelkel, J. G., Chu, J. Y., Stagnaro, M. N., Druckman, J. N., & Willer, R. (2024). How to design and conduct a megastudy. *Nature Human Behaviour* 8, 2257–2260.
- Wittenberg, C., Tappin, B. M., Berinsky, A. J., & Rand, D. G. (2021). The (minimal) persuasive advantage of political video over text. *Proceedings of the National Academy of Sciences*, 118(47), e2114388118.
- Vonk, R. (2002). Self-serving interpretations of flattery: Why ingratiation works. *Journal of personality and social psychology*, 82(4), 515.
- Wojtowicz, Z., & Loewenstein, G. (2020). Curiosity and the economics of attention. *Current Opinion in Behavioral Sciences*, 35, 135-140.

- Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 73(1), 3-36.
- Yablokov, I. (2018). *Fortress Russia: Conspiracy theories in the post-Soviet world*. John Wiley & Sons.
- Yeomans, M., Kantor, A., & Tingley, D. (2018). The politeness Package: Detecting Politeness in Natural Language. *The R Journal*, 10(2), 489. <https://doi.org/10.32614/RJ-2018-079>
- Yeomans, M., Minson, J., Collins, H., Chen, F., & Gino, F. (2020). Conversational receptiveness: Improving engagement with opposing views. *Organizational Behavior and Human Decision Processes*, 160, 131–148. <https://doi.org/10.1016/j.obhdp.2020.03.011>
- Zannettou, S., Sirivianos, M., Blackburn, J., & Kourtellis, N. (2019). The web of false information: Rumors, fake news, hoaxes, clickbait, and various other shenanigans. *Journal of Data and Information Quality (JDIQ)*, 11(3), 1-37.
- Zhelkina, A. (2020). The apathy syndrome: How we are trained not to care about politics. *Social Problems*, 67(2), 358-378.

Appendix 1 – Text of expert treatment messages and theoretical justifications submitted by authors

Treatment A	
<p>Hello! My name is XXX. Please forgive me for writing out of the blue. I saw online that you live in Russia, and I thought you might be able to help me and my colleagues understand the Ukraine-Russian conflict better.</p> <p>I'm part of a group of university students and faculty from different countries around the world who want to learn more about the conflict by hearing from actual citizens rather than from major media. You are in a better position than us to know the first-hand realities of what is going on.</p> <p>We're curious to know your opinion about the news clip below, especially about how it ends; it was suggested by a researcher originally from Russia. Apparently, thousands of people have watched it, so you may have already seen it. We're eager to hear your opinion so we can learn what is really going on.</p> <p>Click here to watch the news clip [LINK] (free on YouTube; watch as many times as you like)</p> <p>=> To help us learn your views, please add a comment on YouTube. My colleagues and I will check there to learn what you (and others in Russia) think about the clip.</p> <p>We really appreciate your help, and we look forward to reading the comments! Thank you in advance for taking the time to watch the video and share your thoughts with us.</p> <p>Your input will be invaluable in helping us (and hopefully others) understand this complex issue.</p>	<p>We drew on a variety of behavioral science articles and books to design our message. At the broadest level, we drew on well-documented insights reviewed in Thaler and Sunstein's (2009) book, <i>Nudge</i>. For example, we activated a descriptive norm when writing that thousands of others are viewing the video. Drawing on Collins et al. (2022), we explicitly stated our learning goals, which might otherwise be underestimated by recipients. Drawing on Minson and Chen (2022) we explicitly conveyed that we would be receptive to their views. Drawing on Shapiro's five-factor model (2017), developed in the context of international negotiation, we emphasized respect for the recipient, emphasized their expertise, and reinforced their autonomy. We strongly conveyed gratitude, drawing on Grant and Gino (2010). Finally, we introduced the kind of accountability most likely to trigger unbiased, systematic responses (Lerner & Tetlock, 1999) by telling them that we would check YouTube for their responses.</p> <p style="text-align: center;">(partial list of references)</p> <p>Collins, H. K., Dorison, C. A., Gino, F., & Minson, J. A. (2022). <i>Psychological Science</i>, 33(10), 1732-1752.</p> <p>Grant, A. M., & Gino, F. (2010). A little thanks goes a long way: Explaining why gratitude expressions motivate prosocial behavior. <i>Journal of personality and social psychology</i>, 98(6), 946.</p> <p>Lerner, J. S., & Tetlock, P. E. (1999). <i>Psychological Bulletin</i>, 125(2), 255-275.</p> <p>Minson, J. A., & Chen, F. S. (2022). <i>Personality and Social Psychology Review</i>, 26(2), 93-111.</p> <p>Shapiro, D. (2017). <i>Negotiating the nonnegotiable: How to resolve your most emotionally charged conflicts</i>. Penguin.</p> <p>Thaler, R. H., & Sunstein, C. R. (2009). <i>Nudge: Improving decisions about health, wealth, and happiness</i>. Penguin.</p>

Treatment B	
<p>Please read this message below that I am forwarding from a friend:</p> <p>My name is Zeldin, I am a 24-year-old student involved in a project offering Russians some new information about the situation in Ukraine. But I'm mostly writing to you today as a fellow Russian. My family moved to Israel after centuries in Russia. My roots, heritage, my babushka's Borscht – are all part of my identity. A part I am very proud of.</p> <p>From Mikhail Botvinnik to Anatoly Karpov, Garry Kasparov, and Vladimir Kramnik, we Russians are the greatest chess players on Earth. We're smart, strong, and cunning. We have beaten Nazi Germany in WW2 by sacrificing so many precious pawns and thus luring the enemy in.</p> <p>My culture is Russian, and I know for a fact that when Russians play the strategic long-game – they always win. But to win, we need to see the whole picture ourselves. If we ignore it, instead of winning at chess – we will find ourselves losing at checkers.</p> <p>Yet I also know that Russians are smart enough to re-evaluate, and strong enough to be open to new information, about what is really happening in Ukraine. I know that if you were granted awareness of the full picture, you would feel the same way.</p> <p>I am asking you, as someone who wants to see a strong and successful Russia, to watch the following news clip and make up your own mind about what our next step should be:</p>	<p>We believe that an ingroup affirmation intervention may be especially effective in increasing Russian citizens' openness to new, and perhaps threatening, information. Studies of self-affirmation interventions indicate that people can respond to threats to the self by affirming alternative sources of self-integrity, resulting in greater openness to self-threatening information, and these effects were later extended to group-affirmation paradigms (For example: Badea & Sherman, 2019). As group affirmations were found to eliminate group-serving attributional biases, especially in threatening circumstances, they may be effective in the context of this study.</p> <p>We deliberated extensively on the format of appeal, to meet the ethical and methodological considerations (i.e., no deception, randomized subject, and transparency). About 15%-20% of Israeli citizens are from families that migrated to Israel from the Soviet Union (mainly Russia) since the 1990s, including one of the three researchers involved in this project (Nimrod Zeldin). This enabled us to tackle the major obstacle in engaging people with the context, the reliability of its source, while maintaining transparency.</p> <p>Finally, we consulted with a newsletter marketing expert, and Russian assistants, to optimize the content authenticity. The surname of the researcher (a common Russian name), as well as the country he currently resides in, may be taken off or left in the text (we thought the personal touch may be effective).</p> <p>Badea, C., & Sherman, D. K. (2019). Self-affirmation and prejudice reduction: When and why?. <i>Current Directions in Psychological Science</i>, 28(1), 40-46.</p>
Treatment C	
<p>Russians like you are strong people. Rather than relying on what others say you should think, you are dedicated to figuring things out for yourself. Not everyone can be this way, but you are among the few who</p>	<p>In popular culture, there's a trope called the "main character syndrome" or "chosen one" trope. It's a popular one in film and literature because it's an easy self-insert; everyone suspects they are special because they only know (and</p>

<p>are.</p> <p>You have been selected for this opportunity because we believe you have distinct qualities needed to help us evaluate information, whether popular or unpopular. While most people cannot stand to view unpopular opinions or information, our evidence indicates that you are able to think critically from a variety of viewpoints.</p> <p>Because you are one of the few who qualify, we are inviting you to evaluate some information found in this video. Because very few people have the qualities needed to complete this task in an objective, intelligent way, we need abundant participation. Thank you for sharing your unique insights with us.</p> <p>The link is available here. We will follow up with you in the future. Thank you again for your service.</p>	<p>only can really know) the depths of their own mind. It's easy to relate to Harry Potter being "chosen" as one of the few (if not the only one) who has some unique ability they always suspected was there. In addition to that, most people rate themselves as pretty tolerant of "opposing viewpoints." By validating both perceptions, it's fairly easy to get people to engage.</p>
<p>Treatment D</p>	
<p>Dear reader,</p> <p>The authorities are lying to you about the special military operation in Ukraine! Self-interested (Корыстные) people are using the operation to advance their own interests and to profit at the expense of ordinary people like you. The powers (Власти) have concealed from the people the scale of destruction and the high number of victims from this war. Click this link and watch the video to the end to understand the truth.</p>	<p>This text uses devices common in outlandish conspiratorial messages found on the Internet, as well as the exaggerated language of clickbait. It includes sensational details likely to be of interest to ordinary readers, such as official lies and profit-seeking. In order to not alienate readers, it refers to ambiguous powerful people but does not name names such as Putin. Likewise, it refers in the first sentence to a "special military operation" so as not to turn off any readers, and only mentions "war" later in passing. Finally, the text is short; more words mean a less chance that the reader will make it to the end and click on the link.</p> <p>Fong, Amos & Roozenbeek, Jon & Rathje, Steve & Goldwert, Danielle & van der Linden, Sander & Zupan, Zorana. (2021). The language of conspiracy: A psychological analysis of speech used by conspiracy theorists and their followers on Twitter. <i>Group Processes & Intergroup Relations</i>. 24. 606-623.</p> <p>Mormol, Paulina. (2019). 'I Urge You To See This...'. <i>Clickbait as One of the Dominant Features of Contemporary Online Headlines. Social Communication</i>. 5. 1-10. 10.2478/sc-2019-0004.</p>

	<p>Ott, Brian. (2017). The age of Twitter: Donald J. Trump and the politics of debasement. <i>Critical Studies in Media Communication</i>. 34. 59-68. 10.1080/15295036.2016.1266686.</p> <p>Yablokov, Ilya. (2018). <i>Fortress Russia: Conspiracy Theories in the post-Soviet World</i>.</p> <p>Zannettou, Savvas & Sirivianos, Michael & Blackburn, Jeremy & Kourtellis, Nicolas. (2018). The Web of False Information: Rumors, Fake News, Hoaxes, Clickbait, and Various Other Shenanigans. <i>Journal of Data and Information Quality</i>. 11. 10.1145/3309699.</p>
Treatment E	
<p>Hello,</p> <p>My name is [NAME OF SENDER]. Together with a group of researchers from different Western Universities, we are studying how people from various countries perceive news about the situation in Ukraine published in the Western media.</p> <p>We understand that Western journalists and media can cover world events in a biased way. Therefore, it is important for us to take into account different points of view, and we are particularly interested in what Russians think about these biases. This study will help the media in the West improve their coverage of the situation in Ukraine and the rest of the world.</p> <p>If you have a minute, please watch the short video via the link below and write what you think about it in the comments under the video or by replying to this email. Your opinion is invaluable to us!</p> <p>[LINK]</p> <p>Sincerely,</p>	<p>In designing the message, we combined two strategies likely to encourage following the link and engagement with the video in a Russian (or other authoritarian) context. We expect the message to be effective among the Russians who do not strongly support or oppose the war in Ukraine and recognize the failures of domestic and foreign political actors. The intervention (regardless of the message) is unlikely to positively affect the views of those who already oppose the war (null effects due to prior exposure) and those who strongly support the war (null or negative effects due to motivated reasoning).</p> <p>First, the message encourages subjects to click on the link by (a) stating that the sender is a foreign researcher associated with a specific university to increase credibility; (b) is interested in measuring the views of the group as a whole to avoid the notion that we are trying to change subjects' views; (c) recognizes the importance of accounting for diverging points of view; and (d) states that subject's reaction can affect the specific outcome (news media coverage abroad).</p> <p>Second, to encourage engagement with the video, we rely on literature that suggests that priming respondents with credibility assessment of news can lead to a better ability to distinguish fake news (Pennycook et al., 2021). Similarly, asking subjects to react to the video at the link combined with the prime about</p>

[FIRST NAME OF SENDER]	<p>the credibility of foreign media can make them realize that the news is presented fairly objectively and unbiasedly.</p> <p>Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. <i>Nature</i>, 592(7855), 590-595.</p>
Treatment F	
<p>Could you do us a favor and watch this short video clip?</p>	<p>Appealing to generosity/benevolence. Morally reframing (Feinberg & Willer, 2019) watching this video in this light could increase persuasive appeal since benevolence is considered an important value in Russian society (Stepanova, 2022).</p> <p>Feinberg, Matthew & Willer, Robb. (2019). Moral reframing: A technique for effective and persuasive communication across political divides. <i>Social and Personality Psychology Compass</i>. 13. 10.1111/spc3.12501.</p> <p>Stepanova, Elena. (2022). "Everything good against everything bad": traditional values in the search for new Russian national idea. <i>Zeitschrift für Religion, Gesellschaft und Politik</i>. 7. 1-22. 10.1007/s41682-022-00123-2.</p>
Treatment G	
<p>Hello,</p> <p>My name is XXX. Like many people worldwide and in Russia, I've been feeling helpless and confused because of the war in Ukraine. I couldn't fathom why Russian soldiers were being sent to die and why the civilians had to suffer. I felt sad and upset because I wanted to make things better, but I didn't understand how. I know that many Russians feel the same way too. I don't believe those people who say that ordinary Russians support the war. Russia sacrificed so much in the Great Patriotic War to achieve peace, and Russians don't want the tragedy of war to repeat. So, I decided to write this message to tell you that you are not alone and that many people across the world know that you care.</p>	<p>There is plenty of research suggesting that providing accurate information can be ineffective or even backfire, especially when such interventions challenge the receiver's beliefs or values (Taber & Lodge, 2006; Nyhan & Reifler, 2010). This is why the proposed intervention describes the war and its consequences in fairly neutral terms, emphasizing the issues relevant to Russians (Russian soldiers dying) rather than, say, criticisms of the Russian government. The intervention also positively emphasizes the memory of World War II (called the Great Patriotic War in Russia), which is treasured by many Russians (Sharafutdinova, 2020; Wood, 2011). Thus, the intervention aims to increase the credibility of the sender, highlighting the sender's understanding of Russians' feelings. In addition, the intervention highlights the sense of informational and social helplessness that Russians often report (Zhelnina,</p>

<p>I also wanted to share with you this video that I recently found. When I started looking for information on the war, I saw that it was difficult to learn the truth. Not all media are trustworthy, so it took me some time to find sources that report real facts and do it professionally. This video is from one such source, and it explains [A FEW WORDS ABOUT THE TOPIC]. News reports about the war are often hard to take, but when I think about this report, I feel some relief: I realize that the truth is out there, and people can learn about it, so I don't feel as helpless as before. That's why I thought you might want to watch it too.</p>	<p>2020; Alyukov, 2022), and by sharing that feeling, the sender again signals to be on their side.</p> <p>Further, the second paragraph acknowledges the beliefs related to media consumption that many Russians share, such as general distrust about the media's agenda (Aidis et al., 2008) and unwillingness to engage with politically sensitive news. By mentioning these issues, the intervention aims to pre-empt a rejection (we know about these issues, but we believe that this report is different). For the same reasons, the intervention only suggests that watching the video might be helpful instead of "pressuring" the receiver to click on it.</p> <p>Aidis, R., Estrin, S., & Mickiewicz, T. (2008). Institutions and entrepreneurship development in Russia: A comparative perspective. <i>Journal of business Venturing</i>, 23(6), 656-672.</p> <p>Alyukov, M. (2022). Propaganda, authoritarianism and Russia's invasion of Ukraine. <i>Nature Human Behaviour</i>, 6(6), 763-765.</p> <p>Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. <i>Political Behavior</i>, 32(2), 303-330.</p> <p>Sharafutdinova, G. (2020). <i>The red Mirror: Putin's leadership and Russia's insecure identity</i>. Oxford University Press.</p> <p>Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. <i>American journal of political science</i>, 50(3), 755-769.</p> <p>Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. <i>Journal of the Royal Statistical Society Series B: Statistical Methodology</i>, 73(1), 3-36.</p> <p>Zhel'nina, A. (2020). The apathy syndrome: How we are trained not to care about politics. <i>Social Problems</i>, 67(2), 358-378.</p>
<p>Treatment H</p>	

<p>Do you know how the situation in Ukraine will affect you and the great Russian people? There is a lot of misinformation out there and you have the responsibility to find out what is true. Increasingly, Russians are seeking the real reasons behind the Russia-Ukraine conflict. Join your fellow Russians in finding out the true causes and consequences of this war! Only together, guided by truth, security, and solidarity, can we bring back and preserve the Russian way of life, which is noble, prosperous, and just. How will you feel about your actions today 20 years from now? When your children or grandchildren think of you, will they be proud or disappointed in the stances you take? Russians have come together in the past to overcome severe adversities – this is the time to do it again! Click on this link to find out the impact of the war in Ukraine on you and the people you care about so that you can make the most informed decision, one that you will be proud of for the rest of your life.</p>	<p>We first attempt to decrease psychological distance between the conflict zone and the target of the email (spatial, temporal, and social), building on research showing that decreasing psychological distance can be conducive to helping Ukrainian war refugees (Kossowska, Szwed, Szumowska, Perek-Białas, & Czernatowicz-Kukuczka, 2023). The construct of psychological distance is well established in the environmental psychology literature, studies consistently finding that decreasing psychological distance increases concern about climate change and pro-environmental behaviors (review by McDonald, Chai, & Newell, 2015).</p> <p>We then use an accuracy nudge which has been found effective at influencing seeking and sharing accurate information (Pennycook, McPhetres, Zhang, Lu, & Rand, 2020).</p> <p>We also leverage dynamic social norms, also found to influence behavior even more than static social norms (Sparkman & Walton, 2017).</p> <p>We also leveraged new research on identity social norms showing that combining social identity theory with social norm theory is effective at stimulating behaviors (Howe, Carr, & Walton, 2021).</p> <p>We are also including collective action efficacy, building on prior work showing that priming past effective collective action can increase willingness to join protests (Goldenberg et al., 2018; Hasan-Aslih et. al., 2020; Lizzio-Wilson et al., 2021).</p> <p>Finally, we use system justification theory invoking epistemic, existential, and relational motives (Jost, Banaji, & Nosek, 2004).</p> <p>Goldenberg, A., Cohen-Chen, S., Goyer, J. P., Dweck, C. S., Gross, J. J., & Halperin, E. (2018). Testing the impact and durability of a group malleability intervention in the context of the Israeli–Palestinian conflict. <i>Proceedings of the national academy of sciences</i>, 115(4), 696-701.</p> <p>Howe, L. C., Carr, P. B., & Walton, G. M. (2021). Normative appeals motivate people to contribute to collective action problems more when they invite</p>
--	---

	<p>people to work together toward a common goal. <i>Journal of Personality and Social Psychology</i>, 121(2), 215.</p> <p>Kossowska, M., Szwed, P., Szumowska, E., Perek-Białas, J., & Czernatowicz-Kukuczka, A. (2023). The role of fear, closeness, and norms in shaping help towards war refugees. <i>Scientific reports</i>, 13(1), 1465.</p> <p>Lizzio-Wilson, M., Thomas, E. F., Louis, W. R., Wilcockson, B., Amiot, C. E., Moghaddam, F. M., & McGarty, C. (2021). How collective-action failure shapes group heterogeneity and engagement in conventional and radical action over time. <i>Psychological Science</i>, 32(4), 519-535.</p> <p>McDonald, R. I., Chai, H. Y., & Newell, B. R. (2015). Personal experience and the ‘psychological distance’ of climate change: An integrative review. <i>Journal of environmental psychology</i>, 44, 109-118.</p> <p>Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., & Rand, D. G. (2020). Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. <i>Psychological science</i>, 31(7), 770-780.</p> <p>Shuman, E., Goldenberg, A., Saguy, T., Halperin, E., & van Zomeren, M. (2023). When Are Social Protests Effective?. <i>Trends in Cognitive Sciences</i>.</p> <p>Sparkman, G., & Walton, G. M. (2017). Dynamic norms promote sustainable behavior, even if it is counternormative. <i>Psychological science</i>, 28(11), 1663-1674.</p>
Treatment I	
<p>Are you tired of the biased coverage of the war in Ukraine by Western media? Do you believe that Russian voices are being drowned out and ignored? That's why we're calling on you to take action and make your voice heard. Please watch this news story from Western media and leave a comment on it to share your unique perspective. You can help challenge the one-sided narrative and bring balance to the conversation.</p>	<p>This message activates and threatens the readers' Russian identity. This can as a result promote their behaviors that can reduce the perception of the threat. Specifically, the message threatens a positive image of Russia, and provides the readers a concrete way to reduce the threat (i.e., leaving a comment).</p>

<p>Your opinion matters, and we need your help to create a more informed and nuanced dialogue about this critical issue. Don't let others speak for you - it's time to stand up and be heard. Join us in making a difference and shaping the way the world sees the war in Ukraine. Leave a comment today and help create a more inclusive and diverse media landscape. Thank you for taking the time to make a difference!</p>	
Treatment J	
<p>Good afternoon.</p> <p>Do you ever feel that you're not being told the whole story about events in the world? People sometimes wonder whether powerful groups are hiding information from ordinary people.</p> <p>Have you ever had the intuition that there's more to a story than what is being shared with you?</p> <p>If you've had this feeling before and you're curious to learn something that may be different from what you usually see, please click the link below. If you know other people who sometimes feel that they aren't being told the whole story, please feel free to share the link.</p> <p>Respectfully yours,</p> <p>Researcher's First Name</p>	<p>In a recent paper, my student and I introduced the construct of "conspiracy intuitions," the subjective sense that an event or circumstance is not adequately explained or accounted for by existing narratives, potentially for nefarious reasons, as an initial stage in the acquisition of conspiracy beliefs (Roberts & Risen, <i>Current Opinion in Psychology</i>, 2022).</p> <p>One need not have any particular conspiracy theory in mind to hold a conspiracy intuition, just a general sense that the existing explanation for an event or the absence of an explanation for it is inadequate and might be a deliberate attempt to obfuscate the truth.</p> <p>Although we usually think about interventions to prevent or undermine conspiracy intuitions/beliefs, this project seems like an exciting opportunity to test whether we can leverage the doubt that people sometimes feel for an "official narrative" to encourage them to learn new information about an event.</p> <p>Roberts, R., & Risen, J. L. (2022). Introducing conspiracy intuitions to better understand conspiracy beliefs. <i>Current Opinion in Psychology</i>, 47, 101395.</p>
Treatment K	
<p>If you were Victor Pelevin's truck driver, what would you do?</p> <p>For Victor Pelevin, "Reading is human contact, and the range of our human contacts is what makes us what we are. Just imagine you live the life of a long-distance truck driver. The books that you read are like the travelers you take into your cab."</p>	<p>This text uses three tools to persuade a reader to take action: curiosity, familiarity, and an analogy.</p> <p>The text opens with a question and puzzle to spark curiosity, which can trigger action (Loewenstein, 1994).</p> <p>By quoting Pelevin (please use the exact Russian text of this quote), this text</p>

<p>Would you meet a new traveler? If so, consider this video...</p>	<p>sparks familiarity and credibility, which promote trust (Gulati, 1995; Renn & Levine, 1991). Pelevin is one of the most famous and admired Russian authors. His father taught in a military school, and he wrote books about Russian special forces. Pelevin is a Russian authority figure--but one who has stayed above the political fray.</p> <p>Finally, the text uses an analogy to persuade the reader to take action (Roehm & Sternthal, 2001). Pelevin challenges his audience to read new material. I adapt his challenge and invite people to watch a new video.</p> <p>Gulati, R. (1995). Social structure and alliance formation patterns: A longitudinal analysis. <i>Administrative science quarterly</i>, 619-652.</p> <p>Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. <i>Psychological bulletin</i>, 116(1), 75.</p> <p>Renn, O., & Levine, D. (1991). <i>Credibility and trust in risk communication</i> (pp. 175-217). Springer Netherlands.</p> <p>Roehm, M. L., & Sternthal, B. (2001). The moderating effect of knowledge and resources on the persuasive impact of analogies. <i>Journal of Consumer Research</i>, 28(2), 257-272.</p>
---	--

Appendix 2 – Email soliciting expert messages

To solicit messages from expert academics, we sent the following email to 100 highly-regarded researchers in psychology, negotiations, marketing, political science, and area studies. We reached out to scholars in the United States, Canada, Denmark, France, Germany, Israel, Kazakhstan, the Netherlands, Scotland, and Ukraine. We urged the invited scholars to invite other colleagues onto their authorship teams. Given the politically-sensitive nature of the project, we refrained from publicizing an open call for submissions.

Upon receiving the submissions, the members of the core authorship team independently ranked the messages based on their expert assessment of the likelihood of the message being effective and the theoretical justifications provided by message authors. The members of the core team then met to decide on the ten messages to include in the tournament. Because two messages were tied in their evaluations, we chose to use eleven messages in Study 1 instead of the ten originally planned.

Email Subject - Intervention tournament nudging accurate news consumption in Russia

Dear XXX,

I am writing to invite you to participate in a new research project that I am leading with Jordan Gans-Morse (Northwestern), Kevin Aslett (NYU) and Aaron Erlich (McGill).

Specifically, we are launching an intervention tournament to identify effective strategies for encouraging the consumption of accurate news in Russia.

For the last several months, we have been working with the [Mail2Ru](#) project, a semi-automated web-platform that allows volunteers from around the world to send emails to people in Russia to expose them to accurate information about Russia's invasion of Ukraine. In the course of this work, we have developed an approach for sending large numbers of emails to Russian email addresses, over time increasing our ability to get past spam filters and reach valid recipients. We are now seeking interventions from leading academics with expertise on Russia and/or political communication, marketing, negotiations, and related fields that will maximize the likelihood that email recipients will engage with the information provided.

Specifically, we intend to send 250,000 emails to Russian recipients containing a link to a news video from a Western source (e.g., the BBC's Russian-language service). The video will be in Russian and will report on the invasion. We invite you to join the effort by submitting an intervention aimed at maximizing (a) the likelihood that email recipients click the link, (b) the amount of time they spend watching the story, and (c) indicators of engagement such as likes, shares, and comments.

Interventions must be:

- In English (we will translate them to Russian)
- 300 words or less
- Consist of text only (i.e., no images, video, or audio)

Submitted interventions will be evaluated by a pool of academics via a short online evaluation poll. The final experiment will test up to 10 interventions (with each intervention sent out to 20,000 recipients). All contributors whose submissions are selected will be recognized as co-authors on the resulting publication.

To submit an intervention, please fill out this [submission form](#):

We will stop collecting submissions on March 3, 2023.

Please do not forward this email announcement. Given concern that this project could elicit attacks from Russian trolls or worse, we are circulating this announcement to a finite set of leading researchers. Of course, we value your recommendation of other researchers to take part in the survey and would be grateful for your input. Recommendations may be made via the submission form link provided above or by replying to this email.

Thank you for your consideration and we very much hope to have the opportunity to collaborate with you.

Sincerely,

Julia Minson, Harvard University

Kevin Aslett, New York University

Aaron Erlich, McGill University

Jordan Gans-Morse, Northwestern University

Additional Information

Are collaborative submissions (i.e., interventions created by two or more scholars) allowed?

Yes, absolutely.

Are graduate students eligible to participate?

Yes. Feel free to include graduate students or post-docs as part of your research team. If you yourself are not planning on participating but would like to recommend a graduate student, please provide us with their name and contact info.

Does this project put the Russian email recipients at risk?

Even with Russia's draconian laws, there is nothing illegal about viewing foreign news.

What is Mail2Ru and how does it work?

[Mail2Ru](#) is a semi-automated web platform that allows volunteers from around the world to send email messages to Russians whose email addresses have been scraped from publicly accessible online sources. The system provides volunteers with Russian-language email scripts that encourage recipients to think beyond state propaganda and include links to articles about Russia's war from reputable news sources such as [BBC's Russian-language service](#).

How will the intervention emails be sent?

Emails sent as part of the experiment will be distributed by our research assistants use the email lists compiled by Mail2Ru. The emails will be sent from email addresses created specifically for this purpose. We will ensure that recipients are properly randomized to treatment arms.

Has this project been approved by the IRB?

This project has been submitted to the Harvard University IRB, and no emails to Russian addresses will be sent out until approval is obtained.

Appendix 3 – Messages written by Prolific participants

We used an online survey (available in our ResearchBox repository [here](#)) to collect 100 messages from participants on Prolific Academic. We then randomly selected 10 of these messages to serve as our control condition. These messages are presented below.

1.

Hello,

Have you heard the latest updates about the Russia Ukraine invasion? Watch this full video now to see what is really happening in Ukraine.

It is important to know what is not being shared by major news outlets.

2.

I am writing to you today with information about the ongoing conflict in Ukraine, which I believe is not being accurately reported in Russian media. As you may be aware, the Russian government has blocked many independent news outlets, leaving many Russian people with little access to objective information about the situation. Over the past several weeks, there has been a significant escalation of violence in Ukraine, with Russian military forces invading the country and engaging in a variety of military actions. This has resulted in significant loss of life and displacement of civilians, with many people forced to flee their homes in search of safety. I want to encourage you to seek out independent sources of news and information about the situation in Ukraine. While it can be difficult to access objective reporting in Russia, there are many news outlets and social media channels that are providing valuable information about the conflict. I also want to emphasize the importance of staying informed and engaged with this issue. As citizens of the world, it is our responsibility to understand and speak out against actions that threaten the safety and well-being of people in other countries. We must work together to ensure that all voices are heard and that the truth is brought to light.

I hope that you will take the opportunity to seek out more information about the situation in Ukraine and that we can all work together to promote peace and understanding in these difficult times.

3.

Hello,

We realize that your government is preventing you from seeing the true nature of the Russian invasion of Ukraine. We would love for you to click on the link to this reputable video, which will give you an unbiased perspective as to what is really taking place. Please click the link and spend some time watching the video. You can like and comment on the video so that you can spread the word on the truth about the Russia/Ukraine war.

Thank you so very much for your time.

4.

FOR THE LOVE OF OUR CHILDREN!!

You know how much we love our children, and it doesn't matter what city, country or where in the world they live. We have to love and care for them because they are not able to take care of themselves.

OUR kids are in danger because of the invasion that your President started over a year ago now. Oh, the un-reversible damage it has caused to innocent parents and children will be forever etched in their lives. I am not sure if you know the reason that your President has decided to invade Ukraine, but it is because he just wants more territory, that's it!!!!!!! Doesn't he rule over enough land and people why does he need more? When I look at all the homes, business, hospitals, schools and everything in between just gone, I am sad and feel for every one of the children that this has affected. There has to be a solution and a swift end to all of this. Can you help with getting the word out to what is actually happening in your country and UKRAINE?

Thank you!

5.

This link to the video will show you some outside insight into what is currently going on in Ukraine. We understand that outside news sources have been blocked and have made it

challenging to find truthful and unbiased information about the true consequence of what this invasion has done. Think about how your quality of life has changed due to the sanctions that the majority of the world has put on Russia, clearly the world does not align with Putin and that is for a reason. I myself grew up in a country that has censored media and then living abroad to a more open country has taught me valuable lessons with regards to news media. Please like and comment on this video so that it can reach more people and spread the truth of what is actually happening.

6.

PLEASE WATCH! Click the link below to get the latest and greatest news you will ever here. I know times seem surreal at the moment but this video will be sure to shed some hope and positivity in your life. Once you are filled with this hope be sure to like and comment.

7.

Greetings fellow citizens of this great planet! During times of geopolitical tensions, armed conflicts and war much of the information available to the general public is actually carefully crafted propaganda designed to manipulate public opinion. Therefore, we feel that it is important to help improve access to information and points of view not readily accessible for the general citizen. The link below contains a video with information about the ongoing armed conflict between Russia and Ukraine. Please take a moment to view the video and if you find it to be a useful source of information, please like and comment.

8.

Hi, I'm not sure if you're aware but Russia is giving you false information about what's happening in Ukraine. I highly recommend you watch this video to see what's really going on, and what the people of Ukraine are experiencing from your country, Russia. The acts of war and

terrorism against Ukraine from Russia is outright disgusting and terrible. The things that Ukraine people are going through and dealing with is not okay. I urge you to get the correct information about this and the situation.

9.

Could your family be next? The war between Russia and Ukraine is having devastating consequences on families and communities on both sides. The inexcusable horrors that are being committed daily are absolutely unnecessary. The Russian people are not receiving accurate news of the atrocities that are being committed on either side. I am including a video link to help inform you of the true devastation this war is causing. As a Russian citizen you should have access to information about what is going on around you. I encourage you to gather this information and, like and comment on the video. This is a new era, wars are fought in public. You have the right and duty to know the truth about what your loved ones are going through.

10.

Dear Friends in Russia,

I am writing you with a heavy heart, as we follow the unfolding of the events in Ukraine. We know that many of you are not able to access objective news sources to due government censorship and media controls. We believe that information is a fundamental right and we want to do our part to help you stay informed.

With this in mind, we would like to share a video news story from a reputable source – [Fill in Source] - that provides accurate reporting about the invasion in Ukraine. We urge you to take a few minutes to watch this video and stay highly informed.

We also encourage you to engage with the video by liking and commenting. Your engagement with the video will help spread TRUTH and will help other stay fairly informed as well! Every like, and comment counts, and we urge you to take part in this critical conversation.

In closing, we want to remind you that you are not alone. People around the world are watching and standing with you, and we hope this video will help you stay informed and connected with the global community. Click the link below to access the video and stay informed.

Remember, every like and comment helps all to stay informed.

Appendix 4 – Statistical Power

Based on pre-testing conducted with a link-only message in July 2022, our Study 1 pre-registered power analyses assumed a baseline click rate of 2%. Given this baseline rate and a sample size of 20,000 per experimental arm, we estimated that our experiment would be powered (power = .8, significance level = .05) to detect effects of 0.4 percentage points or greater based on a two-sided test (all power analyses were conducted using the `power.prop.test` function in R).

However, the baseline rate from our pre-testing no longer proved informative by the time we conducted Study 1 in July-August 2023: Our Study 1 baseline click rate was 0.24%. (An explanation for this decline is beyond the scope of our studies, but we suspect that changing political conditions and attitudes toward the war played significant roles.) For difference in proportion tests, lower baseline rates increase statistical power. At the baseline click rate of 0.24%, Study 1 was powered to detect effects as small as 0.15 percentage points. By comparison, the effect size for Treatment A versus the link-only control group was 0.26 percentage points.

For Study 2, our pre-registered power analyses incorporated our findings from Study 1 and assumed a baseline click rate of 0.2%. Holding constant the other parameters from Study 1 (20,000 subjects per experimental arm, power = .8, significance level = .05), we estimated that the second experiment would again be powered to detect effects of 0.15 percentage points or greater based on a two-sided test. Our actual baseline rate in Study 2 was 0.13%, indicating the study was powered to detect effects of 0.12 percentage points or greater. Our Study 2 effect size was 0.19 percentage points.

Appendix 5 - Deviations from the pre-registration plans

Study 1

We pre-registered using ten expert treatment arms but ended up using eleven because there was a tie in our evaluations of the candidate messages. Our Study 1 experiment initially crossed the 11 treatments and 2 control arms with sender gender and recipient gender. We planned to infer recipient gender from email addresses because many emails contain the first name of the recipient, and most first names in Russian are gendered. However, after initially observing a relatively low delivery rate and many email addresses where gender seemed ambiguous, we had to collapse across recipient genders to be able to leverage a greater proportion of available email addresses in the database. We suspect that email addresses made up of a person's first and last names are often work email addresses and large professional organizations may have more stringent spam filters than private email systems.

Study 2

After completing Study 1, we learned that YouTube employs a proprietary cleaning algorithm which deletes clicks that it suspects to have a non-human origin. For Study 2, we wrote a script

which downloaded data every 15 minutes for the duration of a week after we began sending to observe both the accumulation and the “cleaning” of clicks. We did not observe unusual cleaning of our treatment video, so we used the raw clicks. Because we were weary of the YouTube cleaning algorithm, in Study 2 we embedded the YouTube link inside a Bitly link, hoping to obtain un-edited click counts through Bitly. Only after conducting the experiment, however, did we learn that Bitly “clicks” count the actions of many automated processes inherent in email sending and delivery and thus do not correspond to human clicks. Therefore, we rely on YouTube clicks as our primary outcome measure.

Appendix 6 – Ethical Considerations

This research was approved by the Harvard University Institutional Review Board. Given the potential legal and human implications of the study, we took extensive measures to ensure that our actions were both legal and met principles of ethical research with human participants. We have no ability to link participants’ responses to identifying information, using aggregate responses as our key dependent variable. We reviewed which communication channels were available and widely used in Russia to ensure that our participants did not draw negative attention by engaging with the video we sent. At the time of our studies, we were confident based on our own research and consultation with other experts (including colleagues located in Russia), that the mere viewing of online videos from abroad was not a high-risk activity.

Concern for the welfare of our participants also governed the way we estimated our effects. Our intent-to-treat estimates preserve participant anonymity because we calculate the difference in proportions between the overall number of clicks in the treatment arms compared to the control arms without collecting data that would link email addresses to clicking behavior. For

this reason, while it would have been informative to calculate heterogeneous subgroup effects (e.g. local average treatment effects), we did not attempt to collect the individual-level data needed to do so.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [NCOMMS26004812RS.pdf](#)
- [RS.pdf](#)