Registered Report: How Language Shapes Belief in Misinformation: A Study Among Multilingual Speakers in Ukraine

Abstract

Scholarship has identified key determinants of people's belief in misinformation, but our knowledge predominantly comes from English-language misinformation in the United States. However, in the global media environment, multilingual citizens often consume media in more than one language. To this end, we ask how the language in which misinformation is consumed affects belief in misinformation in multilingual media environments. We suggest that language may pass on specific cues that may affect how bilinguals evaluate misinformation in their less preferred language. In a ten-week survey experiment with bilingual adults in Ukraine, we plan to measure if subjects who are asked to evaluate misinformation in their less-preferred language are less likely to believe it.

1 Introduction

Our cumulative knowledge about belief in misinformation predominantly comes from surveying English-speaking Americans about misinformation written in English from American media sources (Allcott and Gentzkow 2017; Pennycook and Rand 2020; Clayton et al. 2019; Pennycook, Cannon, and Rand 2018).[1] However, the global media environment is complexly multilingual. Half of the global population uses two or more languages or dialects in their daily life (Ansaldo et al. 2008; Grosjean 2010) and, therefore, likely consumes media, including misinformation, in multiple languages from both within and outside of their borders. As the consumption of cross-border media increases and multilingual media with it (PwC-UK 2016), the distribution of false or misleading news in different languages poses substantial political consequences. Reporting news in different languages has the capacity to differentially mobilize populations (Onguny 2019), sometimes based on false or misleading evidence that escalates political violence (Ismail and Deane 2008). Differential belief in misinformation strengthens ethno-political divides (Somerville 2009) and increases levels of affective polarization (Lau et al. 2017; Suhay, Bello-Pardo, and Maurer 2018; Stewart, Plotkin, and McCarty 2021), which weaken the foundations of liberal democracy (Kuklinski et al. 2000). If misperceptions are based on ethnic divisions, policies advocated by politicians will likely accentuate ethnic and economic inequality and may, in some cases, lead to the disenfranchisement of minority groups. Given both these troubling consequences for democracy and the rise of multilingual speakers in a single media market consuming news in different languages, it is imperative we develop a more comprehensive understanding of how news consumers perceive misinformation in different languages. In this registered report we propose a study that addresses the broad research question: Are individuals more or less susceptible to misinformation in their non-preferred language?

Previous research on language and belief in misinformation has focused on identifying a language proficiency effect, but we propose that language may affect belief in misinformation even if individuals are similarly proficient in both languages. For example, language may signal *credibility* or elicit an emotional response, which may affect belief in misinformation written in that language. A linguistic minority may be more skeptical of news written in a different language spoken by the majority, because they have become distrustful of media that communicate using that language or because of tensions with the majority linguistic group. Cues unique to a language and context, such as these, may explain why past studies measuring the effect of language proficiency on belief in misinformation have reported inconsistent results across different languages. Rather than a proficiency effect, we argue that there is a general language cue that affects even

[1] There are a few, but growing list of exceptions. (Mujani and Kuipers 2020; Badrinathan 2020; Rosenzweig et al. 2021)

those that are equally proficient in their less-preferred language. Since previous research has found that emotion affects belief in false news, but not true news (Martel, Pennycook, and Rand 2020), we focus on belief in misinformation; however, we also perform an exploratory analysis measuring the effect of language on belief in true news.

We propose testing whether language cues affect belief in misinformation in a country where most individuals are similarly proficient in two languages. To determine whether or not language indeed affects belief in misinformation regardless of proficiency level, we plan to conduct a survey experiment in Ukraine following the full-scale Russian invasion. In our proposed survey experiment, we will randomly assign bilingual respondents to evaluate false and true news stories in either their preferred or less preferred language (Russian or Ukrainian) directly after the publication of an article. We seek to answer one research question: Are multilingual individuals more skeptical of misinformation produced in their less preferred language?

To present our predicted results for this pre-registered experiment, we use data from a 1-week pilot study prior to Russia's full-scale invasion to simulate the results of a full (proposed) 10-week study. Using this pilot data, we present the predicted effects of reading misinformation in one's less preferred language as well as how our results will be presented in the final paper once we complete the full ten-week study. Our simulated data suggests that minority linguistic groups may be more skeptical of misinformation written in a majority language. It also suggests that the language effect of reading misinformation in one's less preferred language is not stronger among Russian-preferred speakers who have more distrust of the Ukrainian central government, but the language effect of reading misinformation in one's less preferred language may be stronger among Ukrainian-preferred speakers who have a more anti-Russian ideology.

This registered report consists of four more sections. The following section presents the theory that motivates our research questions and hypotheses. The final three sections present the experimental design, proposed methods and how we plan to eventually present our results once we run the study. The discussion of the results will be added upon the proposed study's completion.

2 Theory and Hypotheses

Much of the work on language and misinformation has focused on identifying a language proficiency effect by investigating two modes of cognition: an effortless mode based on heuristics and a more reflective mode based on deliberation (Keysar, Hayakawa, and An 2012; Costa, Vives, and Corey 2017; Corey et al. 2017).

Muda et al. (2021) have suggested that individuals who read news in their less proficient language often use their more reflective mode of cognition and rely less on heuristics. Given that higher levels of cognitive reflection decrease individuals' belief in misinformation (Pennycook and Rand 2019), individuals who evaluate misinformation in their less proficient language could be less likely to believe misinformation, but studies testing language's effect on the belief in misinformation report mixed evidence (Fernández-López and Perea 2020; Muda et al. 2021). Inconsistent results suggest that reading news in one's less proficient language may not have the same effect across each language and context. Features of news identified by other work, such as credibility cues (Flanagin and Metzger 2000; Althaus and Tewksbury 2000; Fogg et al. 2001; Flanagin and Metzger 2007) may explain why some individuals are more susceptible to believing misinformation in their non-native language than others. In addition, language can also invoke certain emotions and increase the salience of ethnic divisions (Pérez and Tavits 2019), which may affect belief in misinformation, but not belief in true news (Martel, Pennycook, and Rand 2020). Building on this work, we move past this proficiency argument and focus on the cues that languages pass on. Cues associated with a language can be particularly influential for those reading in their less preferred language. Individuals are less familiar with this language and are, therefore, more likely to rely on these cues when evaluating suspect information. We, therefore, focus on the effect of evaluating misinformation in one's less preferred language relative to their preferred language.

Language may pass on different cues for numerous reasons. For example, language may pass on a cue to the source of information, which could affect belief in misinformation. News consumers often use the source of information to aid their evaluation of news stories' veracity (Sundar and Nass 2001). Moreover, the information source's country of provenance (which is often imparted by language) can change the effect of information on public opinion (Dragojlovic 2015).[2] Language also has particularly strong priming effects (Pérez 2016), particularly on bilingual and bicultural individuals (Ross, Xun, and Wilson 2002; Trafimow et al. 1997).[3] In a political context, speaking in a language associated with a minority population can increase the political salience of ethnic divisions (Pérez and Tavits 2019), and the language of political ads can pass on cues to voters that affect their sentiment towards the targeted political candidate (Flores and Coppock 2018). The effects of language on consuming information clearly depend on the contextual cues given by the current conditions in a country. For example, in Ukraine, our country of study, we argue it is plausible that different language cues could make both the minority and the majority linguistic groups less likely to believe

[2]Unsurprisingly, Dragojlovic (2015) finds that individuals are more likely to be persuaded by new information from foreign sources that they view more positively. Other work has not found any foreign source effect in any direction when individuals have a negative view of the foreign source (Fisher 2020)

[3]Using a different language also has a powerful effect on people's interpretations of the world (Boroditsky 2006), time (Boroditsky 2001) or even the same event (Fausey et al. 2010).

misinformation in their less-preferred language, albeit for different reasons.

Within a country, minority groups (often but not exclusively ethnic groups) with distinct languages are often skeptical of news that is written in the language spoken by the majority, because mainstream news in the majority language often portrays minority groups in a negative light and over-represent the dominant group that prefers to speak the majority language (Keshishian 2000; Mastro 2009; Tukachinsky, Mastro, and Yarchi 2015). This skepticism can push those in the linguistic minority to consume sectoral or extranational media that often use different languages (Tsfati and Peri 2006). Minority groups' divergence in trust can gain prominence during periods of crisis. For example, after the 2014 Ukrainian crisis, trust in majority language news dropped dramatically among minority Russian speakers in Estonia and Latvia (Vihalemm, Juzefovičs, and Leppik 2019). Indeed, it is likely that during a crisis, mainstream news in the majority language can promote separate narratives that alienate minorities already skeptical of news in the majority language. Therefore, in such situations, it is possible that those who prefer to speak a minority language are less likely to believe misinformation if it is written in the majority language spoken (i.e., their less-preferred language). In Ukraine, we can test if this is indeed the case by surveying those that prefer to speak Russian and measure the effect that reading misinformation in their less-preferred language, Ukrainian, has on belief in that misinformation during a crisis in Ukraine (The 2022 invasion of Ukraine by Russia). Although the status of Russian language is changing quickly in Ukraine, and is a debated question, before the 2022 invasion of Ukraine by Russia, over 40% reported Russian as their language of private life and took opinion surveys in Russian, while 60% reported Ukrainian as their preferred language (Onuch and Hale 2018). Generally, almost the entire population of Ukraine reports high reading proficiency in both languages, given their linguistic similarity.[4] Given their minority status in Ukraine, we expect Russian-preferred speakers in Ukraine to be more skeptical of misinformation written in Ukrainian than in Russian.

International conditions could also create language cues for the majority language speakers. For example, a sustained disinformation campaign by a foreign power that employs a domestic minority language could lower belief in misinformation written in a minority language, because it may be associated with other nations and, therefore, less trusted. Russian language news in Ukraine fits these criteria (StratCom 2015): Ukrainian news consumers who are most proficient in Ukrainian (Vorobiov 2015) and normally consume news in Ukrainian may associate news in Russian with a foreign actor. This cue could reduce belief in reported information in that language. Previous research has partially identified this phenomenon and found that those in Ukraine that prefer to speak Ukrainian are less likely to believe popular pro-Kremlin disinformation

[4]In our pilot study, 86% of our respondents living in Ukraine (230 respondents evenly divided Russian and Ukrainian speaking preferences) self-reported equivalent reading proficiency levels in Ukrainian and Russian.

news stories (which are usually but not exclusively written in Russian) than those that prefer to speak Russian (Erlich and Garner 2021).[5]

We have presented reasons to believe that individuals may be less likely to believe misinformation written in one's less preferred language regardless of its majority or minority status, and we set out to test this hypothesis (see below).

This leads us to propose H1:

(H1) Individuals are less likely to believe a false/misleading article written in their less preferred language than in their more preferred language

However, we have posited different mechanisms for belief in misinformation depending on whether an individual prefers to speak a minority or majority language. We, therefore, proceed in three steps.

First, if there is evidence for H1 for the entire population of Ukraine, we investigate whether there is support for our hypotheses both among minority and majority speakers in the country. To do so, we will test H1 with solely *Russian-preferring* respondents and separately with solely *Ukrainian-preferring* respondents.

Second, while we cannot randomly assign mistrust of the central government, we can test some correlational observable implications of our causal mechanism by hypothesizing that among those who prefer Russian, as their distrust in the central government increases, they will be less likely to believe misinformation in Ukrainian relative to Russian.

H2a Among those who prefer the Russian language, the negative marginal effect of reading news in Ukrainian (relative to Russian) on belief in misinformation will be greater as distrust of the central government increases.

Third, we posited that among those who prefer Ukrainian, as animus towards Russia increases,[6] they will believe misinformation in Russian less, relative to misinformation in Ukrainian. Again, we cannot randomly assign animus (nor should we), but we can examine correlational support for our mechanism. Hence we will test:

H2b Among those who prefer the Ukrainian language, the negative marginal effect of reading news in Russian (relative to Ukrainian) on belief in misinformation will be greater as animus

[5]However, this work does not randomly assign respondents to the language in which they consume news stories or test the proficiency aspect of the theory we suppose.

[6]An updated version of how animus towards Russia will be calculated can be found in Section C of the Supplementary Materials.

towards Russia increases.

In addition to testing these hypotheses, we will also run seven exploratory analyses that: (1) determine if the effects we find hold across articles with different ideological perspectives, (2) test each hypothesis using evaluations of true articles, (3) test if there is a measurable difference in the effect among those speaking the majority language or those speaking the minority language, (4) determine if the effects we find hold across all articles, (5) determine if the effects vary over time, (6) determine if the effects we find hold across four major regions of Ukraine, and (7) determine the effect of language on discernment using evaluations of true and false/misleading articles. A full description of these exploratory analyses and their functional forms is located in the Proposed Methods section.

We believe that our eventual findings can extend outside of Ukraine. For example, Spanish speakers in America, like Russianspeakers, may also be skeptical of mainstream news spoken in English for similar reasons, as mainstream news in the United States often portrays Spanish-speaking groups in a negative light and over-represent the dominant group in society (Tukachinsky, Mastro, and Yarchi 2015). Our findings could also extend to other Post-Soviet countries where there is a fear of Russia and the majority language is not Russian, but there are minority groups of ethnic Russians who prefer to speak Russian. Although our results may not extend to every context in which most individuals are bilingual, we believe they can be informative in many.

3 Experimental Design

To test our hypotheses, we will carry out a 10-week survey experiment in Ukraine. Ukraine is an ideal case to test our hypotheses because most of its citizens are bilingual news consumers in Ukrainian and Russian but prefer one language. Each week, we will sample Ukrainian respondents to ensure geographic, ideological, and linguistic balance among our respondents. Information on who we sample can be found in the Supplementary Materials, Section D. Human subjects in our Ukrainian samples will be recruited by *Info Sapiens*.[7]

Previous research measuring belief in misinformation has yet to integrate important findings about how individuals consume misinformation, limiting inference from these studies. Specifically, misinformation is consumed very quickly after publication (Vosoughi, Roy, and Aral 2018; Starbird et al. 2018), but most research asks respondents to evaluate months- or years-old fact-checked misinformation (Bronstein et al.

[7]Section D, Supplementary Materials details the recruitment of respondents.

2019; Clayton et al. 2019; Pennycook and Rand 2020). To address this limitation, we create a transparent, replicable, and pre-registered news article selection process that sources popular false/misleading and true articles within 24 hours of their publication and subsequently distributes the full articles for evaluation to respondents in Ukraine. Our respondents evaluate these popular articles within 48-96 hours of publication. This process ensures that we are measuring the effect of language on belief in popular misinformation in the time period that individuals are most likely to consume this misinformation.

Our selection method also limits potential researcher selection effects. Past studies investigating interventions affecting belief in misinformation have asked survey respondents to evaluate either popular pieces of misinformation chosen by the researcher (Allcott and Gentzkow 2017; Bronstein et al. 2019; Clayton et al. 2019; Pennycook and Rand 2020), or researcher-designed synthetic news articles (Pennycook, Cannon, and Rand 2018; Moravec, Minas, and Dennis 2018). Both methods risk article selection effects and sampling articles that are unlikely to be representative of popularly-consumed misinformation. This introduces limitations for properly quantifying the effect of interventions on misinformation encountered online. Although we, the researchers, decide on the source list from which we sample, the sources on those lists are chosen by other experts, rather than by ourselves, in an attempt to limit any possible researcher-induced bias.

For each of the study's 10 weeks, we will collect and distribute a new group of five articles for each respondent to evaluate in randomized order. Three of these articles come from political websites known to produce low-credibility news; two articles come from mainstream news sources. Section K, Supplementary Materials, fully describes the news sources we select articles from and this selection process.

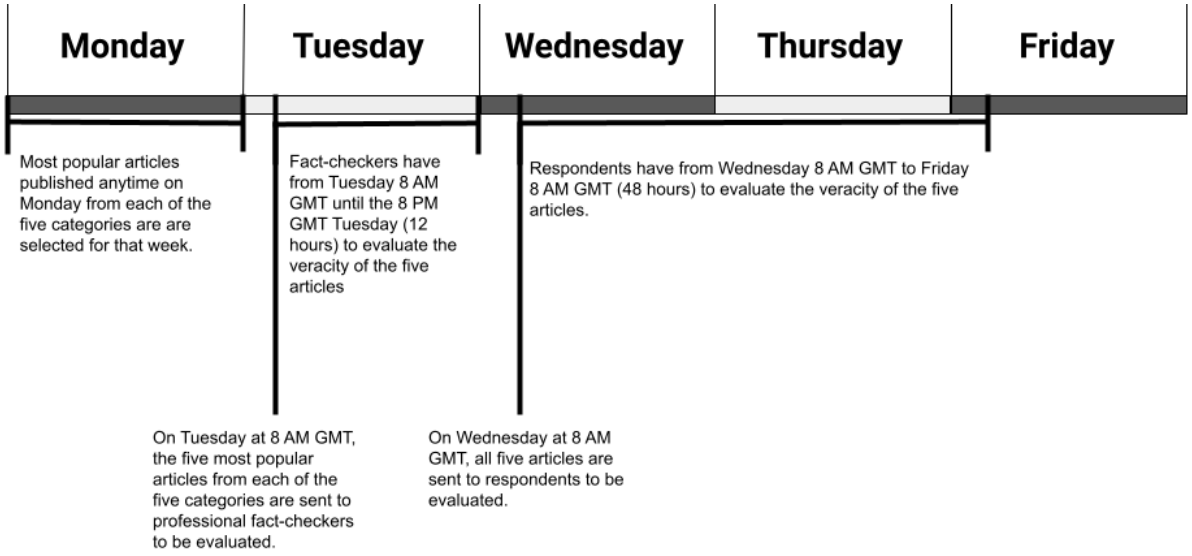
Figure 1 displays the weekly process of selecting articles and distributing these articles to both respondents and professional fact-checkers. Every week, on Tuesday morning, we will source the most popular article published on Monday from each of the five lists of sources.[8] The most popular un-excluded article published in each of the five lists of online news websites is selected and then sent for evaluation to a team of three professional fact-checkers to ascertain whether each article is true, false/misleading, or "could not determine" based on the materials in the article. We use the modal response of the three professional fact-checkers to determine whether an article is coded as true, false/misleading, or 'could not determine'.[9] Then, on Wednesday morning, we will send out the articles to be evaluated by respondents, who will have until Friday

[8]We will exclude some articles. Given that this survey may be distributed in a zone of armed conflict, set forth procedures to avoid "potentially traumatizing topics" (Shesterinina 2019). We will do this by excluding articles that could re-traumatize subjects. We will use our own judgement as well as input from individuals on the ground in Ukraine. If an article is excluded we will move on to the next most popular article in that news feed. The full exclusion protocol can be found in Section A, Supplementary Materials.

[9]Professional fact-checkers rate articles first, so that we can inform our respondents of the veracity of each article they evaluated at the end of the survey.

morning to complete their evaluations. For each article evaluation, we randomly assign the language in which the respondent reads the article (either Ukrainian or Russian).[10] Using this process, respondents evaluate articles within 48-96 hours of publication.[11]

Figure 1: Timeline of Survey Each Week



Given that we focus on belief in misinformation, we only use evaluations of articles that are rated as "false/misleading" by professional fact-checkers. Evaluations of all other articles are not utilized in the main analysis.[12] We leverage the random assignment of the language in which the story is read by respondents, to assign article evaluations to a control or treatment group. Every evaluation of an article read in a respondent's non-preferred language is a 'treated' observation, while each evaluation of an article in a respondent's preferred language is 'not treated.' In table 1, we display the evaluations that we consider in the treatment and control group among Ukrainian-preferring and Russian-preferring speakers.

[10] Articles are translated to Ukrainian or Russian if the article is only available in one language on the website. They are translated by one translator and then checked by another translator.

[11] An example of how the article is viewed by respondents can be found in Section F, Supplementary Materials.

[12] In the exploratory analyses section, we test if the estimated effects hold across articles of different ideological perspectives and true articles.

Table 1: Assignment of Treatment by Preferred Language of Respondent and Language Article is Written In

	Preferred Language of Respondent	
Language of Article Evaluated	Ukrainian	Russian
Ukrainian	Control	Treatment
Russian	Treatment	Control

Given respondent i and news article a, we test each hypothesis using a 4-point ordinal scale ($Y\_ordinal\_4_{ia}$).[13] Additionally, before respondents assess articles, we randomly present half of the respondents with "tips" to help spot false news stories (Section O, Supplementary Materials contains the full set of tips).[14]

Analyzing the "tips" treatment is not a part of this study, but we control for it in robustness checks to ensure that it does not affect our results. We also ask respondents to pass two stand-alone attention checks in the survey prior to experimental manipulation. Removing inattentive respondents threatens internal and external validity (Berinsky, Margolis, and Sances 2014), so we control for their attention in our primary regression models. The text of the attention check questions can be found in Section M of the Supplementary Materials.

To mitigate costs and increase benefits to survey subjects, we debrief survey subjects at the end of the survey to fully explain the design of our experiment, the kind of content individuals were exposed to and why each news article they were exposed to was rated as true or false/misleading by professional fact-checkers. A thorough debriefing of this nature is imperative for survey experiments in armed conflict zones (Baele et al. 2018) and will mitigate any possible effects of showing misinformation to survey subjects and improve their ability to identify misinformation in the future. Research specific to misinformation studies has found that debriefings reduce false beliefs for previously-seen fake stories. In addition, debriefing improves an individual's ability to identify misinformation in the future. Most importantly, any effects of misinformation on planned behaviors were eliminated by the debriefing (Greene and Murphy 2023). In our debriefing, we

[13] Question: "To the best of your knowledge, how accurate is the claim in the above article?
Answers: (A) Very accurate (4); (B) Somewhat accurate (3); (C) Not very accurate (2); (D) Not at all accurate (1)."
Many studies on belief in misinformation, use this measure (e.g., Guess and Munger 2020; Pennycook, Cannon, and Rand 2018; Pennycook and Rand 2019), but as a robustness check, we also test each hypothesis using a categorical measure (Y_true_{ia}) detailed in Section B, Supplementary Materials.

[14] To ensure that the benefits of our survey outweigh any potential costs, we will also expose subjects not initially exposed to these tips at the end of the survey during their debriefing. In conflict areas, it is even more important that we, as researchers, only employ surveys in which the benefits outweigh the costs (Cronin-Furman and Lake 2018; Mazurana, Jacobsen, and Gale 2013) and this increases the benefits to survey subjects.

will follow best practices from Greene and Murphy (2023) and include information about each story that was not true. This message will be reinforced with true information about the topic of the false story that will be written by fact-checkers explaining why the article was false/misleading.

4 Proposed Methods

We use Equation 1 to test H1. Y refers to our main outcome variable — perceived veracity on a four-point ordinal scale in our primary specification for Hypotheses 1 and 2. We consider a sample of individuals randomly drawn from the population to evaluate news articles. Each individual (i) is assigned multiple news articles (a) to be evaluated. The unit of analysis is (ia) — or article evaluated by each individual $\{i, a\}$. The treatment variable (D_{ia}) is a dummy variable (1 if the article evaluated is written in the respondent's less preferred language, 0 otherwise).[15] The τ coefficient represents the estimated treatment effect. K is a vector of article batches, and each k represent one article batch; α_k are the batch coefficients associated with batch fixed effects (B_{ki}). Our covariate associated with preferred language is L_i, and its coefficient is (ν). We specify $\mathbf{Z}_i$ as a collection of additional covariates measured at the respondent level and their associated coefficients γ. These covariates include the attention check, whether respondents were treated with the digital media literacy intervention (orthogonal to this study's treatment), and demographic variables we expect to correlate with our dependent variable: age, gender, income, and education, and region of residence. Given that randomization should make all covariates orthogonal to treatment, we pre-registered the covariates to increase our estimates' precision.[16] For H1, we will use all observations. For H1 as it pertains to our subset of those who prefer Ukrainian and those who prefer Russian, we will separately test two subsets of the data — only those that prefer to speak Ukrainian and only those that prefer to speak Russian — and omit νL_i from Equation 1.

$$Y_{ia} = \sum_{k}^{K} \alpha_k B_{ki} \; + \; \tau D_{ia} \; + \; \nu L_i \; + \; \mathbf{Z}_i \gamma \; + \; \varepsilon_{ia} \tag{1}$$

For H2a and H2b, which we test separately on our two subsets (those who prefer Ukrainian and those who prefer Russian), we include two distinct attitudinal variables (distrust in the central government and

[15] Proficiency is closely-related but distinct from language usage (Andreenkova 2019). We measure language preference (L_i) by assigning each respondent to a preferred language measured by the language in which they choose to take the survey (an exercise in reading that likely mimics print media consumption). However, in robustness check I, we use other self-identified measures of Ukrainian and Russian proficiency (see Section C, Supplementary Materials).

[16] We will present the results of this model unadjusted for covariates in the Supplementary Materials.

anti-Russian attitude),[17] which we label A in Equation 2 and which we interact with treatment. In the full study, we will also examine the linearity of the marginal effect for this and also other models with interaction terms, per Hainmueller, Mummolo, and Xu (2019).

$$Y_{ia} = \sum_{k}^{K} \alpha_k B_{ki} \; + \; \tau D_{ia} \; + \; \lambda A_i \; + \; \zeta(D_{ia} \times A_i) \; + \; \mathbf{Z}_i\gamma \; + \; \varepsilon_{ia} \tag{2}$$

We test our hypotheses using OLS regression models with HC2 robust standard errors in all analyses and report p-values from two-tailed t-tests to predict belief in misinformation (i.e., rating a false or misleading article as true). The treatment (reading the news article in the individual's less-preferred language) is randomly assigned to our unit of observation. Therefore, treatment is not assigned at the cluster unit. Following Abadie et al. (2022, p 31), we do not cluster our standard errors.

For the first two exploratory analyses, we will also use Equation 1 but will estimate coefficients on three subsets for three different ideological perspectives (Exploratory Analysis 1) and subset the data to include only evaluations of true articles (Exploratory Analysis 2). For our third exploratory analysis,[18] we will use Equation 3, which includes an η coefficient on the interaction between our covariate on preferred language (L_i) and our treatment indicator (D_{ia}). To examine the variation of effect sizes across articles (Exploratory Analysis 4), we will modify Equation 1, and replace batch fixed effects with article fixed effects.

$$Y_{ia} = \sum_{k}^{K} \alpha_k B_{ki} \; + \; \tau D_{ia} \; + \; \nu L_i \; + \; \eta(D_{ia} \times L_i) \; + \; \mathbf{Z}_i\gamma \; + \; \varepsilon_{ia} \tag{3}$$

To determine if the effects vary over time, we will inspect whether α_k varies with k (Exploratory Analysis 5). To determine if the effects we find hold across the four major regions, we will utilize equations 1 and 2, but subset the responses by the four major regions (Exploratory Analysis 6). For our final exploratory analysis,[19] we will utilize equations 1 and 2, but substitute the 4-point ordinal scale of belief for a discernment scale (Explanation of this measure can be found in Section B of the Supplementary Materials.) and include all evaluations of false/misleading and true articles.

In our first robustness check we will use other self-identified measures of Ukrainian and Russian proficiency (see Section C, Supplementary Materials). In another robustness check (II), we will run the analysis only

[17] These measures, found in Section C of the Supplementary Materials, have been modified since the pilot given the changing conditions in Ukraine.

[18] Testing if there is a measurable difference in the effect among those speaking the majority language or those speaking the minority language

[19] Determining the effect of language on discernment using evaluations of true and false/misleading articles

including respondents who were not presented with the tips to identify misinformation before they evaluated the articles, which is part of a different study. This treatment is randomly assigned, so it is unlikely to affect results, but we will confirm this is the case. In a third robustness check (III), we will also test each hypotheses using a categorical measure,[20] rather than the 4-point ordinal scale specified earlier.[21] As a fourth robustness check (IV), we will run a restricted analysis with only those who have not seen the story before. Pennycook, Cannon, and Rand (2018) find that prior exposure to misinformation increases belief in it, so running this robustness check will ensure this is not affecting the results we report. In addition, if the sample of respondents who have previously seen the article is large enough we will also provide a restricted analysis with only those who have seen the story previously. In another robustness check (V), we will present all of the main analyses in the paper using only false/misleading articles with a robust mode—which we define as any modal response of fact-checkers that would not change if one professional fact-checker changed their response—to remove articles where there were higher levels of disagreement among professional fact-checkers. As a final robustness check (VI), we will add an order variable to the main specification that will capture the effect, if any, that order may have on rating a news article as true.

5 Simulated Results

To preview what our results may look like, we simulated data for a ten-week study using a pilot study of 230 respondents carried out the week of June 7th, 2021.[22] We randomly sampled from these 230 responses with replacement to create 1,000 simulated datasets of 2,000 hypothetical respondents (200 per day) who evaluated two articles rated as true, one as "could not determine", and two as false/misleading. As specified by our design, we will only use evaluations of the false/misleading articles for the main hypotheses. These articles were the most popular articles from the "Not Anti-Russia low-quality" and "Russia-based mainstream" news streams. The false/misleading article from the "Not Anti-Russia low-quality" news stream described a new tax proposed by Ukrainian President Zelensky that would hurt the Ukrainian economy. The false/misleading article from the "Russia-based mainstream" news stream stated that the Ukrainian soccer team will use a nationalist symbol during the European Soccer Championship that assumes Crimea is a part of Ukraine even though it officially is not.[23]

[20] Even for our dichotomous variable an OLS regression is preferred, because it provides unbiased, reliable estimates of a variable's average effect (Allison 1999; Hellevik 2009; Baetschmann, Staub, and Winkelmann 2015)

[21] Section B and C of the Supplementary Materials contains questions used to measure variables used in all models.

[22] We will not use any of this data in the final analysis, but a balance table for this pilot study is located in Section G, Supplementary Materials. In the final version of the paper, this section will be replaced with one that reports the results from the actual study, as opposed to the simulated study we describe here.

[23] All the articles used in the pilot study are located in Section J, Supplementary Materials

These two stories yield 4,000 simulated evaluations of false/misleading articles, similar to the number we expect to collect during the main study, though we plan to have many more diverse articles during the main study. Using 1,000 simulated datasets of 4,000 evaluations, we run our pre-registered models and present our predicted results for models testing Hypotheses 1, 2a, and 2b and exploratory analyses. Our predicted results for Hypothesis 1 can be found in Figure 2, and results for Hypothesis 2a and 2b can be found in Figure 3. The simulated results we present in this section are suggestive, but we present findings in a fashion we plan to once the full ten-day study is complete.[24] Eventually, we will replace the simulated results in these figures with the results using the data collected during our proposed ten-day study.

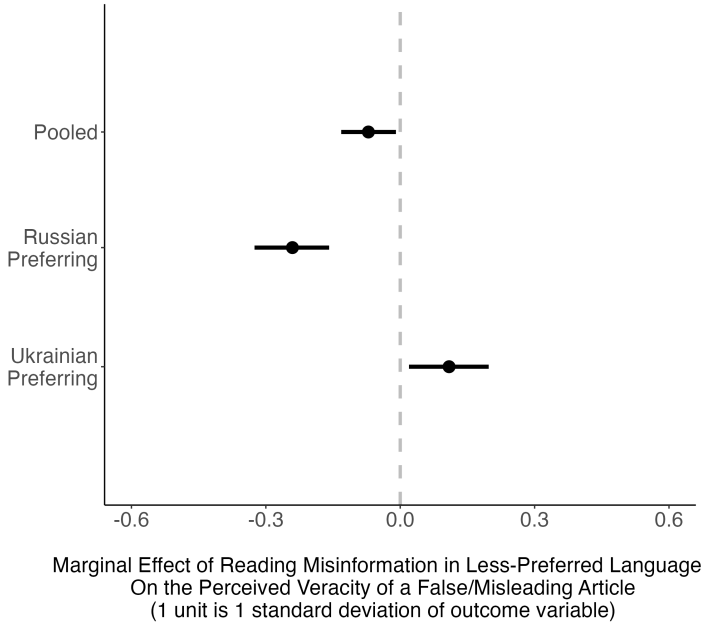
Figure 2 displays results based on data simulated from our pilot study for our model testing Hypothesis 1. Specifically, it presents the estimated effect of reading misinformation written in one's less preferred language on belief of misinformation among all of our respondents (Both those that prefer Ukrainian and those that prefer Russian) as well as among solely those who prefer Russian and solely those who prefer Ukrainian. Each line represents a confidence interval that covers 95 percent of the point estimates reported using these simulated datasets, and the point estimate represents the media simulated effect.[25] Our simulated results suggest weak support for Hypothesis 1 among all the respondents, strong support among Russian-preferred speakers, but no support among Ukrainian-preferred respondents. For all respondents, average belief in misinformation was reduced by -0.07 (95% confidence interval: [-0.002,-.13]) on a 4-point ordinal score when reading it in their less-preferred language. Among Russian-preferred speakers average belief in misinformation was reduced by -0.25 (95% confidence interval: [-0.16,-.33]) on a 4-point ordinal score when reading it in their less-preferred (Ukrainian) language. Among Ukrainian-preferred speakers, we find the opposite effect. On average (median), belief in misinformation increased by 0.11 (95% confidence interval: [0.02,0.21]) on a 4-point ordinal score when Ukrainian-preferred speakers read misinformation in their less-preferred language (Russian).

In addition to these simulated results, we also ran multiple power analyses using `DeclareDesign` (G. Blair et al. 2019) to ensure we would be able to identify marginal effects for Hypothesis 1 using this design. If we are able to sample 200 respondents a day and at least two articles that are rated false/misleading by professional fact-checkers a day as we were in the pilot study, we are confident that we will be able to report an effect size with a Cohen's d of 0.2 or higher (Power > 0.99) for both Ukrainian-speakers and Russian-speakers. Even if we halved our observations, we are still confident that we will be able to report an effect size with a Cohen's d of 0.2 or higher (Power > 0.80). A figure displaying varying levels of power with

[24] The code we wrote in the R statistical program to carry out these analyses on our simulated datasets can be found in Section L of the Supplementary Materials.

[25] We will present coefficients and confidence intervals from our regressions after running the study.

Figure 2: Simulated Treatment Effect of Evaluating *False/Misleading* News Articles in One's Less Preferred Language Across All Respondents (Pooled) Russian-preferring speakers and Ukrainian-preferring speakers



different effect sizes and numbers of observations and a figure displaying varying levels of minimum detectable effects with different numbers of observations at a statistical power of 80% can be found in Section N of the Supplementary Materials. We present a power analysis without covariate adjustments to be conservative since we should have more precision when we run a covariate-adjusted model.

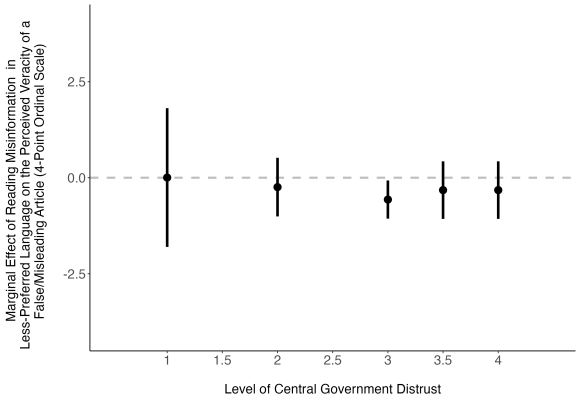
Figure 3a and 3b display our pilot results for our model testing Hypothesis 2a and 2b. Specifically, it presents the marginal effect for reading misinformation written in one's less preferred language on the belief of misinformation on Russian-preferred speakers and Ukrainian-preferred speakers across different levels of central government distrust and anti-Russian ideology respectively. Rather than presenting simulated results, we present results solely from the pilot study because we have so few observations at some levels of our moderators. So, in Figure 3a and 3b each point represents the marginal effect and each line represents the 95 percent confidence intervals. These confidence intervals should be smaller once we run the full study.[26] Our pilot study results provide no substantive support for Hypothesis 2a, but some support for Hypothesis 2b. The language effect of reading misinformation in one's less preferred language is not stronger among Russian-preferred speakers when they have more distrust of the Ukrainian central government (Hypothesis 2a), but the marginal language effect of reading misinformation in one's less preferred language is more negative among Ukrainian-preferring speakers when they have a more anti-Russian ideology (Hypothesis

[26]For the pilot data we show non-linear marginal effects, but we will test for linearity in the main study. Additionally, we do not present marginal effects for levels of the moderators (Central government distrust levels of 1.5 and 2.5 and Anti-Russian attitude levels below 3), because they are too few observations.

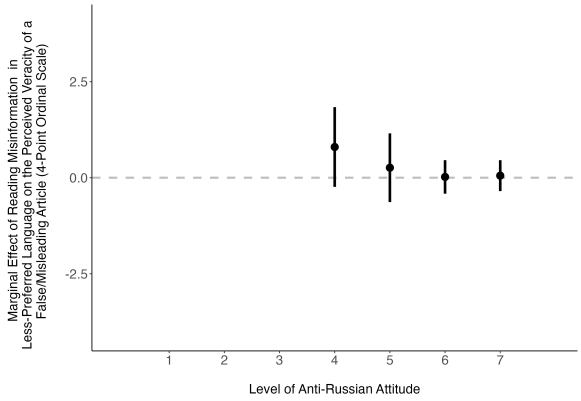
2b).

Figure 3: Marginal effects of evaluating *false/misleading* news articles in one's less preferred language on Russian-preferred speakers and Ukrainian-preferred speakers across different levels of central government distrust and anti-Russian ideology respectively

(a) Marginal effects of evaluating *false/misleading* news articles in one's less preferred language on Russian-preferred speakers across different levels of central government distrust



(b) Marginal effects of evaluating *false/misleading* news articles in one's less preferred language on Ukrainian-preferred speakers across different levels of anti-Russian ideology



5.1 Exploratory Analyses

We do not have sufficient data to present predicted results from Exploratory Analysis 1, but we will present the results after completing the full ten-day survey. In our second exploratory analysis, we present the predicted treatment effect of reading *true* news in their less-preferred language among Ukrainian-preferred and Russian-preferred speakers in Section H of the Supplementary Materials. Our simulated results predict that the effects in support of Hypothesis 1 are **not** present for true articles, but there is support for Hypothesis 2a, but not 2b for true articles.

For our third exploratory analysis, we will strictly report the p-value of the interaction term comparing the language effects of reading misinformation in one's less preferred language on Ukrainian-preferred and Russian-preferred speakers. We found that in almost every simulated analysis, the language effect of reading misinformation in one's less preferred language is stronger for Russian-preferred speakers than Ukrainian speakers and statistically significant at the 95 percent level.

In our fourth exploratory analysis, we present the predicted effects of reading misinformation in one's less preferred language for Hypotheses 1 from our simulated data, for each false/misleading article in Section

H of the Supplementary Materials.[27] This is a disaggregated analysis where we are not concerned about the statistical significance, given the small sample of evaluations per article. Rather we are interested if the marginal effects are in a similar direction across each article. We find in our simluated data that the effects are in same direction for Russian-preferring but not for Ukrainian-preferring respondents for the two articles.[28]

We do not have sufficient data to present the results from Exploratory Analysis 5, but we will present the results after completing the full ten-day survey. For our sixth exploratory analysis, we present simulated results suggesting that the effects of reading misinformation in one's less preferred language for Hypotheses 1 for each region respondents reside in Section H of the Supplementary Materials.[29] Similar to the predicted effects across articles our predicted effects vary across regions.[30]

For our seventh and final exploratory analysis, our simulated results suggest that Russian-preferred respondents may be slightly better able to discern the veracity of news written in Ukrainian, but the effect does not vary across levels of distrust in the central government.[31] Overall, simulated data using our pilot study suggests that linguistic minorities are less likely to believe misinformation written in the majority language, but linguistic majorities are no less likely to believe misinformation written in the minority language. The pilot data (without simulations) also suggests that the language effect of reading misinformation in one's less preferred language may not be stronger among Russian-preferred speakers when they have more distrust of the Ukrainian central government, but the language effect of reading misinformation in one's less preferred language may disappear among Ukrainian-preferred speakers when they have a more anti-Russian ideology.

6 Discussion

This section will be added in the final paper to discuss the implications of our findings and directions for future research.

[27] Due to a low number of observations in our pilot study we do not run our analysis testing Hypotheses 2a and 2b for this exploratory analysis

[28] Figures presenting the coefficients and confidence intervals are available in Section H of the Supplementary Materials. We will explore this issue more fully after we run the study and have all the articles.

[29] Due to a low number of observations in our pilot study we do not run our analysis testing Hypotheses 2a and 2b for this exploratory analysis

[30] Figures presenting the predicted coefficients and confidence intervals are available in Section H of the Supplementary Materials.

[31] Figures presenting the coefficients and confidence intervals are available in Section H of the Supplementary Materials.

7 References

Abadie, Alberto et al. (2022). "When Should You Adjust Standard Errors for Clustering?*". *The Quarterly Journal of Economics* 138.1, pp. 1–35. URL: https://doi.org/10.1093/qje/qjac038.

Allcott, Hunt and Matthew Gentzkow (2017). "Social media and fake news in the 2016 election". *Journal of economic perspectives* 31.2, pp. 211–36.

Allison, Paul D (1999). "Comparing logit and probit coefficients across groups". *Sociological methods & research* 28.2, pp. 186–208.

Althaus, Scott L and David Tewksbury (2000). "Patterns of Internet and traditional news media use in a networked community". *Political communication* 17.1, pp. 21–45.

Andreenkova, Anna V. (2019). "How to Choose Interview Language in Different Countries". In: *Advances in Comparative Survey Methods*. Ed. by Timothy P. Johnson et al. 1st ed. John Wiley & Sons, Ltd, pp. 314–324.

Ansaldo, Ana Inés et al. (2008). "Language therapy and bilingual aphasia: Clinical implications of psycholinguistic and neuroimaging research". *Journal of Neurolinguistics* 21.6, pp. 539–557.

Badrinathan, Sumitra (2020). "Educative Interventions to Combat Misinformation: Evidence from a Field Experiment in India". *American Political Science Review*, pp. 1–17.

Baele, Stephane J et al. (2018). "The ethics of security research: An ethics framework for contemporary security studies". *International Studies Perspectives* 19.2, pp. 105–127.

Baetschmann, Gregori, Kevin E Staub, and Rainer Winkelmann (2015). "Consistent estimation of the fixed effects ordered logit model". *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, pp. 685–703.

Berinsky, Adam J, Michele F Margolis, and Michael W Sances (2014). "Separating the shirkers from the workers? Making sure respondents pay attention on self-administered surveys". *American Journal of Political Science* 58.3, pp. 739–753.

Blair, Graeme et al. (2019). "Declaring and Diagnosing Research Designs". *American Political Science Review* 113 (3), pp. 838–859. URL: https://declaredesign.org/paper.pdf.

Boroditsky, Lera (2001). "Does language shape thought?: Mandarin and English speakers' conceptions of time". *Cognitive psychology* 43.1, pp. 1–22.

— (2006). "Linguistic relativity". *Encyclopedia of cognitive science*.

Bronstein, Michael V et al. (2019). "Belief in fake news is associated with delusionality, dogmatism, religious fundamentalism, and reduced analytic thinking". *Journal of Applied Research in Memory and Cognition* 8.1, pp. 108–117.

Clayton, Katherine et al. (2019). "Real solutions for fake news? Measuring the effectiveness of general warnings and fact-check tags in reducing belief in false stories on social media". *Political Behavior*, pp. 1–23.

Corey, Joanna D et al. (2017). "Our moral choices are foreign to us." *Journal of experimental psychology: Learning, Memory, and Cognition* 43.7, p. 1109.

Costa, Albert, Marc–Llus Vives, and Joanna D Corey (2017). "On language processing shaping decision making". *Current Directions in Psychological Science* 26.2, pp. 146–151.

Cronin-Furman, Kate and Milli Lake (2018). "Ethics abroad: Fieldwork in fragile and violent contexts". *PS: Political Science & Politics* 51.3, pp. 607–614.

Dragojlovic, Nick (2015). "Listening to outsiders: The impact of messenger nationality on transnational persuasion in the United States". *International Studies Quarterly* 59.1, pp. 73–85.

Erlich, Aaron and Calvin Garner (2021). "Is pro-Kremlin Disinformation Effective? Evidence from Ukraine". *The International Journal of Press/Politics*.

Fausey, Caitlin M et al. (2010). "Constructing agency: the role of language". *Frontiers in psychology* 1, p. 162.

Fernández-López, Marıa and Manuel Perea (2020). "Language does not modulate fake news credibility, but emotion does". *Psicológica Journal* 41.2, pp. 84–102.

Fisher, Aleksandr (2020). "Demonizing the enemy: the influence of Russian state-sponsored media on American audiences". *Post-Soviet Affairs* 36.4, pp. 281–296.

Flanagin, Andrew J and Miriam J Metzger (2000). "Perceptions of Internet information credibility". *Journalism & Mass Communication Quarterly* 77.3, pp. 515–540.

— (2007). "The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information". *New media & society* 9.2, pp. 319–342.

Flores, Alejandro and Alexander Coppock (2018). "Do bilinguals respond more favorably to candidate advertisements in English or in Spanish?" *Political Communication* 35.4, pp. 612–633.

Fogg, Brian J et al. (2001). "What makes web sites credible? A report on a large quantitative study". In: *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 61–68.

Greene, Ciara M and Gillian Murphy (2023). "Debriefing works: Successful retraction of misinformation following a fake news study". *PloS one* 18.1, e0280295.

Grosjean, François (2010). *Bilingual*. Harvard university press.

Guess, Andrew and Kevin Munger (2020). "Digital Literacy and Online Political Behavior". *Charlottesville: OSF Preprints. Retrieved April* 13, p. 2020.

Hainmueller, Jens, Jonathan Mummolo, and Yiqing Xu (2019). "How much should we trust estimates from multiplicative interaction models? Simple tools to improve empirical practice". *Political Analysis* 27.2, pp. 163–192.

Hellevik, Ottar (2009). "Linear versus logistic regression when the dependent variable is a dichotomy". *Quality & Quantity* 43.1, pp. 59–74.

Ismail, Jamal Abdi and James Deane (2008). "The 2007 general election in Kenya and its aftermath: The role of local language media". *The International Journal of Press/Politics* 13.3, pp. 319–327.

Keshishian, Flora (2000). "Acculturation, communication, and the US mass media: The experience of an Iranian immigrant". *Howard Journal of Communications* 11.2, pp. 93–106.

Keysar, Boaz, Sayuri L Hayakawa, and Sun Gyu An (2012). "The foreign-language effect: Thinking in a foreign tongue reduces decision biases". *Psychological science* 23.6, pp. 661–668.

Kuklinski, James H et al. (2000). "Misinformation and the currency of democratic citizenship". *The Journal of Politics* 62.3, pp. 790–816.

Lau, Richard R et al. (2017). "Effect of media environment diversity and advertising tone on information search, selective exposure, and affective polarization". *Political Behavior* 39.1, pp. 231–255.

Martel, Cameron, Gordon Pennycook, and David G Rand (2020). "Reliance on emotion promotes belief in fake news". *Cognitive research: principles and implications* 5, pp. 1–20.

Mastro, Dana (2009). "Racial/ethnic stereotyping and the media". *Media processes and effects*, pp. 377–391.

Mazurana, Dyan, Karen Jacobsen, and Lacey Andrews Gale (2013). *Research methods in conflict settings: A view from below*. Cambridge University Press.

Moravec, Patricia, Randall Minas, and Alan R Dennis (2018). "Fake news on social media: People believe what they want to believe when it makes no sense at all". *Kelley School of Business Research Paper* 18-87.

Muda, Rafal et al. (2021). *People Are Worse at Detecting Fake News in Their Foreign Language*. URL: doi:10.31219/osf.io/p8su6.

Mujani, Saiful and Nicholas Kuipers (2020). "Who Believed Misinformation during the 2019 Indonesian Election?" *Asian Survey* 60.6, pp. 1029–1043.

Onguny, Philip (2019). "Electoral violence in Kenya 2007-2008 - the role of vernacular radio". *Journal of African Elections* 18.1, pp. 86–107.

Onuch, Olga and Henry E. Hale (2018). "Capturing ethnicity: the case of Ukraine". en. *Post-Soviet Affairs* 34.2-3, pp. 84–106.

Pennycook, Gordon, Tyrone D Cannon, and David G Rand (2018). "Prior exposure increases perceived accuracy of fake news." *Journal of experimental psychology: general* 147.12, p. 1865.

Pennycook, Gordon and David G Rand (2019). "Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning". *Cognition* 188, pp. 39–50.

— (2020). "Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking". *Journal of personality* 88.2, pp. 185–200.

Pérez, Efrén O (2016). "Rolling off the tongue into the top-of-the-head: Explaining language effects on public opinion". *Political Behavior* 38.3, pp. 603–634.

Pérez, Efrén O and Margit Tavits (2019). "Language heightens the political salience of ethnic divisions". *Journal of Experimental Political Science* 6.2, pp. 131–140.

PwC-UK (2016). "The rise of cross-border news".

Rosenzweig, Leah R et al. (2021). "Happiness and surprise are associated with worse truth discernment of COVID-19 headlines among social media users in Nigeria". *Harvard Kennedy School Misinformation Review*.

Ross, Michael, W. Q. Elaine Xun, and Anne E. Wilson (2002). "Language and the Bicultural Self". *Personality and Social Psychology Bulletin* 28.8, pp. 1040–1050.

Shesterinina, Anastasia (2019). "Ethics, empathy, and fear in research on violent conflict". *Journal of Peace Research* 56.2, pp. 190–202.

Somerville, Keith (2009). "British media coverage of the post-election violence in Kenya, 2007–08". *Journal of Eastern African Studies* 3.3, pp. 526–542.

Starbird, Kate et al. (2018). "Engage early, correct more: How journalists participate in false rumors online during crisis events". In: *Proceedings of the 2018 CHI conference on human factors in computing systems*, pp. 1–12.

Stewart, Andrew, Joshua Plotkin, and Nolan McCarty (2021). "Inequality, identity, and partisanship: How redistribution can stem the tide of mass polarization". *Proceedings of the National Academy of Sciences* 118.

StratCom (2015). *Analysis of Russias Information Campaign against Ukraine.* Tech. rep. NATO StratCom Centre of Excellence.

Suhay, Elizabeth, Emily Bello-Pardo, and Brianna Maurer (2018). "The polarizing effects of online partisan criticism: Evidence from two experiments". *The International Journal of Press/Politics* 23.1, pp. 95–115.

Sundar, S Shyam and Clifford Nass (2001). "Conceptualizing sources in online news". *Journal of communication* 51.1, pp. 52–72.

Trafimow, David et al. (1997). "The Effects of Language and Priming on the Relative Accessibility of the Private Self and the Collective Self". *Journal of Cross-Cultural Psychology* 28.1.

Tsfati, Yariv and Yoram Peri (2006). "Mainstream media skepticism and exposure to sectorial and extranational news media: The case of Israel". *Mass Communication & Society* 9.2, pp. 165–187.

Tukachinsky, Riva, Dana Mastro, and Moran Yarchi (2015). "Documenting portrayals of race/ethnicity on primetime television over a 20-year span and their association with national-level racial/ethnic attitudes".

Vihalemm, Triin, Jānis Juzefovičs, and Marianne Leppik (2019). "Identity and media-use strategies of the Estonian and Latvian Russian-speaking populations amid political crisis". *Europe-Asia Studies* 71.1, pp. 48–70.

Vorobiov, Ievgen (2015). "Why Ukrainians Are Speaking More Ukrainian". *Foreign Affairs*. URL: `https://foreignpolicy.com/2015/06/26/why-ukrainians-are-speaking-more-ukrainian`.

Vosoughi, Soroush, Deb Roy, and Sinan Aral (2018). "The spread of true and false news online". *Science* 359.6380, pp. 1146–1151. URL: `https://science.sciencemag.org/content/359/6380/1146`.

Supplementary Methods and Materials: How Language Shapes Belief in Misinformation: A Study Among Multilingual Speakers in Ukraine

Contents

A Article Exclusion Protocol: 3

B Dependent Variables of Interest 4

C Independent Variables of Interest 6

D Recruitment Information 11

E Regional Divides in Ukraine 12

F Article Format (Ukrainian and Russian) 13

G Balance Table for Pilot Study 15

H Exploratory Analysis 16

H.1 Exploratory Analysis 2: Effect of Language on Belief in True News Articles 16

H.2 Exploratory Analysis 4: Effect Across Articles 17

H.3 Exploratory Analysis 6: Effect Across Four Regions 18

H.4 Exploratory Analysis 7: Substituting Discernment Scale for Veracity Measure 19

I Robustness Checks 20

I.1 Robustness Check 1 - Substituting Language Proficiency for Language Preference 20

I.2 Robustness Check 3 - Replace Dependent Variable with Categorical Measure 22

I.3 Robustness Check 4 - Only Those Who Were Not Exposed To The Article Previously 23

I.4 Robustness Check 6 - Adding the Order Variable to the Model 24

J Articles Evaluated during Pilot Study: 25

K Article Selection Process 26

L Code to Run Models Testing Hypotheses 1–2 (R Statistical Program) 39

M Attention Check Questions 40

N Power Analysis 41

O Treatment: Tips To Spot False News 44

A Article Exclusion Protocol:

Our target content is online written news. Consequently, content that is primarily video, audio, or image based will be excluded. Additionally, content that is explicitly labelled opinion, op-ed, or analysis will be excluded.

Content that primarily concerns sports, satire, celebrity gossip, or other niche content, such as video games, will be excluded from all sources. Any content that is excluded will be recorded, along with the reason(s) for its exclusion and the link to the article made available.

When the researcher selects the most popular articles each day from these streams, they will also determine if the most popular article in that news stream could be potentially harmful. Articles are determined to be potentially harmful if they include any direct reference in the text or accompanying images to death, maiming, sexual violence, torture, or injury or if the misinformation could place people in harm if believed, such as misinformation about evacuation routes and safety. The first author of this paper will read each article and determine if the article meets any of the conditions for exclusion. If so, the article will be withheld, and then the authors will move on to the next most popular article in the news feed.[1]

All articles the author believes do not meet the exclusion criteria (are eligible for selection) will then be sent to three Ukrainian professional fact-checkers. Any of the three fact-checkers can raise a concern about the misinformation causing any emotional or physical harm in addition to the criteria we have created. We will ask them if any are potentially traumatizing.[2] If any are flagged, we will remove the article and move on to the next article.

Some example headlines that would be included or excluded by our criteria are listed below:

Included Headlines (translated):

1. The United States announced a new batch of weapons that will be transferred to Ukraine

2. Within an hour, the Air Force shot down three cruise missiles at once: probably "Caliber" missiles

3. This is our country and our people - Zelensky promised to return Mariupol

4. The United States is secretly buying Russian fertilizer

Excluded Headlines (translated):

1. Armed Forces of Ukraine killed 52 paratroopers from the 31st Special Guards Brigade of the Russian Federation: the battalion commander was also killed

2. Russians intensify the storming of Severodonetsk and suffer losses - General Staff

3. Traitor caught up with karma: Armed Forces of Ukraine killed the commander of a platoon of invaders from Luhansk.

4. In Ukraine, the occupier who threatened to "arrange Bucha" in every city was killed

[1] It is possible that articles about the war pass our exclusion criteria.

[2] The potential to do harm may not be immediately obvious to an external researcher (Fujii 2012; Shanks and Paulson 2022), so it is important that those closest to the situation be consulted.

B Dependent Variables of Interest

Misinformation Rated as True (Y_true_{ia})

The Y_true_{ia} variable is assigned a 1 if an article rated as false/misleading by fact-checkers is rated as true by the respondent. If the article is not rated as false/misleading by fact-checkers this variable is assigned the value of "NA."

Table 1: Veracity Question (Categorical Scale)

Question	Possible Answers
What is your assessment of the central claim in the article?	**(A) True:** The central claim you are evaluating is factually accurate. **(B) Misleading and/or False:** Misleading: The central claim takes out of context, misrepresents, or omits evidence. False: The central claim is factually inaccurate. **(C) Could Not Determine:** You do not feel you can judge whether the central claim is true, false, or misleading.

Perceived Veracity as a Four-Point Ordinal Scale ($Y\_ordinal\_4_{ia}$)

The variable, $Y\_ordinal\_4_{ia}$, is assigned the value of the response chosen from the question below by the respondent (1 through 4).

Table 2: Veracity Question (Ordinal Scale)

Question	Possible Answers
To the best of your knowledge, how accurate is the claim in the above article?	(A) Very accurate (4) (B) Somewhat accurate (3) (C) Not very accurate (2) (D) Not at all accurate (1)

Discernment on a Four-Point Ordinal Scale ($Y\_discernment_{ia}$)

The variable, $Y\_discernment\_4_{ia}$, is assigned the value of the response chosen from the question below by the respondent (1 through 4) for false/misleading articles and true articles.

Table 3: Veracity Question - Ordinal Scale (False/Misleading Articles)

Question	Possible Answers
To the best of your knowledge, how accurate is the claim in the above article?	(A) Very accurate (1) (B) Somewhat accurate (2) (C) Not very accurate (3) (D) Not at all accurate (4)

Table 4: Veracity Question - Ordinal Scale (True Articles)

Question	Possible Answers
To the best of your knowledge, how accurate is the claim in the above article?	(A) Very accurate (4) (B) Somewhat accurate (3) (C) Not very accurate (2) (D) Not at all accurate (1)

C Independent Variables of Interest

Language Preference

We ask the respondent at the beginning of the survey what language they prefer to take the survey in.

We would like you to complete this survey in the language you are most comfortable reading in. Which language is this?
(A) Ukrainian
(B) Russian
(C) Other

The variable, $Pref\_Lang\_UKR_i$, is assigned 1 if the respondent chooses "Ukrainian." The variable, $Pref\_Lang\_RUS_i$, is assigned 1 if the respondent chooses "Russian." The Treatment is $Lang_Treatment$ if the respondent is randomly assigned a news article to evaluate in their non-dominant language.

Age

The Age_i variable is assigned the age of the respondent

Education

The $Education_i$ variable is assigned the education level of the respondent. Possible education levels are as follows: Basic secondary education (up to 9th grade) (1) ; Complete secondary education (grades 10-11) (2) ; Vocational education (vocational school) (3) ; Higher education (bachelor, specialist, master) (4) ; Scientific degree (candidate of sciences and above) (5)

Gender

The $Female_i$ dummy variable is assigned 1 if the respondent self-identifies as female.

Ethnicity

The $Ethnicity_i$ variable. A dummy is assigned 1 if the respondent self-identifies as Ukrainian (multiple allowed).

Ideology of respondent:

In the pre-test, we asked individuals three questions to determine their ideology. The ideology score they received on the ideological scale was an average of their answers in parentheses next to the answer they give.

Given the full-scale invasion, we have changed these questions, the old questions are in red, and the questions that have remained the same or changed are in black. The new questions have an asterisk by them.

Was the dissolution of the USSR a mistake?
(A) Yes (1)
(B) Unsure (0)
(C) No (-1)

What is the best political system of the three listed below?
(A) Political system in Russia (1)
(B) Political system in Ukraine (0)
(C) Political systems in the West (-1)

Which party is closest to you?
(A) Opposition Platform - For Life (1)
(B) Opposition Bloc (1)
(C) For the Future (1)
(D) VO Batkivshchyna (-1)
(E) Holos (-1)
(F) European Solidarity (-1)
(G) Samopomich (-1)
(H) Servant of the People (-1)
(I) Svoboda (0)
(J) Trust (0)
(K) United Centre (0)
(L) Bila Tserkva Together (0)
(M) Other: *textbox* (make the determination based on party)

What is the best of the political systems of the three listed below?*
(A) A political system with a leader who has few checks on his power (1)
(B) A political system with a leader who has some checks on his power (0)
(C) A political system with a leader who has many checks on his power (-1)

Do you support or not support the suspension of 11 political parties with links to Russia?*
(A) Support ban (1)
(B) Neutral (0)
(B) Do not support ban (-1)

Do you support or not support the removal of statues in Ukraine of figures like Pushkin and Catherine the Great*
(A) Support removal (1)
(B) Neutral (0)
(B) Do not support removal (-1)

The variable $UKR\_Ideology_i$ is assigned the average values that corresponds to their response. A value of -3 denotes a very anti-Russia ideological perspective and a value 3 denotes a very Pro-Russia ideological perspective. If a respondent receives an ideological score between -3 and -1 they are considered to have an anti-Russia ideological perspective. If a respondent receives an ideological score between 1 and 3 they are considered to have an pro-Russia ideological perspective.

Proficiency and Usage in Ukrainian relative to Russian:

Language ability can be divided into two related concepts: proficiency and usage (Andreenkova 2019).

In our mainly analysis we proxy for proficiency by categorizing those who choose to take the Study in Ukrainian as $Pref\_Lang\_UKR_i$ and those that choose to take the study in Russian and $Pref\_Lang\_RUS_i$

We also, however, will measure both self assesses usage and proficiency and analyze both of these variables and their relationship to belief in misinformation as well. We will also examine whether the extent of differences in proficiency/usage matter in exploratory analysis. [3]

[3]Since all subjects are unconditionally randomized, we should be balanced on these dimensions as well.

We will create an additive score of the following three questions to measure usage.

What language do you speak in daily life/at home?
(A) Only Russian (1)
(B) Mostly Russian with a few Ukrainian words interspersed (.5)
(C) Equal measure Russian and Ukrainian (0)
(D) Mostly Ukrainian with a few Russian words interspersed (-.5)
(E) Exclusively Ukrainian. (-1)
(F) Mostly some other language (0)

Which language or languages do you usually speak outside home – at work, in the street, and in public places?[4]
(A) Only Russian (1)
(B) Mostly Russian (.5)
(C) Equal measure Russian and Ukrainian (0)
(D) Mostly Ukrainian with a few Russian words interspersed (-.5)
(E) Exclusively Ukrainian. (-1)
(F) Mostly some other language (0)

Which language or languages do you think of as your native? (multiple choice is allowed).[5] We score the question 0 if both Russian and Ukrainian are chosen, -1 if Ukrainian but NOT Russian is chosen, and 1 if Russian but NOT Ukrainian is chosen.
(A) Ukrainian
(B) Russian
(C) Crimean Tatar
(C) Hungarian
(D) Bulgarian
(E) Moldavian
(F) Romanian
(G) Other

We measure proficiency as the normalized sum of differences of the following three questions where we ask respondents about both Ukrainian and Russian.[6]

In your opinion, how well can you SPEAK ... (UKRAINIAN/RUSSIAN)?
(A) I SPEAK fluently (3)
(B) I can SPEAK but with minor difficult (2)
(C) I can SPEAK but with major difficulties (1)
(D) I CANNOT SPEAK his language at all (0)

And how well can you read texts in ... (UKRAINIAN/RUSSIAN)?
(A) I READ fluently (3)
(B) I can READ but with minor difficult (2)
(C) I can READ but with major difficulties (1)
(D) I CANNOT READ this language at all? (0)

[4] If respondents do not work or study, we will omit this question from the calculation.
[5] We take all languages with more that 100,000 estimated speakers.
[6] We adapted these from the "Language Usage Survey in Post-Soviet Republics" (LangUSSR) study cited in Andreenkova 2019 to take into consideration more fine-grained distinctions between Russian and Ukrainian speakers.

And how well can you WRITE texts in ... (UKRAINIAN/RUSSIAN)?
(A) I WRITE fluently (3)
(B) I can WRITE but with minor difficult (2)
(C) I can WRITE but with major difficulties (1)
(D) I CANNOT WRITE this language at all? (0)

We calculate the proficiency of the Ukrainian language ($Prof\_UKR_i$) and Russian language ($Prof\_RUS_i$) by taking the average response to these questions about each language. If only two of these questions are answered we still calculate this ideological score, but normalize it, but if only one or more of these questions are answered we omit this response. To measure the relative proficiency of Ukrainian versus Russian ($Relative\_Proficiency\_UKR_i$) we subtract our measure of their proficiency in the Russian language ($Prof\_RUS_i$) from our measure of their proficiency in the Ukrainian language ($Prof\_UKR_i$). To measure the relative proficiency of their dominant language versus their less proficient language ($Relative\_Proficiency_i$) we subtract our measure of their proficiency in their dominant language from our measure of their proficiency in their non dominant language.

D Recruitment Information

Every week, we will sample 200 respondents evenly from four regions used by Barrington and Herron (2004) (West, Central, South, East), which can be found in the next section (Section J) of the Supplementary Materials (We exclude Russian-controlled Zaporizhzhia Russian-controlled Donetsk, Luhansk, Kherson, Sevastopol, and Crimea). These groups will also be balanced by gender (1/2 who self-identify as male ; 1/2 who self-identify as female), and age (18-24: 12%, 25-34: 27%, 35-44: 33%, 45-55: 28%).

This survey firm recruits individuals through various means. They are paid for their participation in either airline miles or direct transfers of money. Not all respondents are paid the same amount as it is up to both the participant and the vendor (*Infosapiens*) to negotiate terms. The handling of this by *Infosapiens* maintains the anonymity of the respondents in our survey. Survey respondents must answer every question on the survey. This sample size maximizes the power to test hypotheses outlined in this plan but is constrained by funds we have available.

E Regional Divides in Ukraine

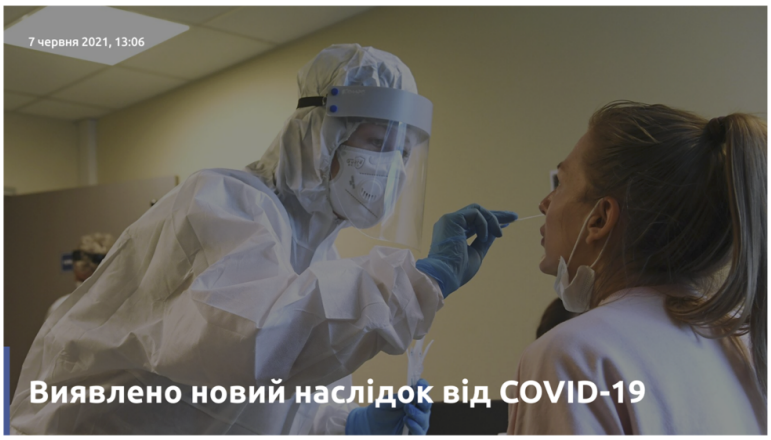
Barrington and Herron (2004) aggregate the following oblasts into four regions:

1. **West:** Volyn, Rivne, Ivano-Frankivsk, Lviv, Ternopil, Chernivetsi, and Zakarpatia.
2. **Center/North:** Chernihiv, Cherkasy, Poltava, Sumy, Kirovohrad, Kyivska Oblast, Kyiv City, Khmelnyskyi, Vinnytsia, and Zhytomyr.
3. **South:** Odesa and Mykolaiv.
4. **East:** Kharkiv (under Ukrainian control), Dnipropetrovsk, Zaporizhzhia (under Ukrainian control), and Donetsk (under Ukrainian control).

Excluded: Krym (Crimea), Sevastopol, Kherson, and Luhansk.

F Article Format (Ukrainian and Russian)

Figure 1: Article from Low-Quality Unclear List (written in Ukrainian)



Дослідники з Бірмінгемського університету стверджують, що захворювання COVID-19 може привести до розвитку синдрому Гієна-Барре.

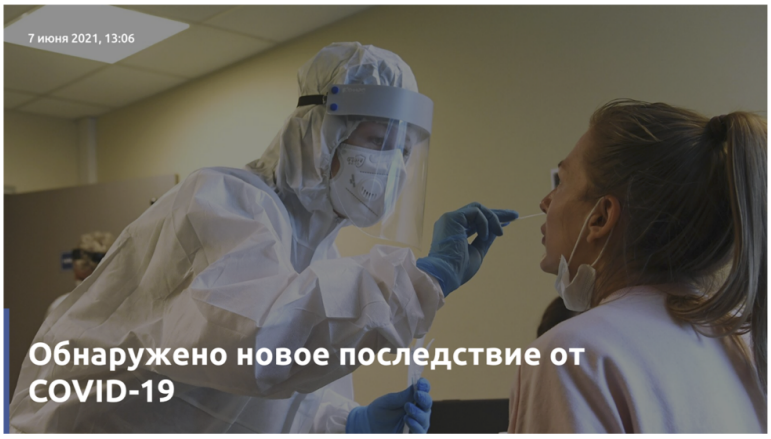
Про це повідомляє кореспондент.

Автор дослідження Алекс Ріхтнер заявив, що антитіла, які організм виробляє після перенесеного коронавируса, схожі з антитілами, що викликають захворювання серця і шкіри.

За результатами дослідження, першими симптомами захворювання Гієна-Барре вважаються оніміння ступень і кистей рук, м'язова слабкість і головний біль, проблеми з координацією. Вони також виявлені в учасників дослідження, які перенесли коронавірус в легкій і важкій формах.

Синдром Гієна-Барре - це аутоімунна запальна полирадикулоневропатия, що протікає в гострій формі і що виявляється млявими парезами, порушеннями чутливості, вегетативними розладами.

Figure 2: Article from Low-Quality Unclear List (written in Russian)



Исследователи из Бирмингемского университета утверждают, что заболевание COVID-19 может привести к развитию синдрома Гийена-Барре.

Об этом сообщает Корреспондент.

Автор исследования Алекс Рихтнер заявил, что антитела, которые организм вырабатывает после перенесенного коронавируса, похожие с антителами, вызывающими заболевание сердца и кожи.

По результатам исследования, первыми симптомами заболевания Гийена-Барре считаются онемение ступней и кистей рук, мышечная слабость и боли, проблемы с координацией. Они также обнаружены у участников исследования, перенесших коронавирус в легкой и тяжелой формах.

Синдром Гийена-Барре — это аутоиммунная воспалительная полирадикулоневропатия, протекающая в острой форме и проявляющаяся вялыми парезами, нарушениями чувствительности, вегетативными расстройствами.

G Balance Table for Pilot Study

Table 5: Balance Table

Group	Observ.	Average Age	Average Income	Gender (Prop. Female)	Average Education	Average Ideology	Prop. Ethnicity (Ukrainian)	Prop. Ethnicity (Russian)
All respondents	422	43.68	3.27	0.50	3.65	-0.02	0.86	0.12
Russian Language Chosen	194	44.49	3.23	0.46	3.61	-0.12	0.73	0.25
Ukrainian Language Chosen	228	42.97	3.31	0.54	3.69	0.05	0.97	0.01
Russian Language Chosen (Control)	106	45.09	3.25	0.46	3.61	-0.17	0.72	0.25
Russian Language Chosen (Treatment)	88	43.77	3.20	0.47	3.60	-0.06	0.75	0.25
Ukrainian Language Chosen (Control)	109	43.29	3.43	0.53	3.80	0.07	0.98	0.01
Ukrainian Language Chosen (Treatment)	119	42.68	3.20	0.55	3.59	0.04	0.97	0.01

H Exploratory Analysis

We discuss the model specifications for all of our models in the main body of the text. We present the results here for those analyses for which our pilot study permits: exploratory analyses 2, 4, 6, and 7. Hence, we omit 1, 3, and 5.

H.1 Exploratory Analysis 2: Effect of Language on Belief in True News Articles

Figure 3: Simulated Treatment Effect of Evaluating *True* News Articles in One's Less Preferred Language Across All Respondents (Pooled) Russian-preferring speakers and Ukrainian-preferring speakers

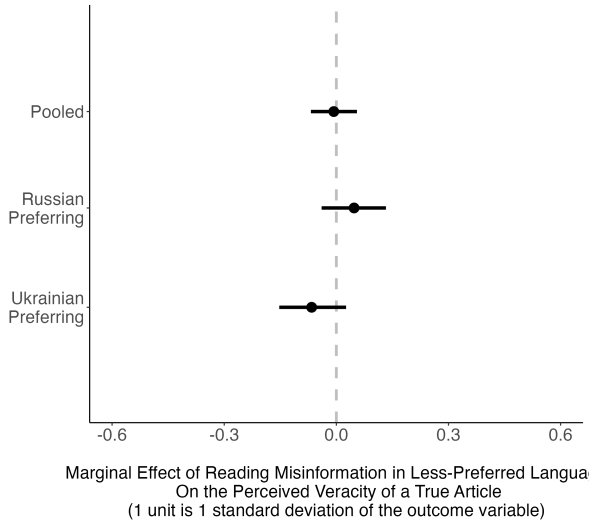
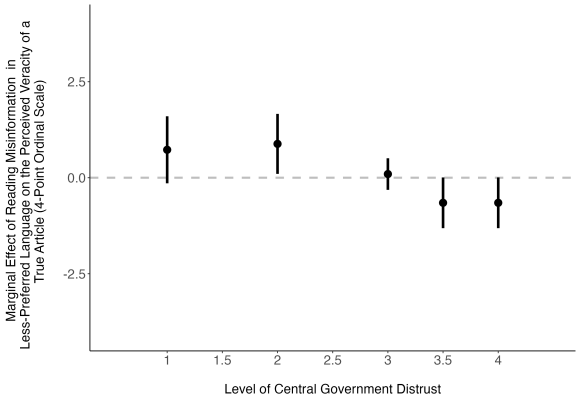
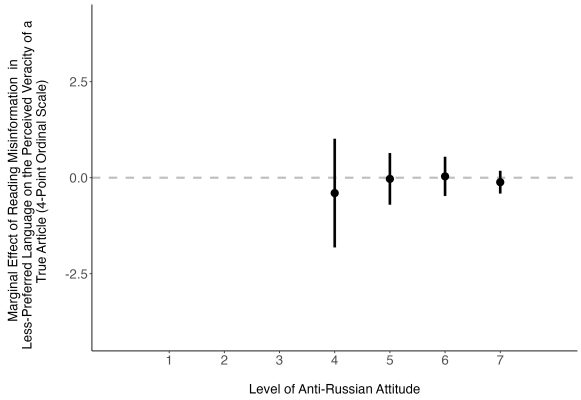


Figure 4: Marginal effects of evaluating *true* news articles in one's less preferred language on Russian-preferred speakers and Ukrainian-preferred speakers across different levels of central government distrust and anti-Russian ideology respectively

(a) Marginal effects of evaluating *true* news articles in one's less preferred language on Russian-preferred speakers across different levels of central government distrust



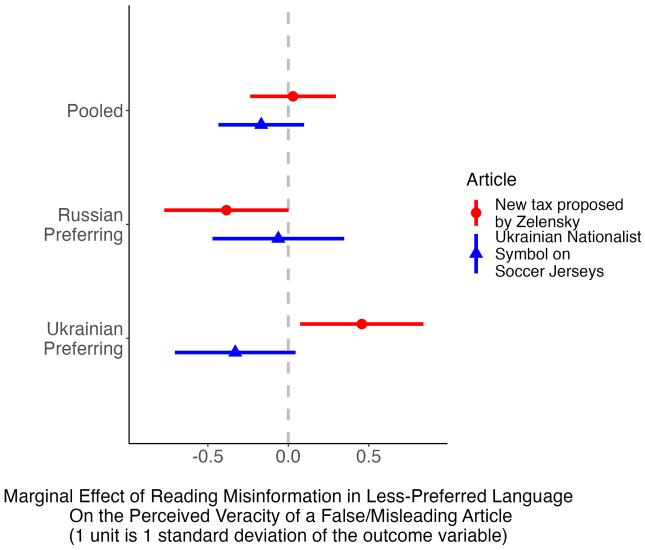
(b) Marginal effects of evaluating *false/misleading* news articles in one's less preferred language on Ukrainian-preferred speakers across different levels of anti-Russian ideology



H.2 Exploratory Analysis 4: Effect Across Articles

Due to a low number of observations in our pilot study we do not run our analysis testing Hypotheses 2a and 2b.

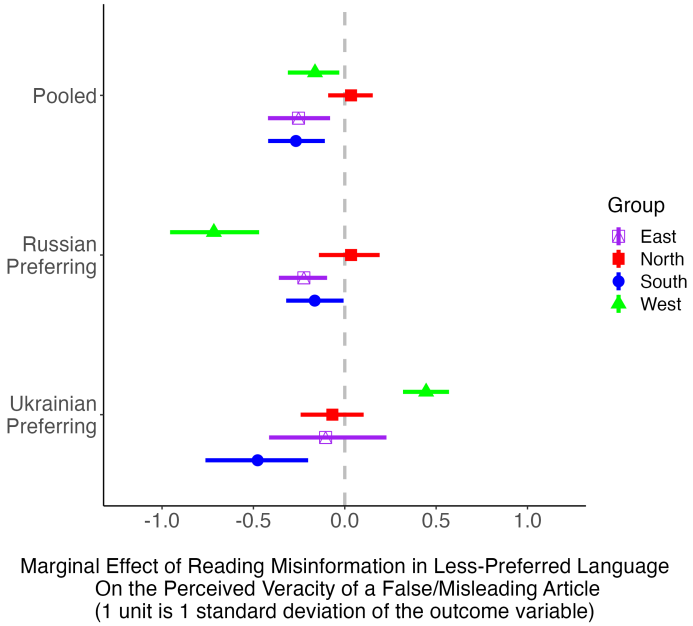
Figure 5: Simulated Treatment Effect of Evaluating *True* News Articles in One's Less Preferred Language Across All Respondents (Pooled) Russian-preferring speakers and Ukrainian-preferring speakers and Articles



H.3 Exploratory Analysis 6: Effect Across Four Regions

Due to a low number of observations in our pilot study we do not run our analysis testing Hypotheses 2a and 2b.

Figure 6: Simulated Treatment Effect of Evaluating *False/Misleading* News Articles in One's Less Preferred Language Across All Respondents (Pooled) Russian-Dominant speakers and Ukrainian-dominant speakers and Regions



H.4 Exploratory Analysis 7: Substituting Discernment Scale for Veracity Measure

Figure 7: Simulated Treatment Effect on Veracity Discernment of Evaluating News Articles in One's Less Preferred Language Across All Respondents (Pooled) Russian-Dominant speakers and Ukrainian-dominant speakers and Regions

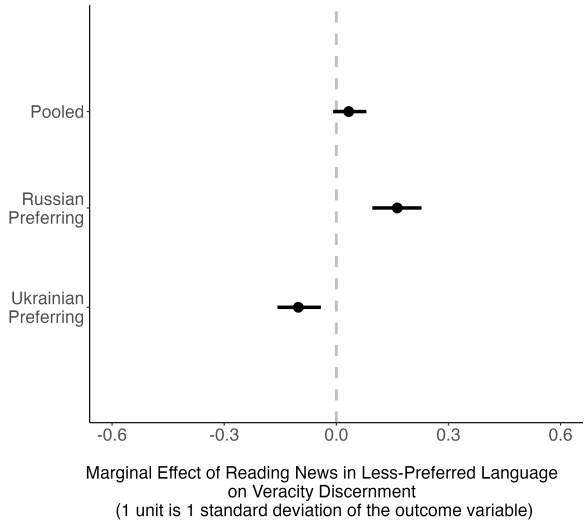
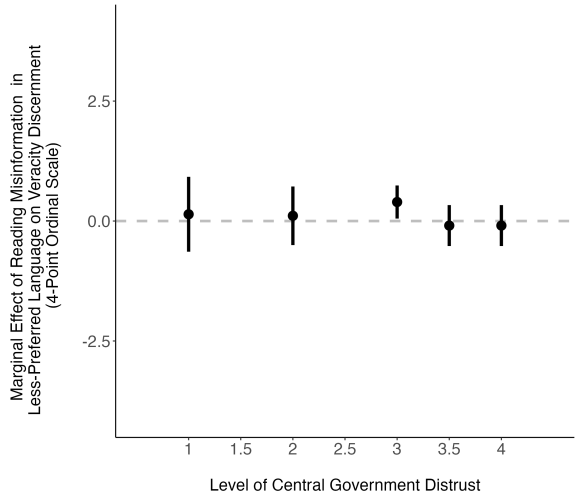
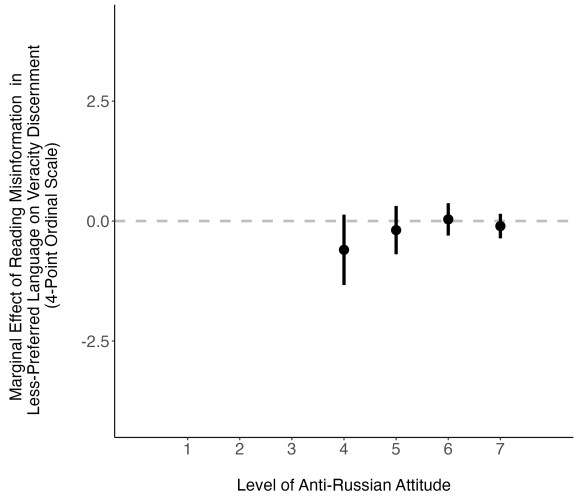


Figure 8: Marginal effects of evaluating *false/misleading* news articles in one's less preferred language on Russian-preferred speakers and Ukrainian-preferred speakers across different levels of central government distrust and anti-Russian ideology respectively

(a) Marginal effects of evaluating news articles on veracity discernment in one's less preferred language on Russian-preferred speakers across different levels of central government distrust



(b) Marginal effects of evaluating news articles on veracity discernment in one's less preferred language on Ukrainian-preferred speakers across different levels of anti-Russian ideology



I Robustness Checks

We discuss the model specifications for all of our models in the main body of the text. We present the results here for those analyses for which our pilot study permits: robustness checks 1, 2, 4, 6, and 7. Hence, we omit 5.

I.1 Robustness Check 1 - Substituting Language Proficiency for Language Preference

Figure 9: Simulated Treatment Effect of Evaluating *False/Misleading* News Articles in One's Less Proficient Language Across All Respondents (Pooled), respondents more proficient in Russian, respondents more proficient in Ukrainian

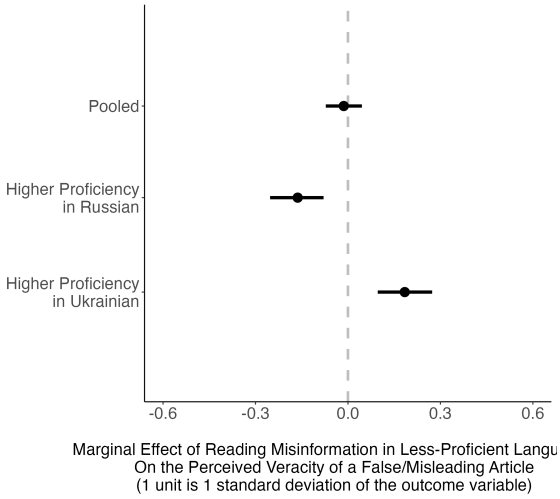
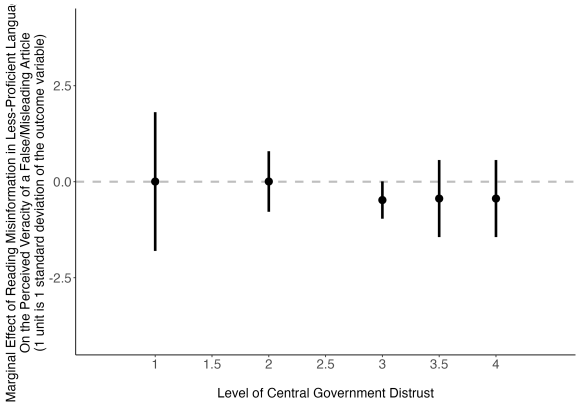
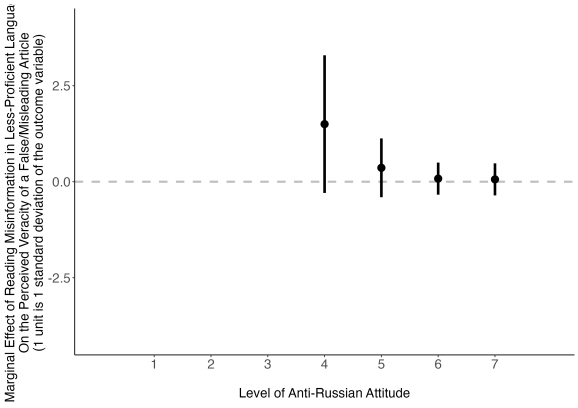


Figure 10: Marginal effects of evaluating *false/misleading* news articles in one's less proficient language on respondents more proficient in Russian and respondents more proficient in Ukrainian across different levels of central government distrust and anti-Russian ideology respectively

(a) Marginal effects of evaluating *false/misleading* news articles in one's less proficient language on respondents more proficient in Russian across different levels of central government distrust



(b) Marginal effects of evaluating *false/misleading* news articles in one's less preferred language on respondents more proficient in Ukrainian across different levels of anti-Russian ideology



I.2 Robustness Check 3 - Replace Dependent Variable with Categorical Measure

Figure 11: Simulated Treatment Effect of Evaluating *False/Misleading* News Articles in One's Less Preferred Language Across All Respondents (Pooled) Russian-preferring speakers and Ukrainian-preferring speakers

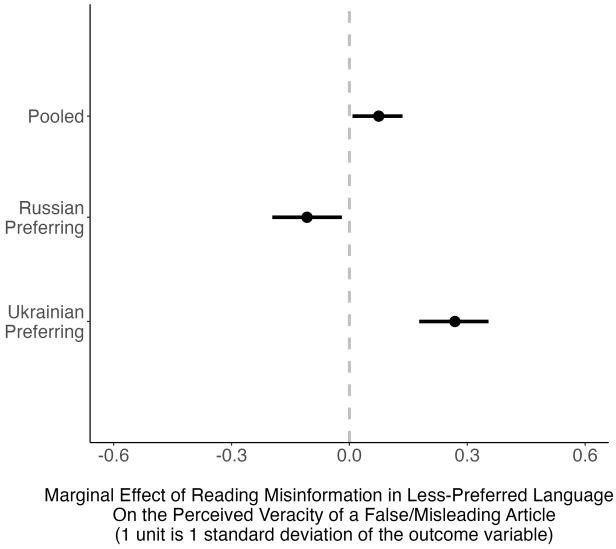
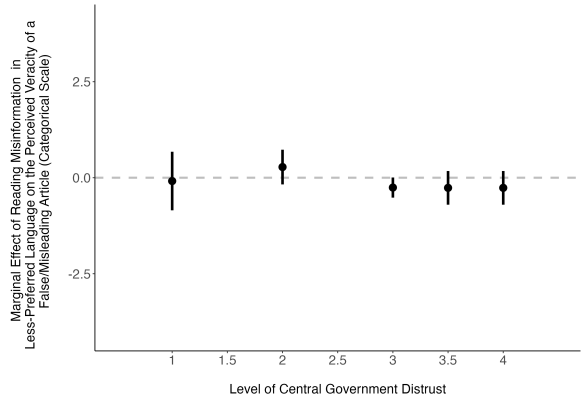
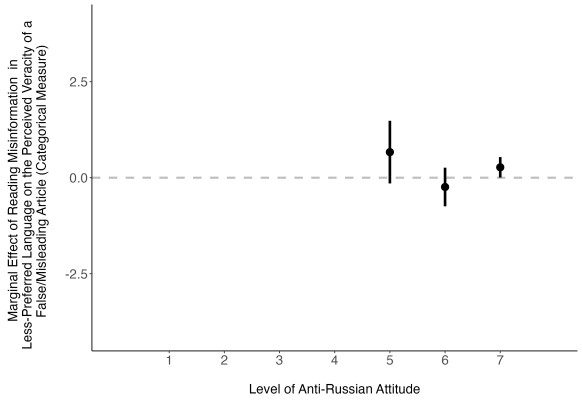


Figure 12: Marginal effects of evaluating *false/misleading* news articles in one's less preferred language on Russian-preferred speakers and Ukrainian-preferred speakers across different levels of central government distrust and anti-Russian ideology respectively

(a) Marginal effects of evaluating *false/misleading* news articles in one's less preferred language on Russian-preferred speakers across different levels of central government distrust



(b) Marginal effects of evaluating *false/misleading* news articles in one's less preferred language on Ukrainian-preferred speakers across different levels of anti-Russian ideology



I.3 Robustness Check 4 - Only Those Who Were Not Exposed To The Article Previously

Figure 13: Simulated Treatment Effect of Evaluating *False/Misleading* News Articles in One's Less Preferred Language Across All Respondents (Pooled) Russian-preferring speakers and Ukrainian-preferring speakers

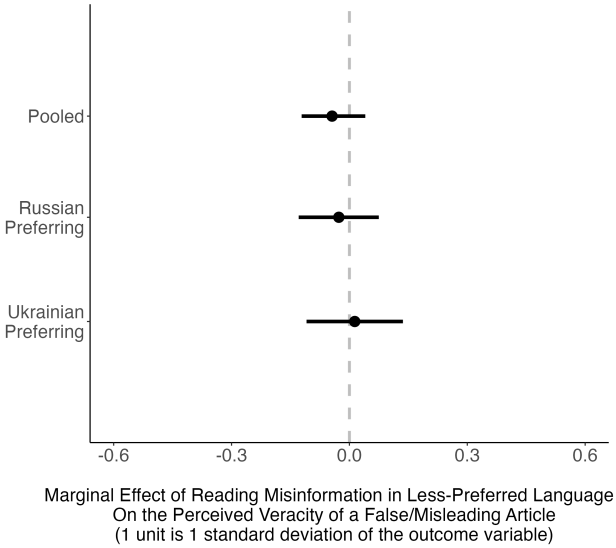
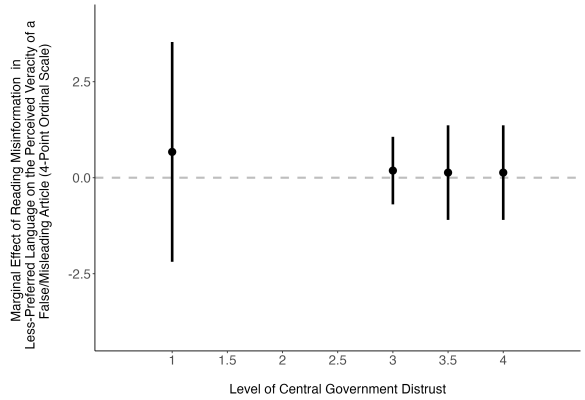
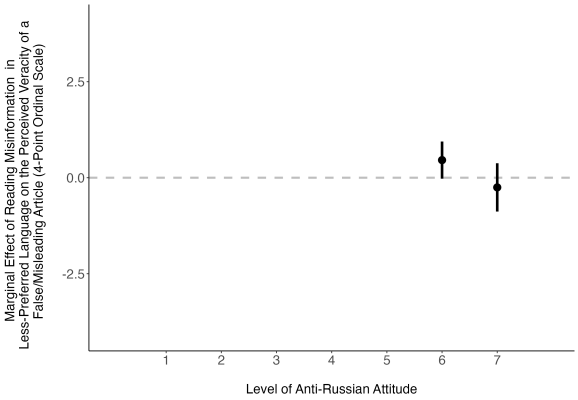


Figure 14: Marginal effects of evaluating *false/misleading* news articles in one's less preferred language on Russian-preferred speakers and Ukrainian-preferred speakers across different levels of central government distrust and anti-Russian ideology respectively

(a) Marginal effects of evaluating *false/misleading* news articles in one's less preferred language on Russian-preferred speakers across different levels of central government distrust



(b) Marginal effects of evaluating *false/misleading* news articles in one's less preferred language on Ukrainian-preferred speakers across different levels of anti-Russian ideology



I.4 Robustness Check 6 - Adding the Order Variable to the Model

Figure 15: Simulated Treatment Effect of Evaluating *False/Misleading* News Articles in One's Less Preferred Language Across All Respondents (Pooled) Russian-preferring speakers and Ukrainian-preferring speakers

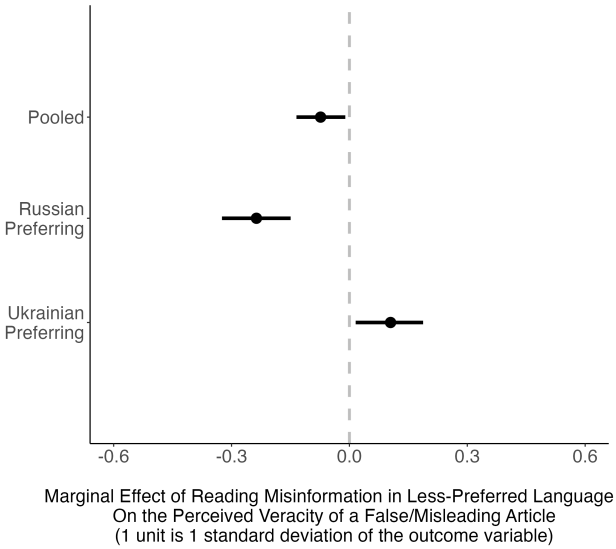
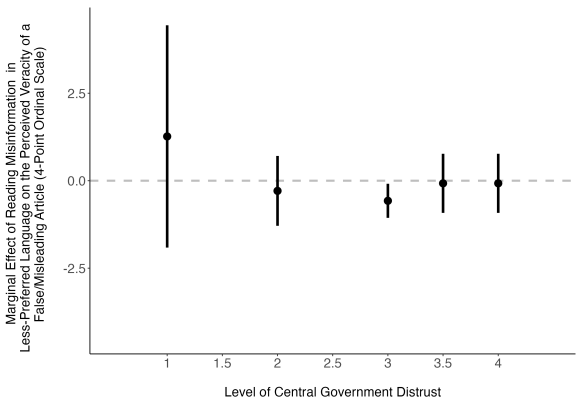
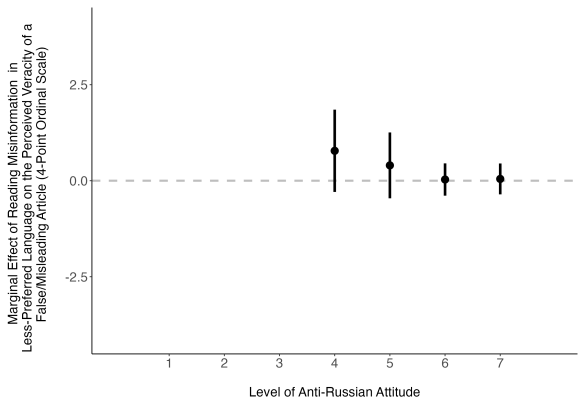


Figure 16: Marginal effects of evaluating *false/misleading* news articles in one's less preferred language on Russian-preferred speakers and Ukrainian-preferred speakers across different levels of central government distrust and anti-Russian ideology respectively

(a) Marginal effects of evaluating *false/misleading* news articles in one's less preferred language on Russian-preferred speakers across different levels of central government distrust



(b) Marginal effects of evaluating *false/misleading* news articles in one's less preferred language on Ukrainian-preferred speakers across different levels of anti-Russian ideology



J Articles Evaluated during Pilot Study:

Table 6: Headlines for Articles Chosen from the Low Quality Liberal News Stream in Study I

News Stream	Date	Headline	Modal Fact-Checker Rating	Topic	Lean of Article
Not Anti-Russia Low-Quality	06/07/21	Zelensky will impose a tax on workers - Kiev experts	False/Misleading	Economics	Neutral
Anti-Russia Low-Quality	06/07/21	The Russian Federation is “unaware” of the shape of Ukraine – Says the Ukrainian Defense Ministry	True	Politics	Anti-Russia
Unclear Low-Quality	06/07/21	A new side-effect of COVID-19 has been discovered	Could Not Determine	Science	Neutral
Russian Mainstream	06/07/21	The Ukrainian national football team will participate at the European Championship with the symbol of a nationalist slogan	False/Misleading	Politics	Not Anti-Russia
Ukrainian Mainstream	06/07/21	He is not worthy! KNU has deprived Lukashenko of the title of honorary doctor	True	Politics	Anti-Russia

K Article Selection Process

To select popular true and false/misleading articles from a diverse set of ideological perspectives we send out the most popular article from five different lists of news sources: (1) Not Anti-Russia low credibility; (2) Anti-Russia low-credibility news sources; (3) General low credibility news sources with an ambiguous or unclear ideological affiliation; (4) Ukraine-based mainstream sources; and (5) Russia-based mainstream sources.

All low-credibility news sources are selected from Texty, but we additionally use trained coders to classify each web-based news source into one of three ideological perspective categories ("anti-Russia", "not anti-Russia", and "unclear") that are based on a longstanding distinction in Ukrainian politics (Erlich and Garner 2021). Although we do not know where all of these news sources are based, we do know that the majority of "anti-Russia" low-credibility news sources are based in Ukraine and the majority of the "not anti-Russia" low-credibility news sources are based in Russia. Coders were asked to use the headlines, the content of its articles, as well as the websites domain and 'about' page to classify websites. [7] Websites were only classified as "not anti-Russia" or "anti-Russia" when at least two of the three coders agreed. If a website did not receive at least two "not anti-Russia' ratings or two "anti-Russia" ratings it was deemed to have an "unclear" ideological lean.[8] In total, during the pre-test, we categorize eight domains as "anti-Russia–low-credibility" news streams, thirty-eight domains as "not anti-Russia– low-credibility" news streams, and nine domains as "unclear–low-credibility" news streams."[9] The prevalence of "not anti-Russia low-credibility" news streams comports with previous research that reports high levels of production of pro-Kremlin misinformation in Ukraine (Snegovaya 2015; StratCom 2015).

Our two mainstream news articles are sourced from a list of popular Ukraine-based mainstream news sites and a list of Russia-based mainstream news sites consumed in Ukraine, both derived from an independent data journalism platform- Texty.[10]

Using these five lists of online news websites we place each list into its own distinct RSS feed (Feedly) and source the most popular articles from each list every week our survey runs. Every week, on Tuesday morning we will source the most popular articles published on Monday from each of the five lists of sources: (1) Not Anti-Russia low credibility news sources (that may publish misinformation) ; (2) anti-Russia low-credibility news sources (that may publish misinformation) ; (3) General low credibility news sources (that may publish misinformation) with an ambiguous or unclear ideological affiliation ; (4) Ukraine-based mainstream sources (that likely publishes true information) ; and (5) Russia-based mainstream sources (that likely publishes true information) consumed in Ukraine. Each article selected will be excluded if it fits our article exclusionary protocol, which can be found in Section A. If it is excluded, we will move on to the next most popular article in the RSS feed.

[7] The complete coding instructions given to these coders are available on the next page.
[8] The agreement scores between coders are available in this section.
[9] This section lists each domain and their ideological perspective.
[10] `https://topic-radar.texty.org/#/`

Coding Instructions for False News Sources

We are asking you to classify news websites by their political and/or partisan orientation. In particular, we are asking you to go through a list of known news websites and classify each one by its political ideology. We expect this to take about five minutes per website.

For each website, please do the following:

- Visit each website.
- Record if the website still appears to be an active news website by confirming that there is at least one new post in the past 30 days .
- Read the headlines, articles of interest and the about page of the website.
- **Classify the website in terms of their ideological leaning:**
  - Anti-Russia
  - Not Anti-Russia
  - Unclear

You can use the following to make a determination about the website's ideological lean:

- Name of the website
- About page
- Slogan of the website
- Article headlines
- Articles and their content

Inter-Rater Reliability Statistics For Coding the Ideological Lean of News Sources

Table 7: Inter-Rater Reliability Statistics For Coding the Ideological Lean of News Sources

Coders	Agreement	Cohen Kappa Score	Categories
01 and 02	0.809	0.730	3
02 and 03	0.865	0.810	3
01 and 03	0.775	0.695	3

Five News Streams for Pilot Study

Table 8: Ukrainian-High Credibility News Domains

Domain
24tv.ua
bykvu.com
censor.net.ua
fakty.com.ua (No RSS Feed)
fakty.ua
gordonua.com
interfax.com.ua
lb.ua
liga.net
nv.ua
pravda.com.ua
rbc.ua
segodnya.ua (No longer active)
tsn.ua
ukrinform.ru (No RSS Feed and no longer active)
unian.net
zn.ua

Table 9: Russian-High Credibility News Domains

Domain
aif.ru
dni.ru
kommersant.ru
kp.ru
lenta.ru
lentainform.com (No RSS Feed and no longer active)
life.ru
newsru.com
pravda.ru
regnum.ru
riafan.ru
russian.rt.com
slovodel.com
svpressa.ru (No RSS Feed)
tass.ru
vz.ru

Table 10: Not Anti-Russia-Low Credibility News Domains:

Domain
fraza.ua (no longer active)
mignews.com.ua (no longer active)
sharij.net
vesti.ua
vz.ua
antimaydan.info (No RSS Feed)
c-inform.info
dnr-pravda.ru
dnr24.com
dnr24.su (no longer active)
donbasstoday.ru
doneck-news.com
dontimes.ru
e-gorlovka.com.ua (No RSS Feed and no longer active)
evening-crimea.com (no longer active)
free-news.su (No RSS Feed and no longer active)
fromdonetsk.net
luga1news.ru (no longer active)
lugansk1.info (no longer active)
miaistok.su (no longer active)
newc.info (no longer active)
news-front.info
novorosinform.org
novorossiy.info (no longer active)
novosti.icu
patriot-donetsk.ru (no longer active)
politnavigator.net
pravdanews.info (no longer active)
ruinformer.com
rusdnepr.ru
rusnext.ru
sevastopol.su
sevnews.info (no longer active)
time-news.net
voenkor.info (No RSS Feed) (no longer active)
voskhodinfo.su
vsednr.ru
xvesti.ru

Table 11: Anti-Russia-Low Credibility News Domains

Domain
agrimpasa.com
antikor.com.ua
finoboz.net
inforesist.org
informator.news (No RSS Feed)
kyiv.press
spektrnews.in.ua
ukranews.com

Table 12: Unclear Low Credibility News Domains

Domain
glavred.info
jizn.info (no longer active)
kompromat1.news
korr.com.ua (no longer active)
lifedon.com.ua (no longer active)
news247.com.ua (No RSS Feed)
podrobnosti.ua (No RSS Feed)
kafanews.com
meridian.in.ua

Five News Streams for Future Study (News sources that are no longer available have been removed)

Table 13: Ukrainian-High Credibility News Domains

Domain
24tv.ua
bykvu.com
censor.net.ua
fakty.com.ua (No RSS Feed)
fakty.ua
gordonua.com
interfax.com.ua
lb.ua
liga.net
nv.ua
pravda.com.ua
rbc.ua
tsn.ua
unian.net
zn.ua

Table 14: Russian-High Credibility News Domains

Domain
aif.ru
dni.ru
kommersant.ru
kp.ru
lenta.ru
life.ru
newsru.com
pravda.ru
regnum.ru
riafan.ru
russian.rt.com
slovodel.com
svpressa.ru (No RSS Feed)
tass.ru
vz.ru

Table 15: Not Anti-Russia-Low Credibility News Domains:

Domain
sharij.net
vesti.ua
vz.ua
antimaydan.info (No RSS Feed)
c-inform.info
dnr-pravda.ru
dnr24.com
donbasstoday.ru
doneck-news.com
dontimes.ru
fromdonetsk.net
news-front.info
novorosinform.org
novosti.icu
politnavigator.net
ruinformer.com
rusdnepr.ru
rusnext.ru
sevastopol.su
time-news.net
voskhodinfo.su
vsednr.ru
xvesti.ru

Table 16: Anti-Russia-Low Credibility News Domains

Domain
agrimpasa.com
antikor.com.ua
finoboz.net
inforesist.org
informator.news (No RSS Feed)
kyiv.press
spektrnews.in.ua
ukranews.com

Table 17: Unclear Low Credibility News Domains

Domain
glavred.info
kompromat1.news
kafanews.com
meridian.in.ua

L Code to Run Models Testing Hypotheses 1–2 (R Statistical Program)

```
#Load libraries:
library(estimatr) library(interflex)


#Hypothesis 1:

#Fit model adjusted for covariates
lm_fit = lm_robust(Y_Outcome ~ Lang_Treatment + RUS_Lang Age + Gender + Income + Education +
Digital_Literacy_Treatment + Region + article_batch, data = data_frame)

#Fit model unadjusted for covariates
lm_fit = lm_robust(Y_Outcome ~ Lang_Treatment + article_batch, data = data_frame)


#Hypothesis 2a:

#Fit model adjusted for covariates
lm_fit = lm_robust(Y_Outcome ~ Lang_Treatment*Central_Government_Distrust + Age + Gender + Income
+ Education + Digital_Literacy_Treatment + Region + article_batch, data = Russian_data_frame)

#Fit model unadjusted for covariates
lm_fit = lm_robust(Y_Outcome ~ Lang_Treatment*Central_Government_Distrust, data = Russian_data_frame)


#Hypothesis 2b:

#Fit model adjusted for covariates
lm_fit = lm_robust(Y_Outcome ~ Lang_Treatment*Anti_Russia_Ideology + Age + Gender + Income +
Education + Digital_Literacy_Treatment + Region + article_batch, data = Ukrainian_data_frame)

#Fit model unadjusted for covariates
lm_fit = lm_robust(Y_Outcome ~ Lang_Treatment*Anti_Russia_Ideology, data = Ukrainian_data_frame)


#Multiplicative Interaction Model Diagnostics (Hypothesis 2a):

out <- interflex(Y = "Eval_Ordinal_Four", D = "Read_Opposite_Lang", X = "Central_Government_Distrust",
Z = c("Age", "Gender", "Income","Education"), FE = "Region", data = Russian_data_frame, estimator
= "binning", main = "Marginal Effects", ylim = c(1, 4), nbins = 6, na.rm=T)

plot(out, xlab = "Moderator:  Central Government Distrust", ylab = "Marginal Effect of Opposite
Languageon Belief in Misinformation (4-Point ordinal scale)", bin.labs = FALSE)

#Multiplicative Interaction Model Diagnostics (Hypothesis 2b):

out <- interflex(Y = "Eval_Ordinal_Four", D = "Read_Opposite_Lang", X = "Anti_Russian_Ideology",
Z = c("Age", "Gender", "Income","Education"), FE = "Region", data = Ukrainian_data_frame, estimator
= "binning", main = "Marginal Effects", ylim = c(1, 4), nbins = 6, na.rm=T)

plot(out, xlab = "Moderator:  Anti-Russian Political Ideology", ylab = "Marginal Effect of
Opposite Language on Belief in Misinformation (4-Point ordinal scale)", bin.labs = FALSE)
```

M Attention Check Questions

1. Most modern theories of decision making recognizes that decisions do not take place ina vacuum. Individual's preferences and knowledge, along with situational variables can greatly impact the decision process. To demonstrate that you've read this much, just go ahead and select both red and green among the alternatives below, no matter what your favorite color is. Yes, ignore the question below and select both of those options. What is your favorite color?

- White
- Black
- Red
- Pink
- Green
- Blue

2. We are also interested in what sections people like to read in the newspaper. This might affect what they learn from articles and how they feel about the issues discussed in them. We also want to see if people are reading the questions carefully. To show that you've read this much, please mark both the Classified and none of the above boxes below. That's right just select these two options only. Regardless of how frequently you read the newspaper, what would you say are your favorite sections to read? (please check all that apply)

- National
- Local
- Real Estate
- Comics
- Classified
- Style
- Sports
- Business
- Science and Technology
- Opinion
- None of the above
- All of the above

N Power Analysis

Using the `DeclareDesign` package (Blair et al. 2019) in R, we estimated the power for different study sizes assuming an average treatment effect (ATE) of different Cohen's D (0.1, 0.15, 0.2) using our pre-registered model. The figures below show that for an ATE of a Cohen's D of 0.2 and 800 observations this model will present a statistically significant effect size over 80% of the time. We plan to collect 2,000 observations in our study, but it is possible we sample less depending on how the articles we sample are evaluated by the professional fact-checkers.

Given that the majority of studies measuring effects of interventions in belief in misinformation report effect sizes above a Cohen's D of 0.2, we feel that this is a proper conservative measure to focus on. For example, many similar recent studies measuring the average treatment effect (ATE) of interventions on belief in misinformation report effect sizes that have Cohen's D larger than 0.3 (Roozenbeek and Van der Linden 2019; Pennycook et al. 2020; Guess et al. 2020).

Figure 17: Statistical power for different combinations of observations and minimum detectable effects for respondent that prefer Ukrainian

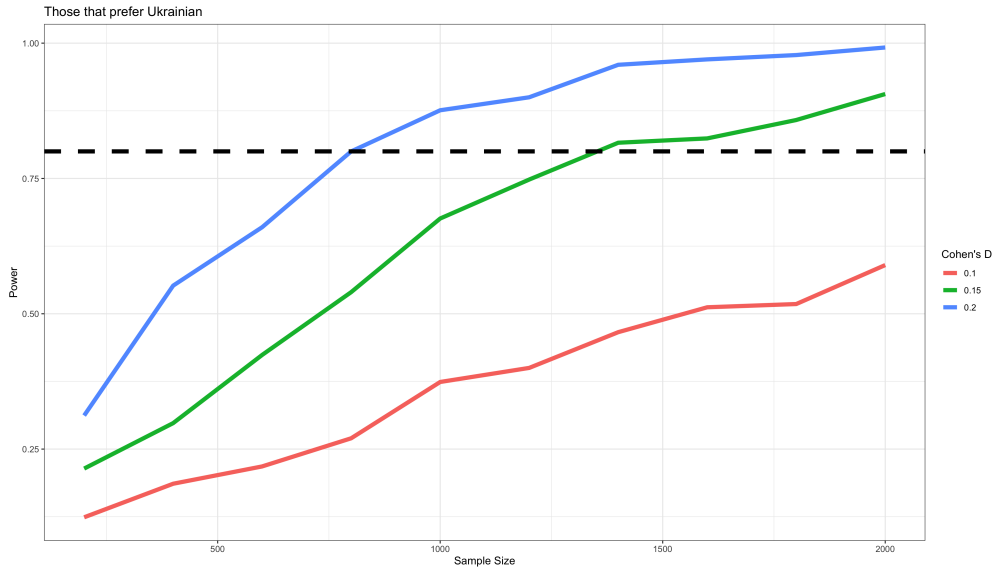
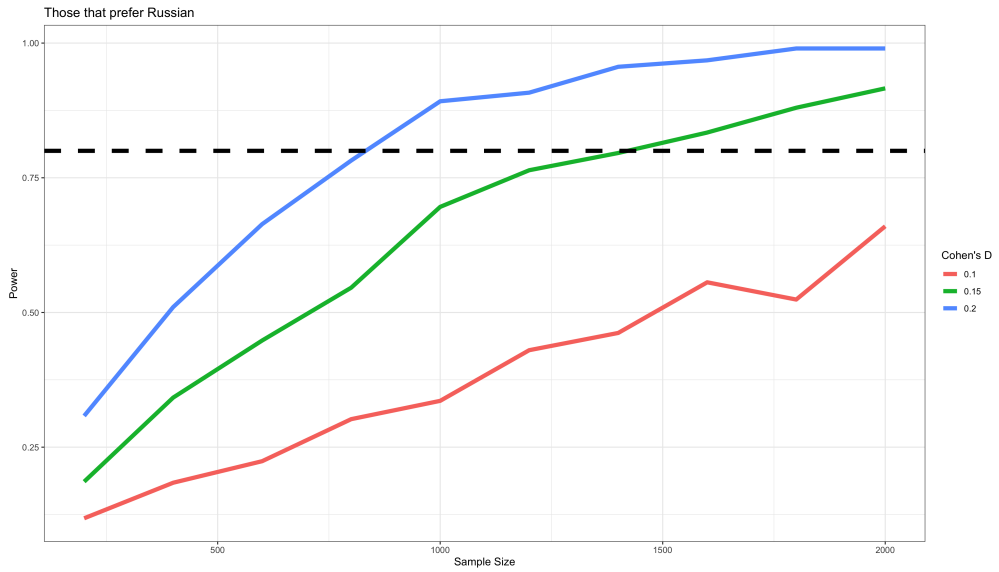


Figure 18: Statistical power for different combinations of observations and minimum detectable effects for respondent that prefer Russian



In Figures 19 and 20 we also estimated the minimum detectable effect for different study sizes assuming a statistical power of 80% using our pre-registered model. These figures show similar results. We are only able to identify an ATE of a Cohen's D of 0.2 with statistical power of 80% at 800 observations. When we have 2000 observations (our predicated number) we can identify an ATE of a Cohen's D of 0.14 with statistical power of 80%.

Figure 19: Minimum Detectable Effects for different combinations of observations and statistical power of 80% for respondents that prefer Ukrainian

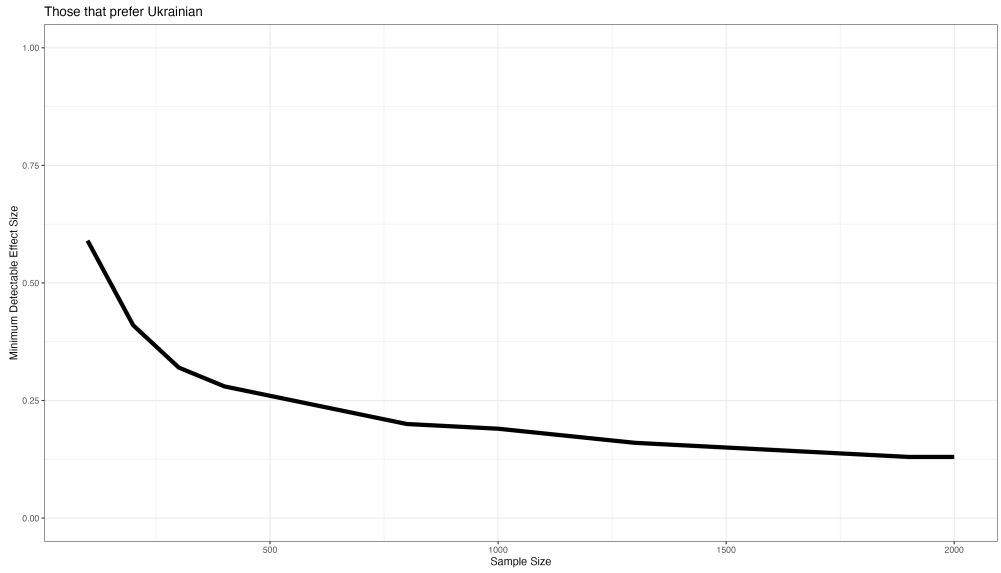
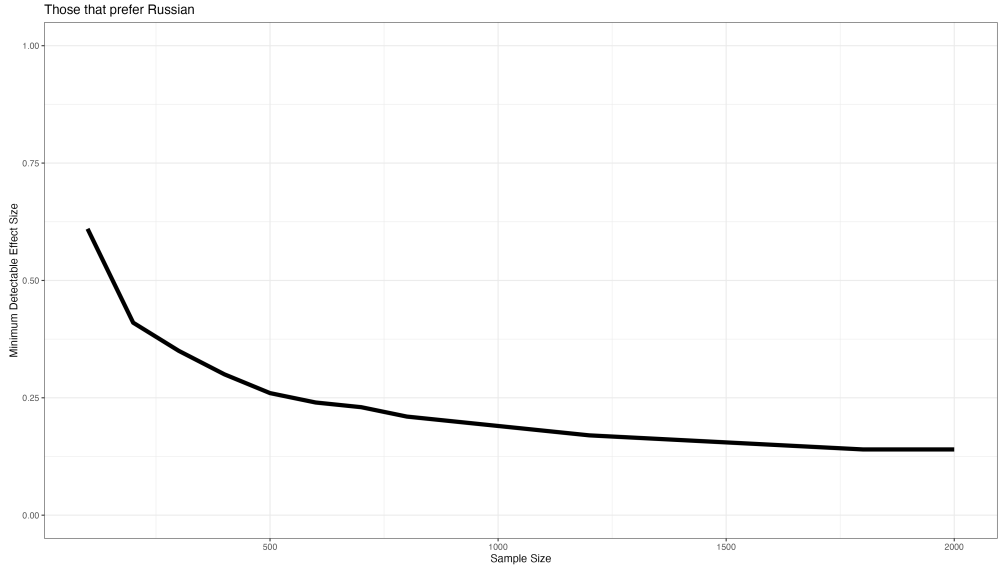


Figure 20: Minimum Detectable Effects for different combinations of observations and statistical power of 80% for respondents that prefer Russian



O Treatment: Tips To Spot False News

Tips to spot fake news some respondents are shown to help them identify misinformation.

Be skeptical of headlines. False news stories often have catchy headlines in all caps with exclamation points. If shocking claims in the headline sound unbelievable, they probably are.

Look closely at the URL. A phony or look-alike URL may be a warning sign of false news. Many false news sites mimic authentic news sources by making small changes to the URL. You can go to the site to compare the URL to established sources.

Investigate the source. Ensure that the story is written by a source that you trust with a reputation for accuracy. If the story comes from an unfamiliar organization, check their "About" section to learn more.

Watch for unusual formatting. Many false news sites have misspellings or awkward layouts. Read carefully if you see these signs.

Consider the photos. False news stories often contain manipulated images or videos. Sometimes the photo may be authentic, but taken out of context. You can search for the photo or image to verify where it came from.

Inspect the dates. False news stories may contain timelines that make no sense, or event dates that have been altered.

Check the evidence. Check the author's sources to confirm that they are accurate. Lack of evidence or reliance on unnamed experts may indicate a false news story.

Look at other reports. If no other news source is reporting the same story, it may indicate that the story is false. If the story is reported by multiple sources you trust, it's more likely to be true.

Is the story a joke? Sometimes false news stories can be hard to distinguish from humor or satire. Check whether the source is known for parody, and whether the story's details and tone suggest it may be just for fun.

Some stories are intentionally false. Think critically about the stories you read, and only share news that you know to be credible.

References

Andreenkova, Anna V. (2019). "How to Choose Interview Language in Different Countries". In: *Advances in Comparative Survey Methods*. Ed. by Timothy P. Johnson et al. 1st ed. John Wiley & Sons, Ltd, pp. 314–324.

Barrington, Lowell W and Erik S Herron (2004). "One Ukraine or many? Regionalism in Ukraine and its political consequences". In: *Nationalities Papers* 32.1, pp. 53–86.

Blair, Graeme et al. (2019). "Declaring and Diagnosing Research Designs". In: *American Political Science Review* 113 (3), pp. 838–859. URL: `https://declaredesign.org/paper.pdf`.

Erlich, Aaron and Calvin Garner (2021). "Subgroup Differences in Implicit Associations and Explicit Attitudes during Wartime". In: *International Studies Quarterly*.

Fujii, Lee Ann (2012). "Research ethics 101: Dilemmas and responsibilities". In: *PS: Political Science & Politics* 45.4, pp. 717–723.

Guess, Andrew M et al. (2020). "A digital media literacy intervention increases discernment between mainstream and false news in the United States and India". In: *Proceedings of the National Academy of Sciences* 117.27, pp. 15536–15545.

Pennycook, Gordon et al. (2020). "Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention". In: *Psychological science* 31.7, pp. 770–780.

Roozenbeek, Jon and Sander Van der Linden (2019). "Fake news game confers psychological resistance against online misinformation". In: *Palgrave Communications* 5.1, pp. 1–10.

Shanks, Kelsey and Julia Paulson (2022). "Ethical research landscapes in fragile and conflict-affected contexts: understanding the challenges". In: *Research Ethics*, p. 17470161221094134.

Snegovaya, Maria (2015). "Putin's information warfare in Ukraine". In: *Soviet Origins of Russia's Hybrid Warfare', Russia Report* 1, pp. 133–135.

StratCom (2015). *Analysis of Russias Information Campaign against Ukraine.* Tech. rep. NATO StratCom Centre of Excellence.