Does voluntary disclosure of polarizing information make polarization deeper? An online experiment on Russo-Ukrainian War

Chapkovski, Philipp and Zakharov, Alexei

University of Bonn, University of Chicago

12 February 2023

Does voluntary disclosure of polarizing information make polarization deeper? An online experiment on Russo-Ukrainian War

Philipp Chapkovski,* Alexei Zakharov†

February 12, 2023

Abstract

Does the animosity toward a holder of an opposite political opinion or the behavior toward someone whose opinion on a divisive issue is unknown depends on whether that opinion was disclosed or withheld voluntarily? In order to study this question, we conducted a pre-registered study in Russia, measuring the pro-war dictators' behavior towards their partners with aligned or conflicting views on the war in Ukraine using give-or-take modification of Dictator Game. In the presence of a large polarisation gap (outgroup discrimination), we did not find that intentional vs. unintentional disclosure of the recipients' positions affected the transfers of the dictators; at the same time, dictators' beliefs about the share of war supporters among experiment participants and the donations made by other dictators were causally affected. Our study is the first one to consider this dimension of social interactions, and contributes to the quickly growing literature on political polarisation.

1 Introduction

The differential treatment of those with contrasting opinions in politically or morally loaded matters is often referred to as "affective polarization" (Iyengar, Sood and Lelkes, 2012) — a phenomenon on the rise in the United States (Iyengar et al., 2019) and in other countries (Boxell, Gentzkow and Shapiro, 2020; Gidron, Adams and Horne, 2020); strong group identities along political loyalties or views (Murray, Plagnol and Corr, 2017; Brañas-Garza, Bucheli and Espinosa, 2020) often result in damaging anti-social behavior, as documented by experimental research (Dimant, 2020).

In this paper we report on a hitherto unstudied aspect of social interactions — how pro- or anti-social behavior toward an individual is affected by which agency chose to disclose that individual's position on polarizing issue. In general people treat those who hide their positions as suspicious, even more than those who chose to disclose an unpleasant truth about themselves (John, Barasz and Norton, 2016). But when we think about declaring openly a sensitive political or moral position, the very act of disclosure

*University of Bonn, Adenauerallee 24–42, 53113 Bonn, Germany(chapkovski@uni-bonn.de)
†University of Chicago, Harris School of Public Policy (zakharov1@uchicago.edu)

may serve as a signal about one's willingness to engage in a conflict. Vice versa, avoidance to declare their position openly may be read as a signal to avoid such a conflict (Nagel, 2004). Previous studies noted that people are unwilling to learn the positions of others (Frimer, Skitka and Motyl, 2017) or to interact with people with opposing views (Settle and Carlson, 2019). In this study we focus on the opposite end of the informational chain — how do people react when the others intentionally (do not) disclose a potentially conflicting position, versus when the (non)disclosure is unintentional?

The fact of (non) disclosure of someone's opinion may lead people to re-estimate their beliefs regarding proportions of polarizing opinions and mutual hostility of each group who hold opposite opinions. The role of beliefs about expected polarization can create an effect of self-fulfilling prophecy. Behavioral dimension of polarization, measured through actual decisions towards conflicting parties can be partially driven by misestimated expectations of this affective polarization by others (Druckman et al., 2022). On average people overestimate the degree of hostility of outgroup members to their own group (Lees and Cikara, 2020). They also tend to have exaggerated beliefs about the polarization in the society, especially regarding animosity of another party (Chambers, Baron and Inman, 2006; Enders and Armaly, 2019). By investigating how the intention to reveal one's position influences the target's beliefs, this study provides additional insights that may help us understand the appearance of the gap between perceived and actual polarization.

Our empirical focus is on Russia — an autocracy with a polarized society (Frye, 2010), with patterns of disagreements regarding political issues among Russians being similar in size to partisan gaps observed in polarized democracies (Shirikov, 2021).

About one quarter of Russian population do not support the war in Ukraine according to the surveys (Center, 2022), although the true figure must be somewhat higher due to a substantial preference falsification (Chapkovski and Schaub, 2022). There is a plentiful anecdotal evidence of an animosity between those who support the war and those who oppose it, sometimes even running within single families (Media, 2022). War supporters enjoy an asymmetric power balance over those who oppose the war, and were known to be actively involved in identifying dissidents, including pupils reporting to authorities on their teachers (Meduza, 2022), neighbors on neighbors (Realii, 2022), and even relatives on one another (Medialeaks, 2022); historically the Soviet repressive regime intensively used citizens' willingness to report their counterparts with whom they disagree (Fitzpatrick, 1996). At the same time, there have been some incidences of animosity in the opposite direction, such as vandalism directed against cars displaying pro-war symbols (Avtovzglyad, 2022).

We argue that holders of the political opinion who believe that they are in minority should be less likely to disclose their positions. Hence, it should be informative if someone's position is concealed voluntarily (rather than by chance), and attract more hostility from holders of the majority opinion. Moreover, a voluntary disclosure of a minority position may be indicative of more intensive preferences, and attract more hostility compared with the case where the position is disclosed involuntarily. To test these arguments, we use an online dictator game with an option to give to or take a part of a partner's endowment ($N = 1594$) (List, 2007).

Variants of the dictator game were used to measure ingroup/outgrop hostility resulting from polariza-

tion of opinions (Carlin and Love, 2013, 2018; Yair, 2020; Dimant, 2020). We augment this design in one important respect — by allowing some of the recipients to voluntarily choose to disclose their positions on the divisive issue (in our case — support or opposition to the war in Ukraine). In the *Forced Reveal* treatment, the decision to provide the recipient's position to the dictator is made by the computer; in the *Recipient reveal* treatment, the decision is made by the recipient. The dictator, before deciding on the distribution, is informed both about the recipient's type (pro-war, anti-war, or unknown), and who made the decision to show/conceal the type: the experimenter or the recipient herself.

We hypothesised that the dictators would be more generous toward recipients with unknown positions if the position was concealed by the experimenter than by the recipient, as the dictators would expect that a smaller share of war opponents would choose to reveal their positions, compared with war supporters, and, as a result, voluntary concealment would be informative of one's position. The dictators sould also contribute less to holders of minority opinion if the opinion was disclosed voluntarily rather than by the experimenter.

We did not find that the pro-war dictators were less generous to recipients with unknown positions if the position of the recipient was concealed voluntarily, rather than by computer (the effect was in the direction that we expected, but not statistically significant). At the same time, the agency that withheld information on the recipient's group identity did causally affect the beliefs of dictators in two ways. If the decision not to reveal was made by the recipient, rather than by computer, then the dictator's belief about the share of war supporters among participants, and the belief about the average donation by other dictators, were both lower.

Our paper contributes to the current literature on polarization by focusing on the agency that discloses or conceals positions on a polarizing topic. An earlier study, using a similar framework, looked at the correlates and consequences of the decision to learn the partner's position (Chapkovski, 2022*a*). Despite the fact that there is evidence how the a possibility to select information by a decision-making agent, increases outgroup discrimination (Akın, 2019), so far little is known how the disclosure of information by a partner in strategic interactions influences a decision-maker. Our paper intends making a first step in closing this gap.

At the same time, some psychological studies have examined the consequences of (non-)disclosure of sensitive information on interpersonal relations. No matter what kind of information is concealed, the overall consensus is that a person who does not want to articulate their position clearly is perceived with suspicion. People who prefer to 'stay out of it', claiming that they do not take any side in political or moral dilemmas, are considered less trustworthy than even those who have a position opposite to one held by a respondent (Silver and Shaw, 2022). When a person tries to conceal their potentially stigmatized identity, regardless of the nature of this identity, they were perceived as less moral and sociable than people who did not (Le Forestier, Page-Gould and Chasteen, 2022). And even when people try to avoid one's envy by hiding the success from their social circle, this concealment provoked harsh negative feelings towards them (Roberts, Levine and Sezer, 2021).

While our study uses attitudes of Russians towards the ongoing war in Ukraine as a group-defining factor, it is not the first paper that examines the reaction on a partner's intention to withheld information.

In a series of psychological experiments, people had to answer sensitive questions with a possibility to abstain from answering them. In a Trust game those who preferred to conceal their answers were considered less trustworthy than revealers, thus first movers transferred them substantially less amounts (John, Barasz and Norton, 2016). Unlike the study by John, Barasz and Norton (2016), our paper creates group identities along political divisions making it potentially insightful for a growing body of polarization studies.

The rest of this paper will be structured as follows. Section 2 contains the theoretical model. In Section 3 we describe our experimental method and data. In Section 4 we provide the main results. Section 5 concludes.

2 Hypotheses and theoretical model

Our first theoretical prediction is that, whenever faced with a recipient who did not disclose one's position, the dictator who belongs to a majority group will donate more if the recipient's position was concealed involuntary, since a recipient belonging to a minority group is less likely to reveal one's group identity to the dictator. The second prediction is based on the assumption that war opponents who strongly (rather than weakly) oppose the war will be more likely to reveal their positions to the dictator. Hence, for a recipient who is known to be a war opponent, the agency that revealed her position to the dictator (the experimenter vs the recipient) will be informative of the recipient's intensity of opposition to the war. As a result, dictators will donate less to recipients known to be war opponents if the decision to reveal the recipient's position was made voluntarily. Our pre-registered hypotheses are as follows:[1]

H1: Donations by majority dictators to unknown recipients will be lower if it was recipient's decision to conceal the identity;

H2: Dictator in-group bias (the difference between donations to recipients with known positions who agree vs. disagree with the dictator) be higher if it was the recipient's decision to reveal the respondent's position (vs. if the decision was made by the computer)

Below we illustrate the argument behind our first hypothesis with a theoretical model. Consider a game between two players: Dictator and Recipient. Recipient can either belong to the same social group as the dictator (denoted by $\theta = 0$), or belong to the opposing group (denoted by $\theta = 1$). The dictator's social group is common knowledge, while the recipient's group identity is private information; denote by $p \in (0, 1)$ the probability that the recipient is from the same group as the dictator.

The timing of the game is as follows:

$t = 1$ The recipient of type θ makes a declaration of one's group identity $d \in \{N, \theta\}$, where N indicates the decision to conceal group identity. For simplicity, we assume that the recipient is not capable of lying, so either $d = N$, or $d = \theta$.

[1] Our experiment was pre-registered at https://aspredicted.org/HBF_G67, with the following stated research question: "Do dictators behave differently in a give-or-take dictator game (DG) when they know that they are not aware of their partner's position regarding polarizing issue because it was actively withdrawn by the partner? Whether this behavior will interact with the position of a dictator?". An additional research question and a hypothesis concerned dictator behavior in a game where a dictator could choose to reveal the position of the recipient (not analyzed in this paper).

$t = 2$ The dictator observes signal $s \in \{N, 0, 1\}$, where N means that the dictator does not observe the recipient's type. With probability $q \in [0, 1]$, the dictator observes only the recipient's choice, so $s = d$; with probability $(1 - q)r$ the dictator observes the recipient's type directly ($s = \theta$), and with probability $(1-q)(1-r)$ the recipient's type is not observed ($s = N$), where $r \in (0, 1)$. Upon observing s, the dictator then chooses the level of donation $x \geq 0$ to the recipient.

Hence, $q = 1$ corresponds to the Recipient Reveal treatment, and $q = 0$ — to the Forced Reveal treatment where the recipient's type is observed with some probability r, with a possibility of an intermediate case.

The dictator derives utility from two sources. First, it is own private consumption; second, the dictator's utility has a warm glow component that is greater if the recipient belongs to the same group as the dictator. Let the dictator's payoff be quasilinear with respect to own consumption:

$$u_D = -x + 2(1 - \theta(1 - a))\sqrt{x},$$

where the parameter $a \in (0, 1)$ denotes how little the dictator cares about the donation made to a member of the opposing group, relative to the donation made to a member of one's own group.

We assume that signaling group identity affects the payoffs of recipients in two ways. First, there is an intrinsic value (or cost) to signaling group identity. Second, signaling one's group identity affects the beliefs of the dictator regarding the recipient's type, indirectly affecting the payoff. So assume that the payoff of the recipient is as follows:

$$u_R = x + \mathbf{1}_{(d=\theta)}v,$$

where v is the (potentially negative) value of expressing group identity. We assume that v is recipient' private information, and is distributed is independently of θ, according to distribution $F(\cdot)$ that is assumed to be uniform over $[a^2 - 1, 1 - a^2]$, with density $f > 0$ over that interval.

The solution concept we are going to use is the Bayesian Nash equilibrium. Denote by w the probability that $\theta = 1$ given $s = N$. Then the donations, depending on the signal observed by the dictator, should be

$$x_0 = 1, \qquad x_1 = a^2, \qquad x_N = (aw + 1 - w)^2. \tag{1}$$

The expected payoffs of the recipient of type θ, depending on whether the group identity is declared, are

$$u_R(0) = (1 - q)rx_\theta + (1 - r + rq)x_N, \qquad u_R(1) = (q + r - rq)x_\theta + (1 - q - r + rq)x_N + v.$$

This gives us the following cutoff values for v:

$$\bar{v}_0 = q(x_N - x_0), \qquad \bar{v}_1 = q(x_N - x_1). \tag{2}$$

A recipient with type (θ, v) will declare his group identity if and only if $v \geq \bar{v}_\theta$. As $x_0 > x_1$, a recipient is more likely to reveal his type if he shares group identity with the dictator.

By the Bayes rule, the probability w is given by

$$w = \frac{(1-p)(1-q)(1-r) + (1-p)qF(\bar{v}_1)}{(1-q)(1-r) + q(pF(\bar{v}_0) + (1-p)F(\bar{v}_1))}. \tag{3}$$

The pure-strategy Bayes-Nash equilibrium will then be any tuple $(w, \bar{v}_0, \bar{v}_1, x_0, x_1, x_N)$ such that (1), (2), (3) hold.

Our theoretical prediction is formulated as follows:

Proposition 1 The following is true.

1. There exists at least one equilibrium.
2. Consider the equilibrium with the smallest possible w. Let x_N be the equilibrium donation to a recipient with unknown position. Then $\frac{\partial w}{\partial q} > 0$ and $\frac{\partial x_N}{\partial q} < 0$.

In the recipient reveal treatment, the probability that a recipient with an unknown type is from the opposing group is larger than in the forced reveal treatment; this directly follows from Proposition 1. Our first hypothesis follows as a result.

In this model, the dictator's type is known to the recipient. We do not expect these results to change if the dictator's type is unknown but the probability θ is sufficiently small. The latter setting corresponds to our experiment, where the group identity of the dictator is unknown to the recipient, and the dictators are recruited from the group having a significant majority.

Equilibrium existence is a standard fixed-point result. The comparative statics result follows from the fact that an individual from an opposing group is more likely to hide one's position. Hence, as we move from Forced Reveal to Recipient Reveal setting, it becomes more likely that an individual with an unknown position belongs to the opposing group. This may not hold if we allow the distribution $F(\cdot)$ to be sufficiently nonuniform, so the values $F(\bar{v}_0)$ and $F(\bar{v}_1)$ change at sufficiently different rates as x_N varies with w.

3 Data and Method

We used an online crowdsourcing platform *Toloka* (Chapkovski, 2022*b*) to recruited a total of 797 participant pairs. In terms of functionality this platform is similar to its competitors, mTurk and Prolific; it provides a quick unique access to the large number of participants from the post-soviet countries, including Russia (for more detailed information on this platform see Section A.1 in the Online Appendix). The subjects played a take-or-give version of the dictator game, where the options of a dictator ranged between giving to the recipient \$0.50 from one's own \$1.00 endowment, and taking the recipient's entire endowment of \$0.50; on top of that, subjects received a \$0.50 participation fee. Average time to complete the experiment was 15.3 minutes.

The experiment was conducted one day following a prescreening survey where the participants were asked about their support for the war. At the beginning of the experiment, subjects were again asked the question about war support: *Please tell whether or not you support the actions of Russian forces in*

Ukraine, with two answer options — *Support* and *Do not support*. A subsequent question elicited the strength of support/opposition to the war.

All dictators invited to participate following the prescreening survey were war supporters, while 427 of the recipients were war supporters and the rest opposed the war. The dictator-recipient pairs were assigned randomly into one of the two treatments. In the *Forced Reveal* treatment, the dictator was randomly informed about the position of her recipients prior to the decision donation. In the *Recipient Reveal* treatment the decision whether to reveal the recipient's position to the dictator was made by the recipient himself. In both treatments, the dictators were informed about the agency (computer in FR, recipient in RR) that decided to show or hide the position of their recipients. The two treatments (as well as the recipient positions within each treatment) were balanced in terms of dictators' observables (see Tables OA1, OA2, and OA3 in the Online Appendix).

A total of 1606 invited subjects completed the experiment: 401 pairs in the Forced Reveal treatment and 402 pairs in the Recipient Reveal treatment.[2] A total of 5 pairs were dropped where the dictator replied "Do not support" to the war support question in the main survey. Among the resulting 1594 subjects, 46.9% were female (53.6% for Russian population); the median age was 36 years (39.5 for Russia); and 48.3% had higher education (26% in Russia in 2015). Sample attrition among dictators following the treatment screen was very small (6 out of 404 in the Forced Reveal treatment and 0 out of 399 in the Recipient Reveal treatment).

When the data collection for both recipients and dictators were completed, we randomly matched recipients and dictators from corresponding treatments to calculate subject payoffs.

4 Results

There is strong evidence of outgroup animosity driven by political polarization. Dictators, on average, took from the recipients who opposed the war or whose position was unknown, and gave a small amount to those who supported the war (see Figure 1). The difference in the giving/taking behavior toward war supporters and war opponents was 14.5 cents or 0.51sd ($p < 0.0001$, two-tailed t-test). The corresponding difference in behavior toward war supporters and recipients with unknown position was also large (12 cents or 0.42sd, $p < 0.0001$, two-tailed t-test), while the difference in behavior toward war opponents and recipients with unknown positions was not significant.

In Table 1, we investigate whether dictator donations depend on the type of the recipient (war supporter, war opponent, or unknown), on the treatment (forced or recipient reveal), and whether the effect of the type of the recipient is different across treatments. The regressions are OLS. In Column 1 we report the baseline model. Demographic controls as in Table OA1 are used in Columns 2 and 3, and in Column 3 we repeat the analysis with additional behavioral controls. We used a slightly modified and abridged version of Information Avoidance Scale (IAS, Howell and Shepperd, 2016); a three-question scale for measuring risk attitudes, adapted from 7-question scale used in Dohmen et al. (2011); and to

[2] An additional 195 dictator-recipient pairs participated in the "Dictator Reveal" treatment, the results of which are not reported in our paper.

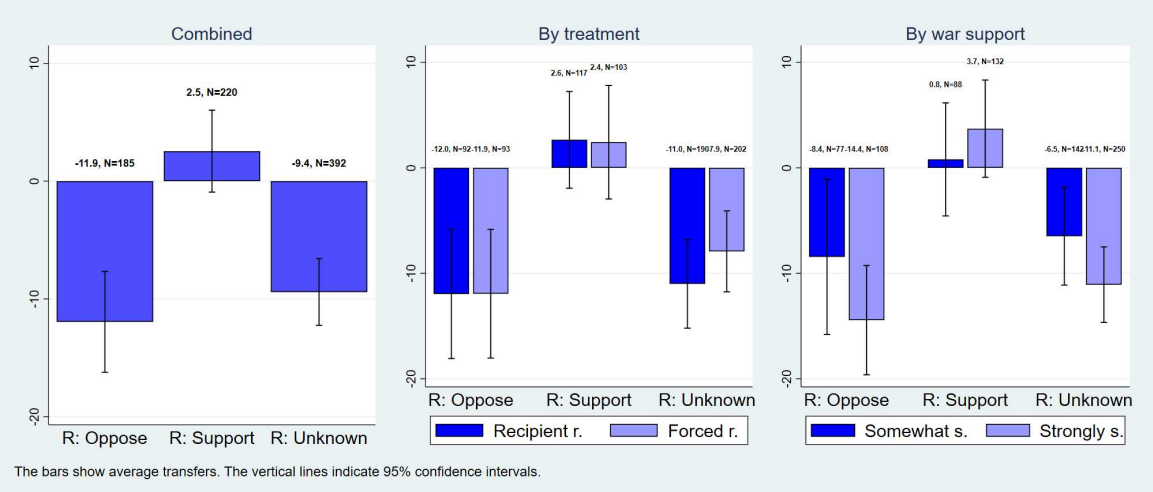

Figure 1: Donation/taking by dictators, depending on recipient position, 95% CI

provide an additional estimate of polarization, we asked a shortened version of set of questions used to compose the Social Distance Index (Druckman et al., 2019).

Dictators (all of whom are pro-war) gave significantly more to pro-war recipients than to anti-war recipients. At the same time, donations to recipients with unknown positions did not significantly differ between forced and recipient reveal treatment. According to our expectations, the amount of donation to recipients with unknown positions has to be larger in the Forced Reveal treatment; this effect is present, but is not statistically significant. Likewise, we did not find that the dictator in-group bias was higher if it was the recipient's decision to reveal her position. We hypothesised the latter effect because the decision to show one's position may be indicative of the intensity of support toward one's position, and, therefore, amplify the dictator's ingroup/outgroup bias; there is evidence that affective polarization is directed against elites to a greater degree than against rank-and-file members of the opposing groups Druckman and Levendusky (2019).

In Columns 4-6 of Table 1 we investigate whether the size of the partisan gap depended on the intensity of the dictator's beliefs; we should expect the dictators with more partisan views to exhibit greater partisan bias. This expectation is partly supported by the data; when demographic controls are included (Column 5), the difference between the amount donated to war supporters vs war opponents is larger for dictators who strongly support (vs somewhat support) the war ($p = .0915$). This difference is positive but not significant at conventional levels in Column 4 or Column 6.

We proceed to investigate the effect that the agency that provided (or withheld) information on recipient position had on dictator beliefs. Our first expectation concerned the effect of the treatment (forced vs recipient reveal) on the dictator beliefs about the position of the recipient, in case the position is not known to the dictator. If an anti-war participant is less likely to reveal her position to the dictator, then a voluntary (vs forced) decision to conceal one's position would be informative of the recipient's position.

The post-treatment survey included two questions: "Out of 100 participants B taking part in this study and supporting the actions of Russian forces in Ukraine, how many do you think agreed to tell

	(1)	(2)	(3)	(4)	(5)	(6)
R: Support	14.61***	15.05***	14.96***	9.462*	9.558*	9.618*
	(3.923)	(3.907)	(3.918)	(5.133)	(5.104)	(5.119)
R: Unknown	0.957	1.090	1.055	0.505	0.628	0.814
	(3.576)	(3.563)	(3.564)	(4.709)	(4.694)	(4.701)
Forced	0.0210	0.393	0.404	0.243	0.576	0.586
	(4.140)	(4.128)	(4.130)	(4.138)	(4.122)	(4.128)
R: Support × Forced	-0.243	-1.164	-1.247	-0.476	-1.373	-1.392
	(5.623)	(5.601)	(5.606)	(5.618)	(5.592)	(5.602)
R: Unknown × Forced	3.058	2.775	2.666	2.661	2.393	2.355
	(5.024)	(5.002)	(5.003)	(5.022)	(4.995)	(5.001)
D: Definitely support				-6.012	-6.741	-5.384
				(4.198)	(4.222)	(4.324)
R: Support × D: Definitely support				8.930	9.592*	9.293
				(5.709)	(5.678)	(5.686)
R: Unknown × D: Definitely support				1.534	1.640	1.167
				(5.135)	(5.119)	(5.134)
IAS (0-1)			-2.350			-1.792
			(2.893)			(2.919)
Risk (0-1)			3.135			3.496
			(4.751)			(4.752)
SDI: pro (0-1)			6.256			5.284
			(5.461)			(5.585)
SDI: contra (0-1)			-5.716			-4.702
			(3.953)			(4.054)
Demo controls	NO	YES	YES	NO	YES	YES
N	797	797	797	797	797	797
R^2	0.0423	0.0672	0.0724	0.0483	0.0744	0.0778

OLS regressions. The dependent variable is the dictator give or take decision (-50 to 50). Demographic controls are as in Table OA1. IAS is the information avoidance scale, 0-1. Risk is the risk preference scale (0-most risk averse, 1-least risk averse). SDI pro/contra is the social distance (0-lowest, 1-highest) to war supporters/opponents.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1: Donation/taking by dictators

Participant A whether he (or she) [1) supports; 2) does not support] the actions of Russian forces in Ukraine?". Across all treatments, the dictators believed that pro-war recipients were more likely than anti-war recipients to reveal their positions (64% vs. 43.6% of $p < 0.0001$, $N = 399$, paired t-test).

In accordance with our expectation, the treatment had a causal effect on the beliefs of dictators about the overall share of war supporters among experiment participants, measured by the question: *Out of 100 participants taking part in this study, how many do you think support the actions of Russian forces in Ukraine?* That is, dictators whose recipients had an option to conceal their positions, believed on average that there are fewer war supporters than those dictators who could see or not see their partner's position due to a computer choice. The answer for dictators in the FR treatment was 71.7, compared with 69.5 in the RR treatment; the difference was statistically significant ($p = 0.0933$, two-tailed t-test).

For the dictators who strongly supported the war, the treatment had a significant effect on dictator beliefs about the even if the position of the recipient was known to the dictator ($p = 0.0615$, $p = 0.0354$, and $p = 0.0505$, respectively, for recipients who oppose the war, support the war, and whose position is unknown, two-sided t; see Figure 2). We conjecture that this can be explained by the contextual

settings in which a recipient can openly declare his/her position regarding the war. The very fact that the recipients have the ability to freely state one's opinion on a divisive issue may signal the dictator that the anti-war minority is not that small as he/she believed.

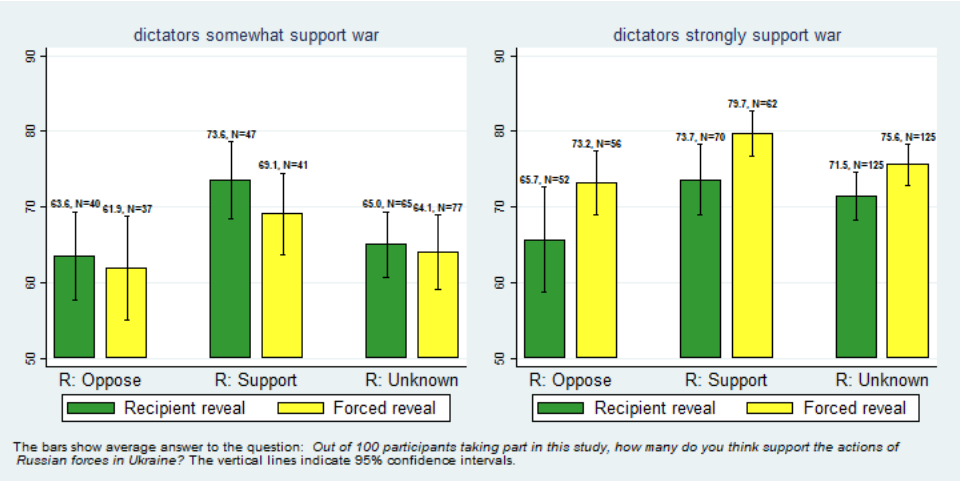

Figure 2: Beliefs about the share of war supporters, depending on treatment and recipient position, 95% CI)

In Table 2 we regress the belief of dictators about the proportion of war supporters on the treatment (forced or recipient reveal) and the recipient's position (war supporter, war opposer, or unknown). We do it separately for dictators who weakly support (Columns 1-3) and strongly support the war (Columns 4-6). For dictators who strongly support the war, the effect of the Forced vs Recipient reveal treatment is significant if the recipient is anti-war or if the recipient's position is unknown: The belief about the share of pro-war supporters is lower if the dictator faces an recipient who voluntarily chose to disclose her anti-war position, or voluntary chose to hide her position.

If the beliefs about the share of war supporters in the population were affected by the treatment, then it is likely that the beliefs about the decisions made by other dictators should be affected as well. The following question was asked after the give/take decision: *At the decision making stage, which decision will the majority of Participants A [the dictators] choose?* The treatment indeed had a causal effect on these beliefs. If the position of the recipient was unknown, the beliefs were significantly higher in the forced reveal treatment ($p = 0.0153$, two-sided t-test)[3]. This effect potentially may shed some light on the causes of the failure to observe treatment effect. While dictators do not distinguish between recipients who hide their views intentionally and those whose views were hidden automatically, they do believe that this intentionality affects other dictators.

In Table 3 we report the OLS regression where the dependent variable is the belief about donations made by other dictators. In the bottom of the table we report the marginal value of the Forced Reveal treatment for dictators who did not know the position of the recipient. In all specifications, the marginal effect was positive and significant at level $p = 0.0302$ or better.

In Columns 3-6 of Table 3 we look at the effect of the intensity of the dictator's preferences on beliefs

[3]See the distribution of answers to this question, broken down by the treatment and the recipient's position in appendix (Figure OA2).

	(1) Weak	(2) Weak	(3) Weak	(4) Strong	(5) Strong	(6) Strong
R: Support	9.996**	9.450**	9.485**	8.003**	9.025***	9.438***
	(4.081)	(4.124)	(4.090)	(3.208)	(3.232)	(3.214)
R: Unknown	1.446	0.815	1.815	5.826**	5.733**	5.926**
	(3.812)	(3.875)	(3.831)	(2.891)	(2.887)	(2.865)
Forced	-1.654	-2.376	-1.881	7.525**	7.264**	7.148**
	(4.327)	(4.465)	(4.415)	(3.374)	(3.392)	(3.362)
R: Support × Forced	-2.820	-1.691	-1.462	-1.456	-2.589	-2.375
	(5.929)	(6.036)	(5.976)	(4.552)	(4.599)	(4.572)
R: Unknown × Forced	0.673	1.590	0.429	-3.397	-2.919	-2.260
	(5.379)	(5.496)	(5.440)	(4.037)	(4.062)	(4.030)
IAS (0-1)			3.211			2.687
			(2.936)			(2.521)
Risk (0-1)			9.248*			-0.184
			(5.441)			(3.685)
SDI: pro (0-1)			-5.851			-7.568*
			(5.952)			(4.516)
SDI: contra (0-1)			15.24***			9.013***
			(4.945)			(3.069)
Demo controls	NO	YES	YES	NO	YES	YES
N	307	307	307	490	490	490
R^2	0.0381	0.0573	0.0982	0.0430	0.0737	0.0983

OLS regressions. The dependent variable is the dictator give or take decision (-50 to 50). Demographic controls are as in Table OA1. IAS is the information avoidance scale, 0-1. Risk is the risk preference scale (0-most risk averse, 1-least risk averse). SDI pro/contra is the social distance (0-lowest, 1-highest) to war supporters/opponents.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2: Beliefs about proportion of war supporters

about actions of other dictators. In one specification we find that the effect of the recipient being anti-war is larger if the dictator is strongly pro-war.

Finally, we analyze the recipient's decision to reveal the position to the dictator. Our theoretical argument was based on the assumption that war opponent recipients would be less likely to reveal their positions than war supporters, and that the recipients who weakly opposed the war were less likely to reveal their position than strong opponents.

In total, 209 recipients out of 399 or 52.4% decided to reveal their opinion, and 190 recipients decided not to reveal. We did not find that war supporters were more or less likely to reveal their positions to the dictator than war opponents. At the same time, the shares of individuals who revealed was high for strong opponents, was much lower for weak opponents, and was intermediate for both strong and weak supporters (See Figure OA1 in the Online Appendix). Therefore our intuition was partly correct — intensity of preferences was positively associated with the decision to show one's position, but only for participants who opposed the war. In the Online Appendix (OA4) we report the marginal logit effects for the recipient decision to reveal one's position to the dictator. There also was a strong positive association with information avoidance (individuals with higher information avoidance are also more likely to reveal their opinion), and with risk preferences (individuals who are more likely to risk are more likely to reveal opinion).

	(1)	(2)	(3)	(4)	(5)	(6)
R: Support	10.19**	10.45**	10.92**	7.598	7.694	8.426
	(4.239)	(4.261)	(4.258)	(5.547)	(5.569)	(5.567)
R: Unknown	-1.109	-0.969	-0.874	-2.522	-2.233	-2.059
	(3.864)	(3.886)	(3.873)	(5.088)	(5.122)	(5.113)
Forced	3.171	3.471	3.403	3.442	3.626	3.556
	(4.473)	(4.503)	(4.488)	(4.472)	(4.498)	(4.489)
R: Support × Forced	-7.402	-8.125	-8.604	-7.664	-8.288	-8.695
	(6.075)	(6.110)	(6.092)	(6.071)	(6.101)	(6.092)
R: Unknown × Forced	4.314	3.786	3.435	3.881	3.422	3.157
	(5.428)	(5.456)	(5.437)	(5.426)	(5.450)	(5.439)
Definitely support/oppose				-7.336	-7.937*	-6.555
				(4.536)	(4.607)	(4.703)
R: Support × Definitely support/oppose				4.739	5.050	4.642
				(6.170)	(6.195)	(6.185)
R: Unknown × Definitely support/oppose				3.181	2.983	2.746
				(5.549)	(5.585)	(5.583)
IAS (0-1)			-5.538*			-4.761
			(3.143)			(3.175)
Risk (0-1)			5.870			6.217
			(5.162)			(5.169)
SDI: pro (0-1)			-0.463			-2.555
			(5.934)			(6.074)
SDI: contra (0-1)			-10.43**			-8.777**
			(4.296)			(4.409)
Forced × (1+R: Unknown)	7.485	7.257**	6.838**	7.323**	7.047**	6.713**
		(3.074)	(3.097)	(3.09)	(3.074)	(3.095)
Forced × (1+R: Support)	-4.231	-4.654	-5.201	-4.222	-4.662	-5.139
	(4.11)	(4.148)	(4.151)	(4.106)	(4.141)	(4.151)
Demo controls	NO	YES	YES	NO	YES	YES
N	797	797	797	797	797	797
R^2	0.0168	0.0239	0.0367	0.0226	0.0308	0.0406

OLS regressions. The dependent variable is the dictator's belief about average donation of other dictators. Demographic controls are as in Table OA1. IAS is the information avoidance scale, 0-1. Risk is the risk preference scale (0-most risk averse, 1-least risk averse). SDI pro/contra is the social distance (0-lowest, 1-highest) to war supporters/opponents.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Beliefs about donations made by other dictators

5 Discussion

Our findings suggest that, in the presence of intergroup animosities, social expectations are affected both by the extent to which people are aware of each other's group identities, and the agency that decides to supply or withhold information on group identity. In particular, it matters for dictator beliefs whether the recipient's position on a divisive issue was concealed voluntarily or by the experimenter; voluntary concealment leads the dictators to believe that experiment participants are more likely to hold a minority opinion, and that other dictators should donate less in the experiment.

We find outgroup hostility by supporters of an authoritarian regime against its opponents, underscoring the importance of peer control over a minority of regime opponents for authoritarian survival (e.g. Geddes et al., 2018). The true size of the partisan gap may be larger due to the fact that the study is conducted in an authoritarian setting where preference misrepresentation is a problem (Chapkovski

and Schaub, 2022) and some dictators (who, by design, were all supposed to be pro-war) were actually anti-war. Our findings may also contribute to the literature studying public support for conflicts (e.g. Gelpi, Feaver and Reifler, 2009), as we document polarization along domestic support/oppositon to an armed conflict.

Our paper is one of the first which examines the polarization generated by the attitudes towards the war in Ukraine. However, the generalizability of its results is limited as it is often the case with online studies: the audience of online surveys tend to be younger, and less attentive than the general population. But this war, and strong negative emotions both supporters and opponents of it in Russia feel towards each, provide researchers with the opportunity to examine what role the human agency plays when outgroup discrimination is expected by both sides of the conflict.

References

Akın, Zafer. 2019. "Dishonesty, social information, and sorting." *Journal of Behavioral and Experimental Economics* 80:199–210.
URL: *https://www.sciencedirect.com/science/article/pii/S2214804318300466*

Avtovzglyad. 2022. "Kak nakazhut voditelya... [How the driver...]." *Avtovzglyad* . Retrieved 2022-07-05.
URL: *https://www.avtovzglyad.ru/obshestvo/socium/2022-03-23-kak-nakazhut-voditelja-za-bukvu-z-na-avtomobile/*

Boxell, Levi, Matthew Gentzkow and Jesse M. Shapiro. 2020. Cross-Country Trends in Affective Polarization. Working Paper 26669 National Bureau of Economic Research. Series: Working Paper Series.
URL: *https://www.nber.org/papers/w26669*

Brañas-Garza, Pablo, Marisa Bucheli and María Paz Espinosa. 2020. "Altruism and information." *Journal of Economic Psychology* 81:102332. Publisher: Elsevier.

Carlin, Ryan E. and Gregory J. Love. 2013. "The politics of interpersonal trust and reciprocity: An experimental approach." *Political Behavior* 35(1):43–63. Publisher: Springer.

Carlin, Ryan E and Gregory J Love. 2018. "Political competition, partisanship and interpersonal trust in electoral democracies." *British Journal of Political Science* 48(1):115–139.

Center, The Levada. 2022. "Konflikt s Ukrainoi [Conflict with Ukraine]." *The Levada Center* . Retrieved 2022-07-05.
URL: *https://www.levada.ru/2022/06/30/konflikt-s-ukrainoj-3/*

Chambers, John R, Robert S Baron and Mary L Inman. 2006. "Misperceptions in intergroup conflict: Disagreeing about what we disagree about." *Psychological Science* 17(1):38–45.

Chapkovski, Philipp. 2022*a*. "Information avoidance in a polarized society." *Journal of Behavioral and Experimental Economics* 101:101940.
URL: *https://www.sciencedirect.com/science/article/pii/S2214804322001112*

Chapkovski, Philipp. 2022*b*. "Interactive experiments in Toloka.". Publisher: Rochester: SSRN.

Chapkovski, Philipp and Max Schaub. 2022. "Solid support or secret dissent? A list experiment on preference falsification during the Russian war against Ukraine." *Research & Politics* 9(2):20531680221108328.

Chen, Daniel L., Martin Schonger and Chris Wickens. 2016. "oTree—An open-source platform for laboratory, online, and field experiments." *Journal of Behavioral and Experimental Finance* 9:88–97.
URL: *http://www.sciencedirect.com/science/article/pii/S2214635016000101*

Dimant, Eugen. 2020. "Hate trumps love: The impact of political polarization on social preferences." *Available at SSRN* .

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp and Gert G. Wagner. 2011. "Individual risk attitudes: Measurement, determinants, and behavioral consequences." *Journal of the european economic association* 9(3):522–550. Publisher: Oxford University Press.

Druckman, James N and Matthew S Levendusky. 2019. "What do we measure when we measure affective polarization?" *Public Opinion Quarterly* 83(1):114–122.

Druckman, James N, Samara Klar, Yanna Krupnikov, Matthew Levendusky and John Barry Ryan. 2022. "(Mis) estimating affective polarization." *The Journal of Politics* 84(2):1106–1117.

Druckman, James, S. Klar, Y. Krupnikov, M. Levendusky and J. B. Ryan. 2019. "The illusion of affective polarization." *See https://www. ipr. northwestern. edu/documents/working-papers/2019/wp-19-25. pdf* .

Enders, Adam M and Miles T Armaly. 2019. "The differential effects of actual and perceived polarization." *Political Behavior* 41(3):815–839.

Fitzpatrick, Sheila. 1996. "Signals from below: Soviet letters of denunciation of the 1930s." *The Journal of Modern History* 68(4):831–866. Publisher: The University of Chicago Press.

Frimer, Jeremy A, Linda J Skitka and Matt Motyl. 2017. "Liberals and conservatives are similarly motivated to avoid exposure to one another's opinions." *Journal of Experimental Social Psychology* 72:1–12.

Frye, Timothy. 2010. *Building states and markets after communism: The perils of polarized democracy*. Cambridge University Press.

Geddes, Barbara, Joseph George Wright, Joseph Wright and Erica Frantz. 2018. *How dictatorships work: Power, personalization, and collapse*. Cambridge University Press.

Gelpi, Christopher, Peter D Feaver and Jason Reifler. 2009. Paying the human costs of war. In *Paying the Human Costs of War*. Princeton University Press.

Gidron, Noam, James Adams and Will Horne. 2020. *American affective polarization in comparative perspective*. Cambridge University Press.

Howell, Jennifer L. and James A. Shepperd. 2016. "Establishing an information avoidance scale." *Psychological Assessment* 28(12):1695. Publisher: American Psychological Association.

Iyengar, Shanto, Gaurav Sood and Yphtach Lelkes. 2012. "Affect, not ideologya social identity perspective on polarization." *Public opinion quarterly* 76(3):405–431.

Iyengar, Shanto, Yphtach Lelkes, Matthew Levendusky, Neil Malhotra and Sean J Westwood. 2019. "The origins and consequences of affective polarization in the United States." *Annual Review of Political Science* 22(1):129–146.

John, Leslie K., Kate Barasz and Michael I. Norton. 2016. "Hiding personal information reveals the worst." *Proceedings of the National Academy of Sciences* 113(4):954–959. Publisher: Proceedings of the National Academy of Sciences.
URL: *https://www.pnas.org/doi/abs/10.1073/pnas.1516868113*

Le Forestier, Joel M., Elizabeth Page-Gould and Alison L. Chasteen. 2022. "Concealment stigma: The social costs of concealing." *Journal of Experimental Social Psychology* 101:104340.
URL: *https://www.sciencedirect.com/science/article/pii/S0022103122000592*

Lees, Jeffrey and Mina Cikara. 2020. "Inaccurate group meta-perceptions drive negative out-group attributions in competitive contexts." *Nature Human Behaviour* 4(3):279–286.

List, John A. 2007. "On the interpretation of giving in dictator games." *Journal of Political economy* 115(3):482–493.

Media, Holod. 2022. "Poshla ty... [Go to...]." *Holod Media* . Retrieved 2022-07-05.
URL: *https://holod.media/2022/02/25/war-family/*

Medialeaks. 2022. "Kak rossiyane donosat... [How Russians inform...]." *Medialeaks* . Retrieved 2022-07-05.
URL: *https://medialeaks.ru/2303tat-str-donosi/*

Meduza. 2022. "V Penze shkolniki... [Schoolchildren in Penza...]." *Meduza* . Retrieved 2022-07-05.
URL: *https://meduza.io/news/2022/04/01/v-penze-shkolniki-napisali-donos-na-uchitelnitsu-rasskazavshuyu-o-voyne-v-ukraine-na-pedagoga-zaveli-ugolovnoe-delo-po-statie-o-feykah*

Murray, Ian, Anke Plagnol and Philip Corr. 2017. "'When things go wrong and people are afraid': An evaluation of group polarisation in the UK post Brexit." *Available at SSRN 3041846* .

Nagel, Thomas. 2004. *Concealment and exposure: and other essays*. Oxford University Press.

Realii, Sever:. 2022. "Putin, poshel... [Putin, go to...]." *Sever: Realii* . Retrieved 2022-07-05.
URL: *https://www.severreal.org/a/putin-poshel-von-iz-ukrainy-kak-pensionerka-donesla-na-pensionerku/31828644.html*

Roberts, Annabelle R., Emma E. Levine and Ovul Sezer. 2021. "Hiding success." *Journal of Personality and Social Psychology* 120:1261–1286. Place: US Publisher: American Psychological Association.

Settle, Jaime E and Taylor N Carlson. 2019. "Opting out of political discussions." *Political Communication* 36(3):476–496.

Shirikov, Anton. 2021. "Fake news for all: misinformation and polarization in authoritarian regimes." *Available at SSRN 3944011* .

Silver, Ike and Alex Shaw. 2022. "When and why "staying out of it" backfires in moral and political disagreements." *Journal of Experimental Psychology: General* 151:2542–2561. Place: US Publisher: American Psychological Association.

Yair, Omer. 2020. "A note on the affective polarization literature." *Available at SSRN 3771264* .

Online appendix for "Hiding one's views: Does pro/anti social behavior depend on who discloses information on group identity?"

Contents

A Data collection and experimental protocol

A.1 The *Toloka* platform

The platform allows to quickly recruit participants for doing small tasks online and processing payments afterwards both in the form of a fixed participation fee and a variable bonus. The tasks vary from image labelling and tagging texts for natural language processing to participating in surveys or behavioral games. A single *task* can be fulfilled by many workers, resulting in several *assignments* for each task. To create a task, an experimenter must first create a *project*: an interface through which Toloka users will communicate, using a code in HTML or JSON format, that usually provides a participant to the link that follows to the survey.

As soon as a project is created, participants are invited to a specific study though opening a *pool*. A pool is a combination of settings, such as a participation fee, number of participants, and some filters that limit access to the study to a specific audience. These filters can be either built-in (such as a region by IP, a participant's country by their registered phone number, and their self-reported nationality, age, gender, educational level and knowledge of languages). Additionally experimenters can assign custom skills to create population subsets that meet any requirements of the researcher. Using this feature we created a custom *skill* that allowed us to invite to our main study only a specific subset of participants based on their answers in the prescreener (see below).

Based to their own reported numbers, Toloka audience has about 245000 active participants in more than 200 countries (Toloka, 2022). Independent estimates evaluate the number of active participants (being online and ready to accept tasks) at from about 6000 during the night to about 20000 active users during working hours (9 AM to 20 PM). About 60% of this population is Russian-speaking (rising to 72% during daytime), with over 80% of them located in Russia. Using the previous large-scale study of Toloka, its Russian-speaking audience is slightly disbalanced towards males (58%) (Chapkovski, 2022*b*).

A.2 Subject recruitment and payoff calculations

In June 1, 2022 we conducted a short pre-screening survey for 5000 participants from the Toloka. There were three filters that limited the participation in the pre-screener: knowledge of Russian language, location in Russia (by IP address), location in Russia (by the self-declared registration form in Toloka). It took on average 1 minute 7 seconds for participants to complete the survey for which they were paid 0.03$. The data was collected in 2 hours and 5 minutes after the start; 23 participants started the survey and then returned the task.

The pre-screening survey consists of the following questions:

Screen 1a. Please tell me whether or not you support the actions of Russian forces in Ukraine (Support/Do not support)

Screen 1b. Please tell me how much you [do not] support the actions of Russian forces in Ukraine? (Definitely [do not] support/Rather [do not] support)

Screen 2. Out of 100 participants taking part in this study, how many do you think **support** the actions of Russian forces in Ukraine?[4]

Screen 3. Please indicate your gender (Male/Female)

Screen 4. If you read this carefully, please select the "Rather agree" (Strongly disagree/Rather disagree/Partly agree, partly disagree/Rather agree/Strongly agree)

Screen 5. Are you currently employed? (Yes/No)

Screen 6. We will invite some of the participants in this study to continue. If you would like us to invite you, please check this box [ ]

The data for all experimental sessions were collected the same day as the prescreening survey, June 1, 2022. We randomly invited participants from the prescreening survey who fulfilled three conditions:

- Accepted to be invited to the main study (without knowing the essense of the study at that point)
- Passed the attention check (Screen 4 above)
- Those who were invited to the sessions for Dictator's role, answered "Support" at the question about their support of the military actions of Russia in Ukraine.

Those who were invited and accepted the invitation for the main part were first shown a brief version of the Informed consent, and were provided with a link that leads to the oTree based server (Chen, Schonger and Wickens, 2016). Upon the completion of the experiment they received a unique participant code to insert at Toloka server, that was used as a confirmation of their participation, so we could transfer their participation fees and additional bonuses.

[4]This question was shown to half of the participants. The other half received an inverted question: Out of 100 participants taking part in this study, how many do you think **do not support** the actions of Russian forces in Ukraine?

A.3 Experimental protocol

Screen 1. Informed Consent

This study is being conducted by a group of independent researchers. If you have any questions, you can always contact us at: XXXX.

Please read the following information carefully:

You are invited to take part in a study that studies how people make decisions in groups. We will ask you to answer a series of questions and make several decisions that may affect your bonus and the bonuses of other participants in this study. The following are important details that you need to know before proceeding with this study.

You must be over 18 years of age to participate in the study.

Your participation is voluntary and you can withdraw from the study at any time.

Your participation in the study will not incur any financial costs on your part.

Some tasks may provide a monetary reward (bonus).

TASK VERIFICATION TIME: It can take us up to three business days to collect all the data, check your answers and calculate bonuses.

If you agree to participate in the study, you are expected to fulfill the obligations associated with the study: you will be required to answer the questions posed to you during the entire period of the study.

This study does not involve any physical risks. The tasks do not require any special physical or psychological skills, nor any special knowledge.

During the course of the research, we may ask you to provide some personal information, such as your gender, educational level, personal income level, etc.

CONFIDENTIALITY: The information you provide will be kept completely confidential. No information that identifies you personally will be presented in published or unpublished works. Data that does not allow identification of a person may be published in the public domain.

If you have any questions regarding this study, please contact us at XXXX

If you agree to the terms of this study and wish to participate, please check the boxes below and click "Next".

I confirm my consent

I understand that it may take several business days to review the assignment.

Screen 2a. Welcome!

Before proceeding with the study, we will ask you to answer a few questions.

Click "Next" to continue.

Screen 2b. Please tell me whether or not you support the actions of Russian forces in Ukraine (Support/Do not support)

Screen 2c. Please tell me how much you [do not] support the actions of Russian forces in Ukraine? (Definitely [do not] support/Rather [do not] support)

Screen 3a. Please read this manual carefully, because your bonus depends on understanding it!

There are TWO types of participants in this study: **Participants A**, whose decisions may affect the bonus of other participants and their own bonus, and **Participants B**, who do not make any decisions that may affect the bonus of other participants and their own bonus. Each participant A is paired with a randomly selected participant B.

You are Participant [A/B].

In this study, Participant A will have to make a decision that will affect his bonus and the bonus of Participant B, with whom he (or she) is paired.

Screen 3b. Please read this manual carefully, because your bonus depends on understanding it!

The decision phase goes like this:

- Participant A is paired with a randomly selected participant B.
- Both of you are allocated $ 0.50.
- Exhibitor A is allocated an additional $ 0.50 for a total of $ 1.00. Participant B does not receive an additional $ 0.50.

Participant A will need to make **one** decision regarding the money received. He or she may decide:

- **Take** up to $ 0.50 from member B and add it to your bonus;
- **Leave** the money allocated to both participants unchanged;
- **Give** up to $ 0.50 of the received money to participant B.

Bonus table

	Take	**Take**	**Take**	**Take**	**Take**	**Leave as is**	**Give**	**Give**	**Give**	**Give**	**Give**
Participant A's decision	**-$0.50**	**-$0.40**	**-$0.30**	**-$0.20**	**-$0.10**	**$0.00**	**$0.10**	**$0.20**	**$0.30**	**$0.40**	**$0.50**
Participant A's bonus	$1.50	$1.40	$1.30	$1.20	$1.10	$1.00	$0.90	$0.80	$0.70	$0.60	$0.50
Participant B's bonus	$0.00	$0.10	$0.20	$0.30	$0.40	$0.50	$0.60	$0.70	$0.80	$0.90	$1.00

Screen 4a. Participant A decides to give 50 cents to Participant B, how much will each of you end up with?

Participant A: ($0.00 $0.50 $1.00 $0.00)

Participant B: ($0.00 $0.50 $1.00 $0.00)

Screen 4b. Participant A decides to to leave the amounts he and Participant B received unchanged, how much will each of you end up with?

Participant A: ($0.00 $0.50 $1.00 $0.00)

Participant B: ($0.00 $0.50 $1.00 $0.00)

Screen 4c. Participant A decides to take away 50 cents from Participant B, how much will each of you end up with?

Participant A: ($0.00 $0.50 $1.00 $0.00)

Participant B: ($0.00 $0.50 $1.00 $0.00)

Screen 4d. You have successfully passed the comprehension test. Click "Next" to proceed to the main part of the study.

Screen 5a (DR treatment - dictator). You now have to decide if **you want to know** how participant B with whom you have been paired answered the question

Please tell me whether or not you support the actions of Russian forces in Ukraine.

Do you want to know the answer to this question? (Yes/No).

NOTE: In any case, **only you** will be able to see participant B's answer to this question. Participant B could not see your answers to any of the questions.

Screen 5b (DR treatment - dictator). You decided **not to learn** how participant B with whom you have been paired answered the question

Please tell me whether or not you support the actions of Russian forces in Ukraine.

Participant B answered [**"Support"/"No not support"**] to the question:

Please tell me whether or not you support the actions of Russian forces in Ukraine.

NOTE: In any case, **only you** will be able to see participant B's answer to this question. Participant B could not see your answers to any of the questions.

Screen 5b' (DR treatment - dictator). You decided **not to learn** how participant B with whom you have been paired answered the question

Please tell me whether or not you support the actions of Russian forces in Ukraine.

Screen 5 (RR treatment - recipient). You now have to decide if participant A with whom you have been paired might learn how you replied to the question about *your support for the actions of Russian forces in Ukraine.*

For every "No" answer we increase our donations to the Podari Zhizn foundation, specializing in helping sick children, by 10 cents.

Do you want Participant A, with whom you have been paired, to learn your answer to that question? (Yes/No)

If you choose **"No"** we guarantee that Participant A does not see your answer. If you choose **"Yes"**, then Participant A **will see** your answer.

Screen 5a (RR treatment - dictator). Participant B, with whom you have been paired, had an opportunity to tell you how he (or she) answered the question about *his or her support for the actions of Russian forces in Ukraine.*

If Participant B **agreed** to tell her answer, you will see the answer on the next screen.

If Participant B **declined** to given an answer, then on the next page you will see the refusal message.

NOTE: In any case, **only you** will be able to see participant B's answer to this question. Participant B could not see your answers to any of the questions.

Screen 5b (RR treatment - dictator, answer provided by R). Participant B, with whom you have been paired, **told you** his or her answer.

He (she) answered [**"Support"/"No not support"**] to the question:

Please tell me whether or not you support the actions of Russian forces in Ukraine.

NOTE: In any case, **only you** will be able to see participant B's answer to this question. Participant B could not see your answers to any of the questions.

Screen 5b' (RR treatment - dictator, answer not provided by R). Participant B, with whom you have been paired, **refused** to tell you his or her answer to the question:

Please tell me whether or not you support the actions of Russian forces in Ukraine.

Screen 5a (FR treatment - dictator) Computer will decide at random whether to tell you how Participant B, with whom you have been paired, answered the question on the support for actions of Russian forces in Ukraine.

If the computer decides to **tell you** the response, you will see the **response** on the next screen.

If the computer decides **not to provide** you with the response, then on the next page you will see the refusal message.

NOTE: In any case, **only you** will be able to see participant B's answer to this question. Participant B could not see your answers to any of the questions.

Screen 5b (FR treatment - dictator, answer provided by the computer). Computer decided **to tell you** the answer of Participant B, with whom you have been paired.

He (she) answered [**"Support"/"No not support"**] to the question:

Please tell me whether or not you support the actions of Russian forces in Ukraine.

NOTE: In any case, **only you** will be able to see participant B's answer to this question. Participant B could not see your answers to any of the questions.

Screen 5b' (FR treatment - dictator, answer not provided by the computer). Computer decided **not to tell you** the answer of Participant B, with whom you have been paired, to the question:

Please tell me whether or not you support the actions of Russian forces in Ukraine.

Screen 6 (dictator) [THE CONTENTS OF Screen 5b or 5b' REPEATED HERE]

Now you need to make a decision about the upcoming bonus and the bonus contained in B.

- $1.00 available.
- Participant B received $0.50.

Your decision:

- You can donate up to $0.50 of your earnings to member B (select the appropriate benefit or prize, zero below).
- You can take up to $0.50 from the winning amount (select any winning amount or zero below).

Your decision: *-$0.50 -$0.40 -$0.30 -$0.20 -$0.10 $0.00 $0.10 $0.20 $0.30 $0.40 $0.50*

Screen 7a. On the following page we will ask you to make several guesses about other participants of this study.

After the data collection is complete, we will gather all responses and check whether your guesses were correct. We will choose **one** of your answers at random, and add the payment to your bonus.

Screen 7b. Out of 100 participants taking part in this study, how many do you think **support** the actions of Russian forces in Ukraine?

(If your response will differ from the true value by no more than 10%, you will receive an additional bonus of $0.25)

Screen 7b (recipients). What do you think will be the decision of Participant A who is paired with you? (-0.5$/-0.4$ $\cdots$ /0.4$ /0.5$).

(If your response will differ from the true value by no more than **10 cents***, you will receive an additional bonus of $0.25)*

Screen 7c (recipients). What do you think will motivate Participant A when deciding whether to give or take money from Participant B (you)?

Screen 7c (dictators). At the decision making stage, which decision will the majority of Participants A choose? (-0.5$/-0.4$ $\cdots$ /0.4$ /0.5$). (If your response will differ from the true value by no more than 10 cents, you will receive an additional bonus of $0.25)

Screen 7d (dictators and recipients, RR treatment). Out of 100 participants B taking part in this study and supporting the actions of Russian forces in Ukraine, how many do you think agreed to tell Participant A whether he (or she) support the actions of Russian forces in Ukraine? (If your

response will differ from the true value by no more than 10%, you will receive an additional bonus of $0.25)

Out of 100 participants B taking part in this study and not supporting the actions of Russian forces in Ukraine, how many do you think agreed to tell Participant A whether he (or she) support the actions of Russian forces in Ukraine? (If your response will differ from the true value by no more than 10%, you will receive an additional bonus of $0.25)

Screen 7e (dictators). Please remember your decision to give or take money from participant B, with whom you were paired. What guided you in making your decision?

Screen 8a. Recall the following question:

Please tell me whether or not you support the actions of Russian forces in Ukraine.

Please choose a position that characterizes you best. (I would avoid learning what would **my friend** thinks regarding this question/Even if it will upset me, I want to know what **my friend** thinks regarding this question.)

Please choose a position that characterizes you best. (I would avoid learning what would **my colleague** thinks regarding this question/Even if it will upset me, I want to know what my colleague thinks regarding this question.)

Please choose a position that characterizes you best. (I would avoid learning what would **a person I don't know well** thinks regarding this question/Even if it will upset me, I want to know what **a person I don't know well** thinks regarding this question.)

Screen 8b. Recall the following question:

Please tell me whether or not you support the actions of Russian forces in Ukraine.

How comfortable are you talking politics with someone who supports the actions of Russian forces in Ukraine. (Not at all comfortable/Not too comfortable/Somewhat comfortable/Extremely comfortable.)

Suppose a son or daughter of yours was getting married. How would you feel if he or she married someone who supports the actions of Russian forces in Ukraine. (Not all all upset/Not too upset/Somewhat upset/Extremely upset.)

Screen 8c. Recall the following question:

Please tell me whether or not you support the actions of Russian forces in Ukraine.

How comfortable are you talking politics with someone who does not support the actions of Russian forces in Ukraine. (Not at all comfortable/Not too comfortable/Somewhat comfortable/Extremely comfortable.)

Suppose a son or daughter of yours was getting married. How would you feel if he or she married someone who does not support the actions of Russian forces in Ukraine. (Not all all upset/Not too upset/Somewhat upset/Extremely upset.)

Screen 8d. Please indicate how much you are willing to take risks. Please choose your answer on a 0-10 scale, where 0 means "not willing to risk at all", and 10 means "very willing to risk".

In general (*0/1/2/3/4/5/6/7/8/9/10*)

In financial matters (*0/1/2/3/4/5/6/7/8/9/10*)

When dealing with strangers (*0/1/2/3/4/5/6/7/8/9/10*)

Screen 8e. Please indicate your age

What is the highest level of education or the highest degree you obtained? (Secondary/Vocational or technical/Incomplete higher/Higher/Multiple degrees or doctoral)

What is your gender (Male/Female.)

Please mark everything that applies to you: (Full employment/Married/Retired/Student/State employee).

How would you best describe your family's financial condition? (We do not have enough money even to buy food/We have enough to buy food, but not to buy clothing or footwear/We have enough for clothing and footwear, but not for small household appliances/We can afford small purchases, but larger ones (computer, washing machine, refrigerator) requires a loan/We can afford to purchase things for our home, but need to save or take a loan to buy a car, dacha, or apartment/We can anything we need without taking a loan.)

Screen 8f. Thank you for your participation in the study!

What do you think is the purpose of this study?

How clear and understandable were the instructions for you? (indicate your answer in the range from 1 = not at all clear to 5 = absolutely clear)

B Additional analysis

B.1 List of variables used

1. **War: Support:** 1 if answered "Support" to "Please tell me whether or not you support the actions of Russian forces in Ukraine", 0 if answered "Do not support".
2. **War: Definitely:** 1 if answered "Definitely [do not] support" to "Please tell me how much you [do not] support the actions of Russian forces in Ukraine?", 0 if "Rather [do not] support".
3. **Beliefs: Own D transfer:** "What do you think will be the decision of Participant A who is paired with you?", -50 to 50
4. **Beliefs: Other D transfers:** "At the decision making stage, which decision will the majority of Participants A choose?", -50 to 50
5. **Beliefs: Support war:** "Out of 100 participants taking part in this study, how many do you think support the actions of Russian forces in Ukraine?", 0 to 100.
6. **Beliefs: Pro/anti war reveal:** "Out of 100 participants B taking part in this study and [supporting/not supporting] the actions of Russian forces in Ukraine, how many do you think agreed to tell Participant A whether he (or she) support the actions of Russian forces in Ukraine?", 0 to 100.
7. **Information avoidance scale:** This is the average (between 0 and 1) of the number of times the respondent chose to avoid learning the response to the question on war support ("I would avoid learning what would [my friend/my colleague/a person I don't know well] thinks regarding this question/Even if it will upset me, I want to know what [my friend/my colleague/a person I don't know well] thinks regarding this question."
8. **Risk scale:** This is the average (normalized to between 0 and 1) of the three risk avoidance scales ("Please indicate how much you are willing to take risks [in general/in financial matters/when dealing with strangers]")
9. **SDI: Pro/contra:** Between 0 and 1. Average of responses to two questions: "How comfortable are you talking politics with someone who [supports/does not support] the actions of Russian forces in Ukraine (1-Not at all comfortable/Not too comfortable/Somewhat comfortable/0-Extremely comfortable)", "Suppose a son or daughter of yours was getting married. How would you feel if he or she married someone who [supports/does not support] the actions of Russian forces in Ukraine (0-Not all all upset/Not too upset/Somewhat upset/1-Extremely upset.)"
10. **Age:** "Please indicate your age", 18 to 99
11. **Education:** "What is the highest level of education or the highest degree you obtained?" 0-Secondary/0.25-Vocational or technical/0.5-Incomplete higher/0.75-Higher/1-Multiple degrees or doctoral.

12. **Income:** "How would you best describe your family's financial condition?" 0-We do not have enough money even to buy food/0.2-We have enough to buy food, but not to buy clothing or footwear/0.4-We have enough for clothing and footwear, but not for small household appliances/0.6-We can afford small purchases, but larger ones (computer, washing machine, refrigerator) requires a loan/0.8-We can afford to purchase things for our home, but need to save or take a loan to buy a car, dacha, or apartment/1-We can anything we need without taking a loan.

13. **Full employment/Married/Retired/Student/State employee:** 1 if mentioned when "please mark everything that applies to you", 0 if not mentioned

B.2 Balance tests

In Table OA1 we report the summary statistics for the dictators, broken down by the treatment. The right column reports the p-values of one-way ANOVA tests for each variable. The null hypothesis is rejected at 5% and 10% levels only for a single variable. Crucially, we do not find that the intensity of support for war was different across the three treatments.

To test that our covariates were globally balanced, we ran a multinomial logit regression where the dependent variable was the treatment category, on all variables in Table OA1. The chi-square statistic was not significant at conventional levels, confirming our expectation. In Tables OA2 and OA3 we repeat the analysis separately for dictators in forced reveal and recipient reveal treatments, comparing the characteristics of dictators whose recipients supported the war, opposed the war, or had an unknown position. Likewise, for neither of the two treatments, we do not find significant difference between these three categories.

	Total	D: Forced reveal	D: Recipient reveal	p
Age	38.85 (11.43)	38.94 (11.73)	38.77 (11.14)	0.829
Education	2.01 (1.16)	1.95 (1.17)	2.07 (1.15)	0.144
Female	0.48 (0.50)	0.48 (0.50)	0.48 (0.50)	0.971
Income	2.75 (1.07)	2.86 (1.01)	2.65 (1.11)	0.005
Full employment	0.61 (0.49)	0.62 (0.49)	0.61 (0.49)	0.849
Married	0.52 (0.50)	0.51 (0.50)	0.52 (0.50)	0.751
Retired	0.11 (0.31)	0.11 (0.32)	0.11 (0.31)	0.812
Student	0.05 (0.21)	0.05 (0.22)	0.04 (0.19)	0.303
State employee	0.10 (0.30)	0.09 (0.28)	0.12 (0.32)	0.131
D: Definitely support	0.61 (0.49)	0.61 (0.49)	0.62 (0.49)	0.806
N	797	398	399	

Each column reports the means of covariates in each group of dictators. The rightmost column reports the p value for the one-way Anova test comparing the means.

Table OA1: Summary statistics and balance test for dictators, depending on treatment

	Total	R: Oppose	R: Support	R: Unknown	p
Age	38.94 (11.73)	39.32 (12.83)	38.22 (11.49)	39.13 (11.36)	0.765
Education	1.95 (1.17)	2.05 (1.23)	1.97 (1.16)	1.89 (1.16)	0.532
Female	0.48 (0.50)	0.45 (0.50)	0.49 (0.50)	0.49 (0.50)	0.852
Income	2.86 (1.01)	2.88 (1.05)	2.86 (1.02)	2.85 (0.99)	0.961
Full employment	0.62 (0.49)	0.65 (0.48)	0.61 (0.49)	0.61 (0.49)	0.829
Married	0.51 (0.50)	0.43 (0.50)	0.54 (0.50)	0.53 (0.50)	0.207
Retired	0.11 (0.32)	0.16 (0.37)	0.09 (0.28)	0.10 (0.31)	0.224
Student	0.05 (0.22)	0.03 (0.18)	0.06 (0.24)	0.06 (0.24)	0.602
State employee	0.09 (0.28)	0.06 (0.25)	0.11 (0.31)	0.08 (0.28)	0.571
D: Definitely support	0.61 (0.49)	0.60 (0.49)	0.60 (0.49)	0.62 (0.49)	0.943
N	398	93	103	202	

Each column reports the means of covariates in each group of dictators. The rightmost column reports the p value for the one-way Anova test comparing the means.

Table OA2: Summary statistics and balance test for dictators, forced reveal treatment, depending on recipient position

	Total	R: Oppose	R: Support	R: Unknown	p
Age	38.77 (11.14)	38.05 (9.95)	39.68 (11.83)	38.55 (11.27)	0.543
Education	2.07 (1.15)	2.01 (1.21)	2.05 (1.13)	2.11 (1.14)	0.776
Female	0.48 (0.50)	0.47 (0.50)	0.51 (0.50)	0.46 (0.50)	0.680
Income	2.65 (1.11)	2.66 (1.13)	2.46 (1.17)	2.76 (1.05)	0.075
Full employment	0.61 (0.49)	0.58 (0.50)	0.56 (0.50)	0.66 (0.47)	0.125
Married	0.52 (0.50)	0.49 (0.50)	0.48 (0.50)	0.56 (0.50)	0.278
Retired	0.11 (0.31)	0.11 (0.31)	0.14 (0.35)	0.09 (0.29)	0.433
Student	0.04 (0.19)	0.02 (0.15)	0.04 (0.20)	0.04 (0.20)	0.662
State employee	0.12 (0.32)	0.11 (0.31)	0.13 (0.34)	0.12 (0.32)	0.904
D: Definitely support	0.62 (0.49)	0.57 (0.50)	0.60 (0.49)	0.66 (0.48)	0.280
N	399	92	117	190	

Each column reports the means of covariates in each group of dictators. The rightmost column reports the p value for the one-way Anova test comparing the means.

Table OA3: Summary statistics and balance test for dictators, recipient reveal treatment, depending on recipient position

C Factors affecting revealing decision

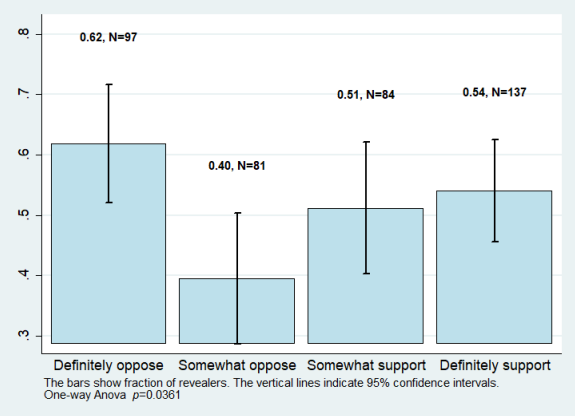

Figure OA1: Share of recipients choosing to reveal one's position, 95% CI)

	(1) Oppose	(2) Oppose	(3) Oppose
Somewhat oppose	-0.223***	-0.215***	-0.211***
	(0.0734)	(0.0753)	(0.0675)
Somewhat support	-0.107	-0.0987	-0.188***
	(0.0735)	(0.0745)	(0.0729)
Definitely support	-0.0784	-0.0862	-0.247***
	(0.0652)	(0.0665)	(0.0739)
IAS (0-1)			0.378***
			(0.0538)
Risk (0-1)			0.268**
			(0.115)
SDI: pro (0-1)			-0.428***
			(0.0981)
SDI: contra (0-1)			-0.0479
			(0.0992)
Demo controls	YES	YES	YES
N	399	399	399
Pesudo R^2	.0165	.0293	.1335

Marginal effects for logit regressions are reported. The dependent variable is the recipient's decision to reveal opinion to the dictator. Demographic controls are as in Table OA1.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table OA4: Recipient decision to reveal one's type

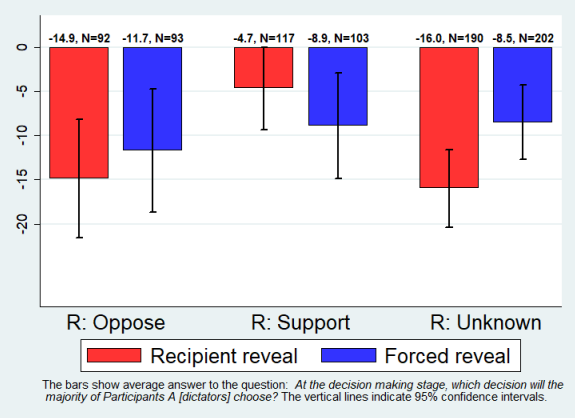

Figure OA2: Beliefs about donations by other dictators, depending on treatment and recipient position, 95% CI)

D Ethics and competing interests statements

The research project was successfully certified by the expedite review procedure of GfeW. The authors have no competing interests to declare.

E Proofs of results

Proof of Proposition 1. Denote

$$H = \frac{(1-p)(1-q)(1-r) + (1-p)qF(\bar{v}_1)}{(1-q)(1-r) + q(pF(\bar{v}_0) + (1-p)F(\bar{v}_1))}.$$

In equilibrium we must have $H = 0$. It must be satisfied for some w, as $H < 0$ for $w = 0$ and $H > 0$ for $w = 1$. Differentiating, we obtain

$$\frac{\partial H}{\partial q} = -\frac{p(1-p)(1-r)(F(\bar{v}_1) - F(\bar{v}_0)) + p(1-p)q(1-q)(1-r)f(\bar{v}_1 - \bar{v}_0) + p(1-p)qf(\bar{v}_1 F(\bar{v}_0) - \bar{v}_0 F(\bar{v}_1))}{(1-q)(1-r) + q(pF(\bar{v}_0) + (1-p)F(\bar{v}_1)))^2}.$$

This value is negative as $\bar{v}_0 < 0$. Now, $H < 0$ if $w = 0$. So, in the generic case we must have $\frac{\partial H}{\partial w} > 0$ if w is the smallest value satisfying $H = 0$. From the implicit function theorem it follows that $\frac{\partial w}{\partial q} > 0$ and, hence, $\frac{\partial x_N}{\partial q} < 0$. **Q.E.D.**

Appendix references

Chapkovski, Philipp. 2022. "Interactive experiments in Toloka.". Publisher: Rochester: SSRN.

Chapkovski, Philipp and Max Schaub. 2022. "Solid support or secret dissent? A list experiment on preference falsification during the Russian war against Ukraine." *Research & Politics* 9(2):20531680221108328.

Toloka. 2022. "Toloka for social sciences." *Toloka* . Retrieved 2022-07-05.
URL: *https://toloka.ai/toloka-for-social-sciences/*