

Trust in Survey Results. A Cross-country Replication Experiment

Ádám Stefkovics^{1,2} · Zoltán Kmetty^{1,3}

¹HUN-REN Centre for Social Sciences

²Századvég Foundation

³ELTE Faculty of Social Sciences

Amidst a flood of new survey results, people selectively determine which results to trust. These decisions may be shaped by perceived methodological quality, the survey's sponsor or platform, and various individual political and non-political characteristics. This study examined the factors that drive trustworthiness assessments of survey results in the two countries. Study 1 replicated the study of Stadtmüller et al. (2022), Study 2 extended the original study by applying the same experimental design to a more contested and politicized issue (immigration). In online panel surveys, respondents were asked to rate hypothetical survey results. Survey quality information, sponsors, and survey results varied in the vignettes. The results show that most of the findings of Stadtmüller et al. (2022) travel well across survey topics and countries. Individual characteristics explain more than survey quality information, larger sample sizes enhance trust, and individuals with higher levels of education attribute more importance to methodological information, regardless of the topic, in both countries. Our study brought nuance to the original findings by showing that the extent to which individuals incorporate survey quality information depends on the degree of controversy and politicization associated with the survey topic.

Keywords: survey; trust; data quality; replication

1 Introduction

Surveys play a pivotal role in democratic processes by providing a means to measure public opinion, allowing policymakers and decision-makers to make informed choices that align with the needs and preferences of the people, thus fostering a more inclusive and participatory democracy (Brodie et al. 2021; Gallup and Rae 1940; Newport et al. 2013; Shapiro 2011; Verba 1996). With the spread of online polling methods (Couper 2013), there may be more survey results in news headlines than ever before (Madson

and Hillygus 2020). Consequently, people are constantly exposed to a plethora of survey findings that may shape their everyday decisions and behaviors. Findings about the public's position on certain issues (e.g., candidate chances) aid individuals in reevaluating their existing beliefs (Boudreau and McCubbins 2010; Chia and Chang 2017) and adjusting their actions accordingly (Alabrese 2022; Ansolabehere and Iyengar 1994; Blais et al. 2006; Dawson 2022; Morwitz and Pluzinski 1996).

Nevertheless, for survey results to effectively foster democratic processes and serve as reliable reference points, they are contingent on the trust placed in them by the public. There are several reasons why people would choose not to trust survey results.

For example, people may be concerned about the methodology of the survey and the quality of the data (Salwen 1987). The rapidly growing use of non-probability-based online survey methods that do not necessarily enable accurate generalizations on broader populations, has contributed to a perceived decline in the overall quality of survey research (Baker et al. 2016; Stadtmüller et al. 2022).

Supplementary Information The online version of this article (<https://doi.org/10.18148/srm/2025.v19i2.8267>) contains supplementary material.

Corresponding author: Ádám Stefkovics, HUN-REN Centre for Social Sciences, Budapest, Hungary (Email: stefkovics.adam@tk.hu)

However, because of the decreasing response rates worldwide, even probability-based surveys encounter challenges in maintaining quality standards.

Another reason for the lack of trust may be the failure to provide transparent information regarding the survey. Despite the importance of transparency and available standards for disclosure of survey methodology (e.g., AAPOR 2021; Seidenberg et al. 2023), both media (Bhatti and Pedersen 2016; Meyer 1990; Miller and Hurd 1982; Portilla et al. 2016) and academia (Stefkovics et al. 2024; von Hermann and Lemcke 2017) tend not to comply with these standards.

A related potential source of distrust can be the biases journalists introduce in the interpretation of survey results. Examples of this include unsubstantiated reports of change due to the demand for horse race coverage (Bhatti and Pedersen 2016; Larsen 2021).

Finally, people may distrust survey results regardless of their methodological background due to motivated reasoning or political motivation. Research shows that people tend to reject information that contradicts their current beliefs and accept those that align with their viewpoints (Bolsen et al. 2014; Kunda 1990; Kuru et al. 2017, 2020; Redlawsk 2002; Tsifti 2001). Irrespective of the results, individuals may perceive poll results as unreliable because of a lack of trust in the credibility of their source or because of perceived partisan leaning of their source (Chia and Chang 2017; Searles et al. 2018). Moreover, politicization of the survey research field can create a perception among many individuals that the survey takers or the survey results are subject to political manipulation, leading to biased outcomes (Lelkes et al. 2017; Madson and Hillygus 2020). Amidst declining trust in science and the growing political polarization of the media landscape in Western democracies (Huber et al. 2019; Krause et al. 2019; Lorenz-Spreen et al. 2023), it is plausible that such perceptions emerge.

A handful of studies addressed the formulation of trust in surveys (Kuru et al. 2017, 2020; Madson and Hillygus 2020; Stadtmüller et al. 2022). Kuru et al. (2017) experimentally manipulated survey reports about gun control and abortion and varied the poll outcome (support/oppose), the media outlet disseminating the survey (FoxNews/MSNBC), and the extent of methodological information provided (none/extensive). They found evidence for motivated reasoning, as the preexisting attitudes of respondents were the strongest factor in determining the perceived credibility of poll results, while they reported non-significant findings about the role of the news source and mixed findings about the role of the methodological details. The findings of Madson and Hillygus (2020) corroborated these results. In two studies, they found that respondents evaluated polls on both candidates' chances and policy issues as more credible when the majority opinion in the survey matched their views.

In another study, Kuru et al. (2020) presented the results of two election polls about candidate preferences in the 2016 US presidential election and varied whether the polls exhibited consistent or divergent results regarding the leading candidates and whether the polls consistently demonstrated high or low methodological quality. Respondents tended to perceive polls as more credible when they indicated that their preferred candidate was in the lead. They also found a moderating effect of education: highly educated respondents were better at accurately identifying high-quality polls, whereas less educated respondents' bias decreased when presented with polls of varying methodological quality.

In contrast to Kuru et al. (2017, 2020), Stadtmüller et al. (2022) examined trust in surveys in the case of a non-salient, less contested issue. They varied the results of a survey on tax allowances for commuters due to rising gasoline prices, the survey sponsor, the sample balance, the sample size, and the sampling method. Their results indicated that perceptions of trustworthiness were minimally influenced by survey quality information compared to respondent characteristics, although sample size and balance mattered. They also found that the importance of methodological details increased in accordance with the cognitive abilities of the respondents.

With this paper, we contribute to the literature by attempting to replicate the vignette experiment of Stadtmüller et al. (2022) in two other countries (the US and Hungary) and extend their approach to a more controversial issue (effects of migration). Choosing the US as a location was justified given that a substantial number of previous research studies in this field have been conducted there, making it a good benchmark. Our study can show if one can make generalizable claims based on findings from such a well-studied country. Hungary, on the other hand, presents a unique political and social landscape, which is different from many Western countries. For instance, Hungary's democracy is relatively younger compared to established democracies like Germany or the US. The maturity of a democracy can significantly influence public attitudes towards institutions (Badescu 2004) which may include polling institutions and surveys in general.

Study 1 aimed to determine the extent to which the findings of Stadtmüller et al. (2022) replicate and generalize to other countries. On the other hand, Study 2's main question was whether the formation of trust in surveys remains consistent when examining both contested and uncontested issues. While previous findings have provided suggestive evidence regarding this question, our study was specifically designed to offer more precise comparisons by conducting identical experiments simultaneously in the same countries. The data collection and the study were pre-registered be-

fore the data collection. The pre-registration for the study is available here: <https://osf.io/v2x9w>

2 Hypotheses

The formulation of our hypotheses was guided by the research questions raised in Stadtmüller et al. (2022)'s original study. Their first question was related to the relative importance of survey quality information for the perceived trustworthiness of the survey result. Their results and other earlier findings (Kuru et al. 2017, 2020) suggest that the relative importance of survey quality information is smaller than the role of individual characteristics. Individual characteristics cover gender, age, highest level of education, and trust in science. We hypothesized that:

H1: Individual characteristics will explain more of the variance of trust in survey results than vignette characteristics (i.e. survey quality information) in both studies.

Even if individual characteristics carry more weight in influencing the development of trust in survey results, methodological quality may still moderate individual biases (Kuru et al. 2017, 2020) and certain aspects of survey quality may be more relevant than others (Stadtmüller et al. 2022). Stadtmüller et al. (2022) reported that sample size and sample representativeness significantly affected perceived trustworthiness; thus, we hypothesized that:

H2: Sample size and sample representativeness will have the strongest effect on trust in survey results among survey quality information. A larger sample size and when the survey is described as representative will indicate higher levels of trust.

Earlier evidence suggests that the use of survey quality information is not uniform in every segment of society (Kuru et al. 2017, 2020; Stadtmüller et al. 2022). In particular, those with higher cognitive ability may be more likely to be familiar with and pay more attention to the methodological details of a survey, since cognitive abilities form the fundamental basis for engaging in deep information processing and understanding of incoming messages (Ackerman 1988; Chaiken 1980). More educated respondents were found to be more likely to identify high-quality polls accurately (Kuru et al. 2020), and give more relevance to survey quality information (Stadtmüller et al. 2022), thus we hypothesize that:

H3: Those with lower levels of education or lower levels of cognitive ability will use survey characteristics

less when assessing trust in survey results (i.e. vignette characteristics will explain less of the variance of trust in survey results).

The first three hypotheses were applied to both studies. In H4, we turn specifically to Study 2 and the impact of changing the original study's topic to a more contested and salient issue. First, as the level of politicization increases regarding the survey results, the role of individual (e.g., political) differences may become more significant in shaping the development of trust in the survey results. Accordingly, the survey quality aspects may matter less in salient politicized issues. Second, building on the findings of Kuru et al. (2017, 2020), we expected to find signs of motivated reasoning when assessing trust in the survey results. In particular, respondents' ideological and issue positions were expected to have a strong direct effect on perceived trustworthiness.

H4: In Study 2, individual political variables will explain the most of the variance of trust in survey results

We did not formulate a hypothesis on country differences, we rather focused on the change of variance explained by individual characteristics when the topic is politicized (see the difference between Study 1 and Study 2). Since political variables were not included in Study 1, we decided not to compare the two studies regarding explained variance.

3 Methods and Data

3.1 Experimental Designs

The experimental design in Study 1 replicated that of Stadtmüller et al. (2022). We used the same vignettes: descriptions of a fictitious survey on whether to grant commuters higher tax benefits as a result of escalating gasoline prices (see Table A1 for an example). Five dimensions (four with two and one with 3-levels) were varied in the text: the *survey sponsor*, *results of the survey*, *sample balance*, *sample size*, and *sampling method* (see Table 1). The results of the survey were used as a between-subject dimension only, whereas all other dimensions were allowed to vary within respondents as well. Thus, each respondent saw either '61% good' or '61% bad' vignettes. This design option was part of the original study thus it was important for us not to deviate from it. We believe the original study's authors' rationale to use survey result as a between-subject dimension was that the original study focused on the the role of methodological information and not on how much it affects confidence when the results are closer to what we think about

Table 1*Summary of Study 1 and Study 2*

| Study | Dimension | Level | Design |
|---------|-----------------------|--|--------------------|
| Study 1 | Survey source/sponsor | AAA; ATRI/Hungarian Autoclub; TDRC | Within and between |
| | Survey result | 36% favor higher tax allowances; 72% favor higher tax allowances | Between only |
| | Sample balance | Nothing mentioned; “representative” | Within and between |
| | Sample size | 100; 1000; 1.000 respondents | Within and between |
| | Sampling method | Nothing mentioned; Random selection added | Within and between |
| Study 2 | Survey source/sponsor | CNN; Fox News/Magyar Nemzet; HVG | Within and between |
| | Survey result | 61% say it is bad; 61% say it is good | Between only |
| | Sample balance | Nothing mentioned; “representative” | Within and between |
| | Sample size | 100; 1000; 1.000 respondents | Within and between |
| | Sampling method | Nothing mentioned; Random selection added | Within and between |

Note: AAA=The American Automobile Association, ATRI=American Transportation Research Institute, TDRC=Transportation Development Research Center

the world. In the original study, the first two contextual dimensions were manipulated to assess the vignettes’ ecological validity, while three other methodological dimensions were identified as “important quality indicators by survey statisticians” (Stadtmüller et al. 2022). Admittedly some of these measures are rarely used as survey quality indicators in science (sample size, “representativity”), yet they may still serve as familiar reference points for an average member of the general population. That is because these terms are indeed frequently mentioned in news and public discourse concerning the credibility and accuracy of polls and surveys. We needed to deviate from the original texts of Stadtmüller et al. (2022) in one of the dimensions and customise the names of the survey sponsors to align with each country’s context.¹ No changes were made to other dimensions. This design resulted in 48 unique vignettes. Each respondent received a random set of four vignettes and was asked to assess their level of trust in the survey results using a 7-point scale, ranging from 1 (not at all trustworthy) to 7 (completely trustworthy).² Note, that respondents had assessed their level of trust for four individual vignettes and not, for instance, contrasting two survey reports as in a conjoint experiment.

Study 2 used the same experimental design, but the survey descriptions had different contextual characteristics (see Table 1). The survey in Study 2 was about the perceived economic impact of immigration (61% of *Americans/Hungarians say that it is generally good/bad for the American/Hungarian economy that people come to live here*

from other countries). Our goal in using immigration as a central issue was to provide a more politicized context. Recently, debates surrounding immigration policies, border control, and the integration of migrants have sparked passionate discussions, and parties across the political spectrum have formulated distinct stances on immigration in both Europe and the United States (Baker and Edmonds 2021; Buonfino 2004; Grande et al. 2019). Specifically, Donald Trump in the US (Campani et al. 2022; Flores 2018) and Viktor Orbán in Hungary (Bíró-Nagy 2018; Boda and Rakovics 2022) prominently placed immigration as a central issue in their political agenda. We chose 61% instead of the proportion used in Study 1 (72 vs. 36%), because we suspected that tighter results would appear as a more realistic outcome. Accordingly, we used *FoxNews* and *CNN* in the US and *Magyar Nemzet* (Hungarian Nation) and *HVG* (Weekly World Economy) in Hungary as survey sponsors.³ Besides this, the methodological dimensions were constant between Studies 1 and 2.

3.2 Measures for Controlling Purposes

The measures used by Stadtmüller et al. (2022) were adapted for our study. We measured age (in years), gender (male coded as 1, and female coded as 2), highest level of education (high school or lower and diploma), and trust in

¹ In contrast to Stadtmüller et al. (2022) we used real research institutions to make the vignettes sound more realistic.

² ‘Do not know’ or ‘Refused’ options were not displayed, instead respondents were allowed to skip the question.

³ *Magyar Nemzet* is among the most longstanding daily newspapers in Hungary. After several recent owner changes, *Magyar Nemzet* is now one of the leading pro-government (conservative) media outlets (Batorfy and Urban 2020) in Hungary. *HVG*, another longstanding newspaper, in contrast, is recognized for its left-leaning political stance and consistently critical viewpoints toward the current government.

science (1–4 graded scale).⁴ In addition to the strict replication analysis, we used additional measures. Stadtmüller et al. (2022) used the educational level as a proxy for cognitive abilities. Although research has shown that cognitive ability is strongly correlated with educational level (Ceci 1991), we decided to add a more finite measure of mathematical cognition to our survey and implemented the Subjective Numeracy Scale (SNS) (Fagerlin et al. 2007).⁵ We calculated the mean values of the scale items. The Cronbach Alpha value was 0.86 in Hungary and 0.81 in the US. For Study 2, political interest (on a scale of 1 to 5), left–right ideological position (on a scale of 1 to 7), and issue position (beliefs about the perceived economic impact of immigration with a standard question of the ESS⁶) were measured. The full questionnaire can be found in the pre-registration.

3.3 Analytical Strategy

3.3.1 Study 1 – Replication Analysis

In our replication analysis (Study 1), we followed the approach of Stadtmüller et al. (2022). Multilevel mixed-effects linear regression models were fitted hierarchically. This allowed us to test H1 by differentiating between the relative importance of within- and between-subject variance, as well as the variance explained on different analytical levels. We started with null models and then added vignette- (Model 1), and respondent-level characteristics (Model 2). We proceeded by adding interactions between the vignette-level variables (Model 3). The same model was used to test H2. To test H3, this approach was replicated by splitting the sample into low and highly-educated respondents. We deviated from the original study's approach in two ways. First, the relatively small sample size did not allow for an analysis of variance (ANOVA). Second, due to the problem of sample sizes and skewed distribution of the educational level variable (towards the highly educated), we faced limitations in confidently dividing the sample into

three educational groups. As a result, we opted to use two groups. As an extension of the original study, we replicated this last step using numeracy skills (SNS scores) to split the sample. We split the sample using mean SNS scores. As a robustness check, we also report the estimates of multilevel mixed-effects ordinal logistic regression models and the results of models where only the first vignettes answered by respondents were included in the analysis.

3.3.2 Study 2 – Extension

To ensure comparability, we used the same analytical approach used in Study 1. Nevertheless, our interest extends to examining the extent to which individual political characteristics may modify these findings. To test H4, we used an additional model after including the individual-level variables of Study 1 and added political interest and the left–right position. Left–right was categorized into left (1–3), middle and undecided (4 and DK/NA answers), and right (5–7). We used left–right as a factor variable in our models.

To measure issue agreement, we calculated whether the respondent's issue position matched the survey result (1) or not (0). We considered matching as 0–3 or 7–10 scores (strong leaning) for the issue question.

Although we did not develop specific hypotheses about the potential moderating effects of the political variables, we ran some additional exploratory analyses. Based on Kuru et al. (2017), we expected to find that political interest and the left–right position would moderate the effect of motivated reasoning on trust in survey results. We expected that motivated reasoning among those with higher levels of political interest and those on the right side of the political spectrum would matter more to trust in surveys. We tested these effects with interaction terms between political interest, left–right ideology, and the match between the respondent's issue position and the survey result. We also examined whether the sponsor of the study and the political-ideological position of the respondents interacted with each other. We carried out the analysis using the *lme4* (Bates et al. 2015) and related packages in R. See the R code of the analysis in the Data Availability section.

3.4 Data

Both Study 1 and Study 2 drew on two online surveys conducted in the United States and Hungary. The U.S. dataset utilized in this study was derived from the Harvard Digital Lab for Social Science (DLABSS), a non-probability-based panel operated by Harvard University. The DLABSS consists of voluntary survey respondents who are primarily recruited through social media and other freely available

⁴ For those who did not answer that question, we imputed the mean value of trust in other institutes and actors (politicians, political parties, media). The Cronbach Alpha of the four questions was 0.78 in both Hungary and the US.

⁵ In the US survey, we inadvertently omitted one item [When you hear a weather forecast, do you prefer predictions using percentages (e.g., “there will be a 20% chance of rain today”) or predictions using only words (e.g., “there is a small chance of rain today”)?] from the SNS scale. Consequently, we only utilized the complete items that were available in both countries.

⁶ Would you say it is generally bad or good for [country]'s economy that people come to live here from other countries? 0 – Bad for the economy, 10 – Good for the economy

sources. The pool of DLABSS volunteers is rapidly expanding and currently comprises approximately 30,000 individuals. Studies (Strange et al. 2019) have demonstrated that volunteer panels such as DLABSS can successfully replicate both classic and contemporary social science findings, while also exhibiting high levels of overall response quality that are comparable to paid subjects. As outlined in the pre-registration, the determination of this sample size relied on an estimated power analysis, employing an ANOVA design with repeated measures and within-between interactions, using G*Power (Faul et al. 2007). Based on the a priori calculations, to achieve a 0.95 power with a 0.15 (small) effect size, a minimum sample size of 432 was required in each study, adding up to 864 for the whole survey. 14,009 panel members were invited, and 1085 of them started the survey (participation rate: 12.9%). 1037 members completed the survey, yielding a completion rate of 95.6%. After data curation (removing respondents with missing data), 877 respondents were included in our sample (447 in Study 1 and 430 in Study 2). The fieldwork was conducted between May 14 and May 30, 2023.

The survey for the Hungarian study was conducted through an online polling company, NRC, using a non-probability access panel. The NRC panel comprises members of over 140,000 individuals. We employed quota sampling using quotas for gender, age, and geographical region. The respondents of NRC surveys receive regular incentives from the company. The target sample size was based on the same power calculation (864). 11,091 members were invited, 977 started the survey (participation rate: 8.8%), and 900 completed it (completion rate: 92.1%). After data curation, we had 884 respondents in the analyzed sample (444 in Study 1 and 440 in Study 2). Fieldwork was carried out between May 16 and May 23, 2023.

Importantly, compared to Stadtmüller et al. (2022), our study's sample size and statistical power were lower. Due to the lower statistical power, we could only detect more robust correlations; thus, we expected to find fewer significant correlations in our data. However, this lower statistical power does not cause hypothesis-testing problems. The data used in this study is openly available for further analysis. See the Data Availability section.

4 Results

4.1 Main Results

The mean of the trustworthiness question used as a dependent variable was 4.11 in the Hungarian commuting vignettes and 3.76 in the migration vignettes, respectively in a 1-7 scale. Thus, the political question yields a lower

mean trustworthiness score. This difference can also be observed in US data (3.6 vs 3.39). The vignette-level results also show that respondents in the Hungarian sample gave higher trust ratings on average than those in the US, regardless of the survey. However, the overall measured value is significantly lower than the mean of 4.5 obtained in the Stadtmüller et al. (2022) study.

Our first hypothesis was that individual characteristics explain more of the variance in trustworthiness ratings than vignette characteristics do. In the Hungarian sample measuring the trustworthiness of the commuting survey results, 81 percent of the variance was explained by individual characteristics. In contrast, 73 percent is explained by the individual level in the more divisive migration topic⁷. Similarly, in the US sample, although the variance explained at the individual level is lower (63 percent in both vignette designs), individual differences matter more than survey quality information. Thus, H1 was confirmed, and the findings of the original study were replicated in both countries and across the two topics (see null models in Tables 2 and 3).

Including vignette features increased the model's explanatory power in both countries and studies but to different degrees. In the Hungarian commuting vignettes, the total explanatory power was less than 1 percent after including survey characteristics, compared to 7.7 percent in the US sample that examined the same question. In Study 2, the total explanatory power was 6 percent after adding the survey information variables in both countries.

Based on H2, we expected to find that the sample size and sample representativeness matter the most among the survey quality information (see Models with vignette dimensions in Tables 2 and 3). Across both studies and consistently in both countries, as the sample size mentioned in the survey description increased, the perceived trustworthiness of the survey results also increased. The B values suggest that the sample size had a more substantial effect in the US survey, but no significant differences in B values were observed along the survey topic. The importance of representativeness was significant (and positive) only in the Hungarian commuting study. Mentioning the random selection of the respondents yielded mixed results, with two significant and positive effects out of the four. Comparing the B values, survey size consistently had the most substantial effect. Overall, the results partly confirmed H2 and replicated the results found by Stadtmüller et al. (2022) regarding sample size but not sample balance. Concerning other survey information, the survey result was only significant in Study 2 in Hungary, where the mean of survey trust significantly increased (by 0.76 points) when the results of the survey indicated a negative economic effect of immi-

⁷ The two explained variances were significantly different from each other based on bootstrap estimates

Table 2

Results of Study 1 – replication

| | Null model (HU) | | | Model with vignette dimensions (HU) | | | Model with vignette dimensions and controls (HU) | | | Null model (US) | | | Model with vignette dimensions (US) | | | Model with vignette dimensions and controls (US) | | |
|------------------------------------|-----------------|-----------|----------|-------------------------------------|-----------|----------|--|-----------|----------|-----------------|-----------|----------|-------------------------------------|-----------|----------|--|-----------|----------|
| | <i>b</i> | <i>SE</i> | <i>p</i> | <i>b</i> | <i>SE</i> | <i>p</i> | <i>b</i> | <i>SE</i> | <i>p</i> | <i>b</i> | <i>SE</i> | <i>p</i> | <i>b</i> | <i>SE</i> | <i>p</i> | <i>b</i> | <i>SE</i> | <i>p</i> |
| Constant | 4.11 | 0.08 | < 0.001 | 4.18 | 0.26 | < 0.001 | 2.37 | 0.51 | < 0.001 | 3.6 | 0.06 | < 0.001 | 3.04 | 0.24 | < 0.001 | 1.62 | 0.47 | 0.001 |
| <i>Vignette dimensions</i> | | | | | | | | | | | | | | | | | | |
| Survey results: 36% in favor | – | – | – | –0.16 | 0.15 | 0.286 | –0.2 | 0.14 | 0.151 | – | – | – | 0.02 | 0.13 | 0.905 | 0.05 | 0.12 | 0.689 |
| Sponsor: TDRC (HU) / ATRI (US) | – | – | – | –0.07 | 0.04 | 0.081 | –0.07 | 0.04 | 0.097 | – | – | – | –0.24 | 0.04 | < 0.001 | –0.25 | 0.04 | < 0.001 |
| Sample size: 1000 | – | – | – | 0.24 | 0.05 | < 0.001 | 0.24 | 0.05 | < 0.001 | – | – | – | 0.69 | 0.05 | < 0.001 | 0.69 | 0.05 | < 0.001 |
| Sample size: 10000 | – | – | – | 0.40 | 0.05 | < 0.001 | 0.40 | 0.05 | < 0.001 | – | – | – | 1.17 | 0.06 | < 0.001 | 1.17 | 0.05 | < 0.001 |
| Representativeness: mentioned | – | – | – | 0.09 | 0.04 | 0.022 | 0.08 | 0.04 | 0.029 | – | – | – | 0.04 | 0.04 | 0.426 | 0.03 | 0.04 | 0.465 |
| Random selection: mentioned | – | – | – | –0.05 | 0.04 | 0.234 | –0.04 | 0.04 | 0.241 | – | – | – | 0.15 | 0.04 | 0.001 | 0.14 | 0.04 | 0.001 |
| <i>Controls</i> | | | | | | | | | | | | | | | | | | |
| Age (in years) | – | – | – | – | – | – | 0.00 | 0.00 | 0.935 | – | – | – | – | – | – | –0.01 | 0.00 | 0.054 |
| Gender: Male | – | – | – | – | – | – | 0.39 | 0.15 | 0.008 | – | – | – | – | – | – | 0.1 | 0.13 | 0.466 |
| Education: Diploma | – | – | – | – | – | – | –0.44 | 0.15 | 0.004 | – | – | – | – | – | – | 0.14 | 0.14 | 0.3 |
| Trust in science | – | – | – | – | – | – | 0.64 | 0.09 | < 0.001 | – | – | – | – | – | – | 0.59 | 0.09 | < 0.001 |
| <i>Variances of random effects</i> | | | | | | | | | | | | | | | | | | |
| Variance: constant | 2.36 | | | | 2.37 | | | 2.09 | | 1.63 | | | | 1.69 | | | 1.49 | |
| Variance: residual | 0.56 | | | | 0.53 | | | 0.53 | | 0.97 | | | | 0.71 | | | 0.71 | |
| Proportion of Level 1 variance | 19% | | | | 18% | | | 2% | | 37% | | | | 30% | | | 32% | |
| Proportion of Level 2 variance | 81% | | | | 82% | | | 80% | | 63% | | | | 7% | | | 68% | |
| <i>Model fit</i> | | | | | | | | | | | | | | | | | | |
| Variance explained (Level 1) | – | | | | 5% | | | 5% | | – | | | | 27% | | | 27% | |
| Variance explained (Level 2) | – | | | | 0% | | | 11% | | – | | | | 0% | | | 9% | |
| Variance explained (overall) | – | | | | 1% | | | 1% | | – | | | | 8% | | | 15% | |
| Deviance | 5281 | | | | 5201 | | | 5145 | | 5929 | | | | 5506 | | | 5451 | |

Note: TDRC=Transportation Development Research Center, ATRI=American Transportation Research Institute

Table 3

Results of Study 2 – extension

| | Null model (HU) | | | Model with vignette dimensions and controls (HU) | | | Null model (US) | | | Model with vignette dimensions (US) | | | Model with vignette dimensions and controls (US) | | |
|------------------------------------|-----------------|-----------|----------|--|-----------|----------|-----------------|-----------|----------|-------------------------------------|-----------|----------|--|-----------|----------|
| | <i>b</i> | <i>SE</i> | <i>p</i> | <i>b</i> | <i>SE</i> | <i>p</i> | <i>b</i> | <i>SE</i> | <i>p</i> | <i>b</i> | <i>SE</i> | <i>p</i> | <i>b</i> | <i>SE</i> | <i>p</i> |
| Constant | 3.76 | 0.08 | < 0.001 | 1.76 | 0.28 | < 0.001 | – | 0.62 | 0.971 | 3.39 | 0.07 | < 0.001 | 2.77 | 0.27 | < 0.001 |
| <i>Vignette dimensions</i> | | | | | | | | | | | | | | | |
| Survey results: 61% say it is bad | – | – | – | 0.76 | 0.16 | < 0.001 | 0.69 | 0.14 | < 0.001 | – | – | – | 0.02 | 0.14 | 0.908 |
| Sponsor: HVG/Fox | – | – | – | 0.26 | 0.05 | < 0.001 | 0.27 | 0.05 | < 0.001 | – | – | – | –0.11 | 0.05 | 0.05 |
| Sample size: 1000 | – | – | – | 0.29 | 0.06 | < 0.001 | 0.29 | 0.06 | < 0.001 | – | – | – | 0.55 | 0.07 | < 0.001 |
| Sample size: 10000 | – | – | – | 0.44 | 0.06 | < 0.001 | 0.45 | 0.06 | < 0.001 | – | – | – | 0.97 | 0.07 | < 0.001 |
| Representativeness: mentioned | – | – | – | 0.01 | 0.05 | 0.88 | 0.01 | 0.05 | 0.847 | – | – | – | 0.06 | 0.05 | 0.278 |
| Random selection: mentioned | – | – | – | 0.14 | 0.05 | 0.006 | 0.13 | 0.05 | 0.007 | – | – | – | 0.11 | 0.05 | 0.054 |
| <i>Controls</i> | | | | | | | | | | | | | | | |
| Age (in years) | – | – | – | – | – | – | –0.02 | 0.01 | 0.003 | – | – | – | 0.00 | 0.00 | 0.975 |
| Gender: Male | – | – | – | – | – | – | 0.08 | 0.15 | 0.587 | – | – | – | 0.15 | 0.14 | 0.294 |
| Education: Diploma | – | – | – | – | – | – | –0.47 | 0.15 | 0.002 | – | – | – | 0.19 | 0.15 | 0.205 |
| Trust in science | – | – | – | – | – | – | 0.73 | 0.11 | < 0.001 | – | – | – | 0.59 | 0.11 | < 0.001 |
| <i>Variances of random effects</i> | | | | | | | | | | | | | | | |
| Variance: constant | 2.56 | – | – | – | 2.41 | – | – | 1.91 | – | – | 2.01 | – | – | 1.98 | 1.48 |
| Variance: residual | 0.93 | – | – | – | 0.87 | – | – | 0.87 | – | – | 1.17 | – | – | 1.01 | 1.01 |
| Proportion of Level 1 variance | 27% | – | – | – | 27% | – | – | 31% | – | – | 37% | – | – | 34% | 5% |
| Proportion of Level 2 variance | 73% | – | – | – | 74% | – | – | 69% | – | – | 63% | – | – | 66% | 59% |
| <i>Model fit</i> | | | | | | | | | | | | | | | |
| Variance explained (Level 1) | – | – | – | – | 7% | – | – | 7% | – | – | – | – | – | 14% | 14% |
| Variance explained (Level 2) | – | – | – | – | 0% | – | – | 25% | – | – | – | – | – | 0% | 26% |
| Variance explained (overall) | – | – | – | – | 6% | – | – | 2% | – | – | – | – | – | 6% | 22% |
| Deviance | 5936 | – | – | – | 5825 | – | – | 5726 | – | – | 6036 | – | – | 5824 | 5707 |

gration. The results were mixed among the sponsors of the survey. Significant effects were found in both studies in the US, and Study 2 in Hungary. However, B values were relatively low, indicating that who commissioned the survey had little overall impact on perceived trustworthiness (see the “Model with vignette dimensions” columns in Tables 2 and 3). We will return to the impact of the sponsor at the end of the results section.

Control variables were included in the third modelling step. In both countries and across the two studies, general trust in science increased trust in survey results. No clear trend could be identified for the other demographic variables. Lower levels of education decreased trustworthiness ratings in Hungary regardless of the topic. This result was consistent with that reported by Stadtmüller et al. (2022). The results presented above demonstrate consistent robustness when tested across multilevel mixed-effect ordered logistic regression models (see Tables A12 and A13). We further tested the robustness of our findings by changing the educational level to the SNS scores. Our findings remained consistent in these models, although, in contrast to education, SNS scores were not associated with perceived trustworthiness, indicating that education is not an explicit proxy for SNS (see Tables A2 and A3). We also tested how robust the results were when analysing only the vignettes that respondents received first. Vignette ordering was not available in the Hungarian database, so we could only test this on the US sample. The results obtained were consistent with the models run on the full vignette set for both topics (see Table A14).⁸

4.2 Results by Education Level

In H3, we hypothesized that survey-quality information is more critical for highly educated people or people with higher cognitive abilities. In all cases, the models run for higher education groups had higher variance at the vignette level than for lower education groups, and perceived trustworthiness was better explained by the vignette-level variables (see Tables A6 and A7). Moreover, in the higher-educated subsamples, the B values for the sample size were generally higher than those in the corresponding rows in the lower-educated groups. In contrast, no similar differences were found between the low and high SNS score groups (Tables A8 and A9). When education was used as a proxy

for cognitive skills, H3 was confirmed and replicated the results of the original study.

4.3 The impact of Political Variables

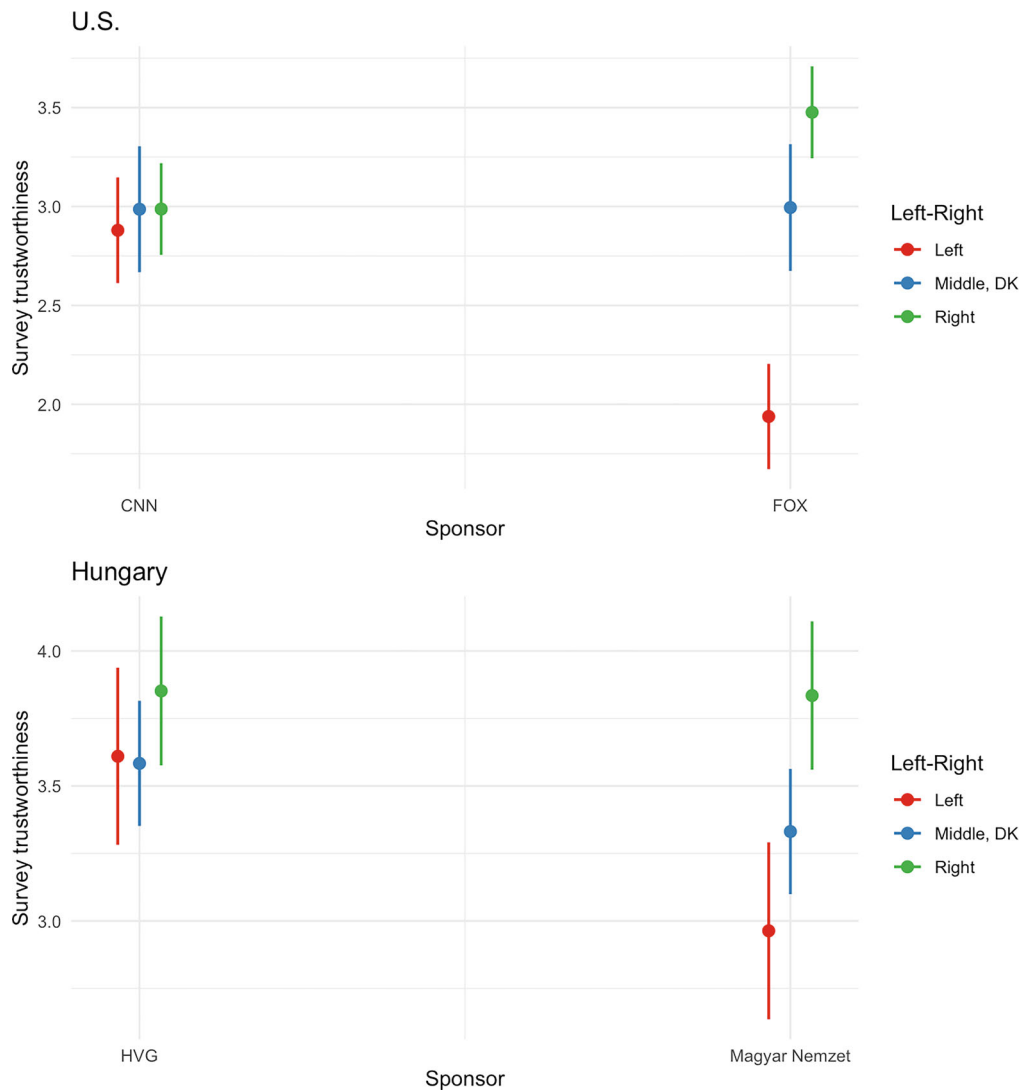
Our final (fourth) hypothesis was that individual-level political variables would be the most important determinants of trustworthiness ratings when a topic is more contested. Our hypothesis is not confirmed in the Hungarian case; the role of the political variables is indeed important but not larger than that of the other individual-level variables. In the Hungarian sample, the total explanatory power increased from 6 to 15.8 percent when we included non-political background variables and 20.3 percent with political background variables. In the US data, however, the overall explanatory power increased from 6 to 8.2 percent after including non-political individual variables, and 21.7 percent when we included political controls (see Table 3 and Table A10). These are partial confirmations of H4.

An additional sign of the political factors at play is that respondents in both countries showed a higher likelihood of trusting the results when the majority's stance in the text was that immigration has a negative economic impact on the country. Moreover, respondents had a higher trust in the survey results if their views matched the results from both countries. The B value of 1.4 in the US is considered exceptionally high, given that the range of the dependent variable was only six units. On the left-right dimension, we found that right-wingers had higher trust in the survey results than left-wingers in both countries. Political interest was significant only in the Hungarian sample, with people with higher political interest showing higher trust in the results.

As a complementary analysis, we examined whether the interaction of agreement with the other two political background variables increased the explanatory power of our models. However, these interactions were not significant (Table A11).

The final question we examined was whether the survey's sponsor and the respondents' political affiliation interacted with each other to determine the trustworthiness of the survey. We expected that those on the political right would have lower trust when they see results from CNN/HVG and leftists would have lower trust when they have to evaluate a survey sponsored by Fox News or Magyar Nemzet. We added survey sponsor and left-right scale interaction terms to our models. The interaction term was significant in both countries. We could only observe substantial differences in the case of right-leaning media outlets (see Fig. 1 for the marginal predictions). People with left political orientation gave a trust score under 2 to surveys sponsored by FOX, while right supporters gave nearly 3.5 scores for the

⁸ In the fourth step of the modeling, we added the interaction of the survey characteristics. However, these interaction terms were not significant in any of the models (Tables A4 and A5). These results are partly consistent with the findings of Stadtmüller et al. (2022), who also reported null or weak effects.

**Fig. 1**

The interaction between survey sponsor and the left-right ideological position of the respondent in Study 2

same study. We could find the same pattern in the Hungarian sample but with a smaller difference between the two political sides.

5 Discussion

This study examined the factors that drive trustworthiness assessments of survey results in contested and uncontested topics across two countries. Study 1 replicated the study by Stadtmüller et al. (2022) and investigated the degree to which individuals depend on information about survey quality and the specific indicators they prioritize when expressing their perceived trust in a survey result. Study 2 ex-

tended the original study by asking the same questions in a more contested and politicized issue. Our findings show that most results of Stadtmüller et al. (2022) travel well across these survey topics and the two countries. As for the main findings, when evaluating the trustworthiness of survey results, individual characteristics hold more weight compared to survey quality information, larger sample sizes enhance trust, and highly educated individuals assign more importance to methodological information in both countries, regardless of the topic. Nevertheless, when the survey topic was more politicized, the respondent's political characteristics and issue position strongly shaped their trust judgments, which resonates well with findings from research on

motivated reasoning (Kuru et al. 2017, 2020; Madson and Hillygus 2020).

The significance of methodological information persisted in Study 2, suggesting that people consider survey quality information even when their own political or issue position biases their perceptions. In particular, sample size significantly impacted trust. This evidence suggests that across different cultures, the sample size might be one methodological aspect of a survey consistently considered in trust evaluations, while other factors such as representativeness might not be universally applied. Our findings regarding the moderating role of education corroborate the earlier results. Those with higher levels of education tended to attribute greater importance to survey quality information when evaluating their level of trust. Nevertheless, the difference between the two groups was rather small, indicating that some methodological details of a survey are available during judgments, even for low-educated populations. The findings underscore the significance of transparency and methodological disclosure in survey reporting (Bhatti and Pedersen 2016) as they can enhance trust directly or potentially mitigate partisan-based motivational biases (Kuru et al. 2017). Furthermore, even if our findings suggest that survey quality information is judged and taken into consideration by many, enhancing survey literacy and educating individuals on how to discern the validity of survey results remain crucial tasks in science communication (Stadtmüller et al. 2022).

It remains unclear whether differences between education groups reflect differences in cognitive skills. We found that subjective numeracy skills were unrelated to trust assessment. Future research should use other measures of cognitive abilities (e.g., Cognitive Reflection Test [CRT] and Need for Cognition) and replicate our analysis.

Our study adds nuance to the findings of the original study by showing that the extent to which individuals utilize survey-quality information is contingent upon the level of controversy and politicization surrounding the survey topic. These results resonate well with some earlier experiments (Kuru et al. 2017, 2020; Madson and Hillygus 2020) and findings of the motivated reasoning literature (Bolsen et al. 2014; Kunda 1990; Redlawsk 2002; Tsfati 2001; Chia and Chang 2017; Searles et al. 2018). Particularly, in the US survey, individual-level political variables accounted for most of the variance (other individual characteristics [e.g. trust in science] had stronger explanatory power in Hungary). The American results align well with the dynamics of polarization observed in American society (Iyengar and Westwood 2015).

This leads us to the issue of generalizability. It was assumed that cultural factors are likely to have an impact on the level of trust individuals have in scientific, polling, or media institutions, their knowledge about and consideration

of survey quality information, and their inclination towards motivated reasoning. This study showed that some of the basic heuristics of trust assessment are common in these three countries, while there are also relatively small differences in the weights of the different factors. This continues to be an important avenue for future studies to examine the development of trust in surveys in cross-national settings with greater cultural diversity.

Our study comes with some limitations, which offer opportunities for future research. First, for sample representativeness and random selection, we manipulated the *disclosure* of these characteristics instead of using more direct (representative vs. non-representative) comparisons. We assume that Stadtmüller et al. (2022) chose not to use categories such as non-representative or non-random selection it is unlikely that survey reports in real-world scenarios would feature such descriptions. Yet, it is true that some respondents may have assumed that those surveys with no information were representative or applied random selection as well. Future research could experiment with stronger manipulations. Second, the extent to which our findings can be extrapolated to the general population remains unclear. Both panels used in this study were nonprobability-based. Members of such panels are typically more highly educated and may be more interested in public affairs, indicating that the way they perceive survey results can be systematically different from that of the non-panel member population. The use of probability-based panels can help tackle this issue. Second, although the vignettes covered the most important indicators of survey quality, future research could explore additional factors that may affect the public. For instance, online surveys are often connoted with lower quality in scientific (Baker et al. 2010) and public discourse (Cohn 2019), which suggests that the survey mode may carry significant implications for credibility assessments.

In conclusion, the development of trust in the survey results is a complex mental process that involves the consideration of several factors. Trustworthiness assessments are shaped by perceived methodological quality, the topic of the survey, the sponsor or platform on which the survey was published, and several political and non-political individual characteristics. Our research aided in determining the significance of these elements and the degree to which the development of trust in survey results can be universally applied across diverse cultures.

Acknowledgements We thank Századvég Foundation and Harvard University for funding the data collection.

Ethics approval Ethical review and approval were waived for this Hungarian study because data was anonymized within the fieldwork of the study. The results do not allow identification of the individuals involved in the study. The authors and the fieldwork agency managed all information collected in accordance with the General Data Protection Regulation (GDPR).

The US study was approved by Harvard University's Committee on the Use of Human Subjects as exempt as protocol IRB23-0212.

Author Contributions Conceptualization Á.S. and Z.K.; theoretical background: Á.S.; methodology: Á.S. and Z.K.; formal analysis, Z.K.; data curation, Z.K.; writing—original draft preparation Á.S. and Z.K.; writing—review and editing: Á.S. and Z.K.; funding acquisition: Á.S.

Conflicts of Interest The authors declare no conflict of interest.

References

- AAPOR (2021). AAPOR Code of Professional Ethics and Practices. https://aapor.org/wp-content/uploads/2022/12/AAPOR-2020-Code_FINAL_APPROVED.pdf
- Ackerman, P.L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General*, 117, 288–318. <https://doi.org/10.1037/0096-3445.117.3.288>.
- Alabrese, E. (2022). National polls, local preferences and voters' behaviour: Evidence from the UK General Elections. *SSRN Scholarly Paper.*, 4265932 <https://doi.org/10.2139/ssrn.4265932>.
- Ansolabehere, S., & Iyengar, S. (1994). Of horseshoes and horse races: Experimental studies of the impact of poll results on electoral behavior. *Political Communication*, 11(4), 413–430. <https://doi.org/10.1080/10584609.1994.9963048>.
- Badescu, G. (2004). Social trust and democratization in the post-communist societies. In *Social capital and the transition to democracy* (pp. 134–153). Routledge. Retrieved November 21, 2023, from <https://www.taylorfrancis.com/chapters/edit/10.4324/9780203428092-12/social-trust-democratization-post-communist-societies-gabriel-badescu>.
- Baker, J. O., & Edmonds, A. E. (2021). Immigration, presidential politics, and partisan polarization among the american public, 1992–2018. *Sociological Spectrum*, 41(4), 287–303. <https://doi.org/10.1080/02732173.2021.1900760>.
- Baker, R., et al. (2010). Research synthesis: Aapor report on online panels. *Public Opinion Quarterly*, 74(4), 711–781. <https://doi.org/10.1093/poq/nfq048>.
- Baker, R., et al. (2016). *Evaluating survey quality in today's complex environment*. American Association for Public Opinion Research.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Batorfy, A., & Urban, A. (2020). State advertising as an instrument of transformation of the media market in Hungary. *East European Politics*, 36(1), 44–65. <https://doi.org/10.1080/21599165.2019.1662398>.
- Bhatti, Y., & Pedersen, R. T. (2016). News reporting of opinion polls: Journalism and statistical noise. *International Journal of Public Opinion Research*, 28(1), 129–141. <https://doi.org/10.1093/ijpor/edv008>.
- Biró-Nagy, A. (2018). Politikai lottóötös: A migráció jelentősége a magyar politikában, 2014–2018. In B. Böcskei & A. Szabó (Eds.), *Várakozások és valóságok. Parlamenti választás 2018* (pp. 269–291). Napvilág Kiadó: MTA TK PTI.
- Blais, A., Gidengil, E., & Nevitte, N. (2006). Do polls influence the vote? In H. Brady & J. Johnston (Eds.), *Capturing campaign effects* (pp. 263–279). Ann Arbor, MI: University of Michigan Press.
- Boda, Z., & Rakovics, Z. (2022). Orban Viktor 2010 es 2020 közötti beszedeinek elemzése. *Szociológiai Szemle*, 32(4), 46–69. <https://doi.org/10.51624/SzocSzemle.2022.4.3>.
- Bolsen, T., Druckman, J. N., & Cook, F. L. (2014). The influence of partisan motivated reasoning on public opinion. *Political Behavior*, 36(2), 235–262. <https://doi.org/10.1007/s11109-013-9238-0>.
- Boudreau, C., & McCubbins, M. D. (2010). The blind leading the blind: Who gets polling information and does it improve decisions? *The Journal of Politics*, 72(2), 513–527. <https://doi.org/10.1017/S0022381609990946>.
- Brodie, M., Parmalee, L. F., Brackett, A., & Altman, D. E. (2021). Polling and democracy. *Public Perspective*, 12, 10–24.
- Buonfino, A. (2004). Between unity and plurality: The politicization and securitization of the discourse of immigration in Europe. *New Political Science*, 26(1), 23–49.
- Campani, G., Fabelo, C. S., Rodriguez Soler, A., & Savin, S. C. (2022). The rise of Donald Trump right-wing populism in the United States: Middle american radicalism and anti-immigration discourse. *Societies*. <https://doi.org/10.3390/soc12060154>. Article 6.
- Ceci, S. J. (1991). How much does schooling influence general intelligence and its cognitive components? a reassessment of the evidence. *Developmental Psychology*, 27, 703–722. <https://doi.org/10.1037/0012-1649.27.5.703>.
- Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, 39, 752–766. <https://doi.org/10.1037/0022-3514.39.5.752>.

- Chia, S.-C., & Chang, T.-K. (2017). Not my horse: Voter preferences, media sources, and hostile poll reports in election campaigns. *International Journal of Public Opinion Research*, 29(1), 23–45. <https://doi.org/10.1093/ijpor/edv046>.
- Cohn, N. (2019). *No one picks up the phone, but which online polls are the answer?* The New York Times. <https://www.nytimes.com/2019/07/02/upshot/online-polls-analyzing-reliability.html>
- Couper, M.P. (2013). Is the sky falling? new technology, changing media, and the future of surveys. *Survey Research Methods*, 7(3), 145–156. <https://doi.org/10.18148/srm/2013.v7i3.5751>.
- Dawson, S. (2022). Poll wars: Perceptions of poll credibility and voting behaviour. *The International Journal of Press/Politics*. <https://doi.org/10.1177/19401612221087181>.
- Fagerlin, A., Zikmund-Fisher, B.J., Ubel, P.A., Jankovic, A., Derry, H.A., & Smith, D.M. (2007). Measuring numeracy without a math test: Development of the subjective numeracy scale. *Medical Decision Making*, 27(5), 672–680. <https://doi.org/10.1177/0272989X07304449>.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>.
- Flores, R.D. (2018). Can elites shape public attitudes toward immigrants?: Evidence from the 2016 US presidential election. *Social Forces*, 96(4), 1649–1690. <https://doi.org/10.1093/sf/soy001>.
- Gallup, G., & Rae, S.F. (1940). *The pulse of democracy: The public-opinion poll and how it works*. New York: Simon & Schuster.
- Grande, E., Schwarzbozl, T., & Fatke, M. (2019). Politicizing immigration in Western Europe. *Journal of European Public Policy*, 26(10), 1444–1463. <https://doi.org/10.1080/13501763.2018.1531909>.
- von Hermanni, H., & Lemcke, J. (2017). *A review of reporting standards in academic journals – a research note*. Survey Methods: Insights from the Field (SMIF). <https://doi.org/10.13094/SMIF-2017-00006>.
- Huber, B., Barnidge, M., de Zuniga, G.H., & Liu, J. (2019). Fostering public trust in science: The role of social media. *Public Understanding of Science*, 28(7), 759–777. <https://doi.org/10.1177/0963662519869097>.
- Iyengar, S., & Westwood, S.J. (2015). Fear and loathing across party lines: New evidence on group polarization. *American Journal of Political Science*, 59(3), 690–707. <https://doi.org/10.1111/ajps.12152>.
- Krause, N.M., Brossard, D., Scheufele, D.A., Xenos, M.A., & Franke, K. (2019). Trends—americans' trust in science and scientists. *Public Opinion Quarterly*, 83(4), 817–836. <https://doi.org/10.1093/poq/nfz041>.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108, 480–498. <https://doi.org/10.1037/0033-2909.108.3.480>.
- Kuru, O., Pasek, J., & Traugott, M.W. (2017). Motivated reasoning in the perceived credibility of public opinion polls. *Public Opinion Quarterly*, 81(2), 422–446. <https://doi.org/10.1093/poq/nfx018>.
- Kuru, O., Pasek, J., & Traugott, M.W. (2020). When polls disagree: How competitive results and methodological quality shape partisan perceptions of polls and electoral predictions. *International Journal of Public Opinion Research*, 32(3), 586–603. <https://doi.org/10.1093/ijpor/edz035>.
- Larsen, E.G., & Fazekas, Z. (2021). *Characteristics of opinion poll reporting: Creating the change narrative*. In *Reporting public opinion: How the media turns boring polls into biased news* (pp. 53–81). https://doi.org/10.1007/978-3-030-75350-4_4.
- Lelkes, Y., Sood, G., & Iyengar, S. (2017). The hostile audience: The effect of access to broadband internet on partisan affect. *American Journal of Political Science*, 61(1), 5–20. <https://doi.org/10.1111/ajps.12237>.
- Lorenz-Spreen, P., Oswald, L., Lewandowsky, S., & Herwig, R. (2023). A systematic review of worldwide causal and correlational evidence on digital media and democracy. *Nature Human Behaviour*. <https://doi.org/10.1038/s41562-022-01460-1>.
- Madson, G.J., & Hillygus, D.S. (2020). All the best polls agree with me: Bias in evaluations of political polling. *Political Behavior*, 42(4), 1055–1072. <https://doi.org/10.1007/s11109-019-09532-1>.
- Meyer, P. (1990). Presidential address: Polling as political science and polling as journalism. *Public Opinion Quarterly*, 54(3), 451–459. <https://doi.org/10.1093/poq/54.3.451>.
- Miller, M.M., & Hurd, R. (1982). Conformity to AAPOR standards in newspaper reporting of public opinion polls. *Public Opinion Quarterly*, 46(2), 243–249. <https://doi.org/10.1086/268716>.
- Morwitz, V.G., & Pluzinski, C. (1996). Do polls reflect opinions or do opinions reflect polls? The impact of political polling on voters' expectations, preferences, and behavior. *Journal of Consumer Research*, 23(1), 53–67. <https://doi.org/10.1086/209470>.
- Newport, F., Shapiro, R.Y., Ayres, W., Belden, N., Fishkin, J., Fung, A., & Warren, M. (2013). Polling and democracy: Executive summary of the aapor task

- force report on public opinion and leadership. *Public Opinion Quarterly*, 77(4), 853–860. <https://doi.org/10.1093/poq/nft039>.
- Portilla, I., Cano, E., & Barometro, C.I.S. (2016). The inclusion of methodological information in poll-based news: How do Spanish newspapers deal with political opinion polls? *Journal of Applied Journalism & Media Studies*, 5(3), 351–367. https://doi.org/10.1386/ajms.5.3.351_1.
- Redlawsk, D.P. (2002). Hot cognition or cool consideration? testing the effects of motivated reasoning on political decision making. *The Journal of Politics*, 64(4), 1021–1044. <https://doi.org/10.1111/1468-2508.00161>.
- Salwen, M.B. (1987). Credibility of newspaper opinion polls: Source, source intent and precision. *Journalism Quarterly*, 64(4), 813–819.
- Searles, K., Smith, G., & Sui, M. (2018). Partisan media, electoral predictions, and wishful thinking. *Public Opinion Quarterly*, 82(S1), 888–910. <https://doi.org/10.1093/poq/nfy006>.
- Seidenberg, A.B., Moser, R.P., & West, B.T. (2023). Preferred reporting items for complex sample survey analysis (PRICSSA). *Journal of Survey Statistics and Methodology*. <https://doi.org/10.1093/jssam/smz040>.
- Shapiro, R.Y. (2011). Public opinion and American democracy. *Public Opinion Quarterly*, 75(5), 982–1017. <https://doi.org/10.1093/poq/nfr053>.
- Stadtmüller, S., Silber, H., & Beuthner, C. (2022). What influences trust in survey results? Evidence from a vignette experiment. *International Journal of Public Opinion Research*. <https://doi.org/10.1093/ijpor/edac012>.
- Stefkovics, Á., Harrison, C., Eichhorst, A., & Skinnion, D. (2024). *Are We Becoming More Transparent?* Survey Reporting Trends in Top Journals of Social Sciences, *International Journal of Public Opinion Research*, 36(2) 1–6, <https://doi.org/10.1093/ijpor/edae013>.
- Strange, A.M., Enos, R.D., Hill, M., & Lakeman, A. (2019). Online volunteer laboratories for human subjects research. *PLOS ONE*, 14(8), e221676. <https://doi.org/10.1371/journal.pone.0221676>.
- Tsfati, Y. (2001). Why do people trust media pre-election polls? Evidence from the Israeli 1996 elections. *International Journal of Public Opinion Research*, 13(4), 433–441. <https://doi.org/10.1093/ijpor/13.4.433>.
- Verba, S. (1996). The citizen as respondent: Sample surveys and American democracy. *American Political Science Review*, 90(1), 1–7. <https://doi.org/10.2307/2082793>.