

RESEARCH ARTICLE  

Is a Picture Worth 280 Characters?: Contextually Realistic Graphics vs. Plain Text in Survey Experiments¹

Benjamin Norwood Harris  and Erik Lin-Greenberg 

Massachusetts Institute of Technology, Cambridge, MA, USA

Corresponding author: Benjamin Norwood Harris; Email: harrisb@mit.edu



Abstract

As survey experiments have become increasingly common in political science, some scholars have questioned whether inferences about the real world can be drawn from experiments involving hypothetical, text-based scenarios. In response to this criticism, some researchers recommended using realistic, context-heavy vignettes while others argue that abstract vignettes do not generate substantially different results. We contribute to this debate by evaluating whether incorporating contextually realistic graphics into survey experiment vignettes affects experimental outcomes. We field three original experiments that vary whether respondents are shown a realistic graphic or a plain text description during an international crisis. In our experiments, varying whether respondents are shown realistic graphics or plain text descriptions generally yields little difference in outcomes. Our findings have implications for survey methodology and experiments in political science – researchers may not need to invest the time to develop contextually realistic graphics when designing experiments.

Keywords: survey methodology; experimental research; mass media and political communication; public opinion; international relations

Survey experiments have become increasingly common in political science as tools to study events where real-world data are scarce or difficult to observe. Although survey experiments offer a time and cost-effective means of data generation, scholars have raised questions about their use and limitations (Hyde 2015; Barabas and Jerit 2010). As a result, a growing stream of scholarship explores whether and

¹The experiments conducted for this project were preregistered with the OSF Registry, available at https://osf.io/ktu3e/?view_only=e1e6ecb5c2cd4a13bc768ce5eacd70f8 and https://osf.io/kwqjn/?view_only=9b537db2fd7749748613daf3417c5604.

  This article has earned badges for transparent research practices: Open data and Open materials. For details see the [Data Availability Statement](#).

© The Author(s), 2024. Published by Cambridge University Press on behalf of American Political Science Association. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

how experiment design – particularly the content and structure of vignettes – affects experimental outcomes and external validity. Some research has focused on vignette format, assessing whether experiments featuring longer simulations of news articles yield different outcomes from those featuring shorter narratives (Kreps and Roblin 2019). Other studies have examined whether the degree of abstraction in experiment vignettes affects outcomes (Dafoe, Zhang, and Caughey 2018; Brutger et al. 2021). Still other projects analyze whether video formatting or the degree of humor affect outcomes (Young et al. 2018). These studies offer valuable insights but overlook important questions about the visual presentation of information. Does a vignette that is modeled to look like a real-world source, like a newspaper article or a politician’s Tweet, result in different levels of respondent attentiveness or produce different substantive outcomes than traditional plain text surveys?

As graphic design services and software become more readily accessible, political scientists – as well as communication and psychology scholars – have increasingly incorporated realistic depictions of newspaper articles, social media posts and misinformation, and media reports into surveys (Dill, Sagan, and Valentino 2022; Smetana, Vranka, and Rosendorf 2023; Green-Riley, Kruszewska-Eduardo, and Fu 2021; Bode and Vraga 2015). We call such depictions contextually realistic graphics. While these lifelike representations may more closely mirror real-world stimuli than the text-based scenario narratives found in most survey instruments, political scientists have yet to fully explore whether realistic graphic design affects survey outcomes. Instead, most existing research on graphics studies whether evocative imagery affects various political outcomes (Gadarian 2014; Green-Riley, Kruszewska-Eduardo, and Fu 2021). Yet, as scholars consider enhancing the realism of experimental designs, understanding whether and how treatment formatting affects outcomes becomes ever more important.

In this note, we aim to make a methodological contribution that helps political scientists navigate survey design. We field three original survey experiments that vary whether a vignette is presented as plain text or as a contextually realistic graphic – either of a presidential tweet, a leaked government document, or a newspaper article. We find that, on average, vignette format has little effect either on substantive outcomes or on retention of key details. To be sure, our experiments feature three specific contexts, limiting the generalizations we can draw from our findings. Consistency in findings across the three experiments, however, suggests that researchers need not devote the time or resources to develop contextually realistic graphics for their experiments.

Contextual realism and survey design

As survey experiments have become more common in political science, some scholars have questioned whether accurate inferences about the real world can be drawn from experiments involving text-based, hypothetical scenarios (Hyde 2015; Barabas and Jerit 2010; Egami and Hartman 2022). In response, some researchers have recommended that scholars carefully craft experimental vignettes to be as realistic as possible using language and imagery that places the vignette in the proper context – in a way that approximates real-world settings (Aguinis and Bradley 2014; Alekseev, Charness, and Gneezy 2017; McDonald 2020). Steiner et al. summarize the logic of contextual realism proponents, writing that “highly

contextualized vignettes increase the construct validity, that is, the degree to which the vignettes measure what we intend to measure” (Steiner et al. 2017, 54).

However, a wave of recent scholarship casts doubt on whether contextual realism actually affects survey outcomes (Sauer, Auspurg, and Hinz 2020; Shamon, Dülmer, and Giza 2019). Kreps and Roblin (2019) find that presenting vignette text as a mock news story or as plain text does not affect respondents’ opinions of support for conflict. More generally, Brutger et al. (2021) conclude that respondents typically provide similar responses regardless of whether a survey features abstract or highly realistic vignettes. In short, the debate over whether contextual realism affects substantive outcomes remains unresolved.

Several elements associated with contextual realistic treatments might moderate substantive outcomes. Past research has studied variations in vignette length and framing (Kreps and Roblin 2019), level of vignette detail (Brutger et al. 2021), and vignette specificity (Dafoe, Zhang, and Caughey 2018), there has been less attention on a vignette’s graphic design. On one hand, an experimental treatment’s graphical realism could affect respondent engagement in ways that shape substantive outcomes. For instance, a treatment that more realistically captures real-world stimuli might more fully engage participants, bolstering their buy-in and the amount of thought they dedicate to answering questions (McDermott 2002). In turn, this could produce stronger treatment effects compared to plain text vignettes. Alternately, more complex realistic graphical representations could be more cognitively taxing for respondents (Skulmowski and Rey 2020), leading to decreased attentiveness and weaker treatment effects.

On the other hand, a treatment’s graphical realism might have little effect on substantive outcomes. Because graphical representation only affects whether treatments are presented in a visually realistic manner versus plain text (i.e., respondents receive identical information), there may be no difference in how respondents receive the treatment. In other words, graphical realism may not moderate treatment effects. Given the findings of recent studies on vignette format and abstraction (Kreps and Roblin 2019; Brutger et al. 2021), we predict that there will be no difference in substantive outcomes between respondents who see a contextually realistic graphic and those who see plain text.

H₁: The graphical realism of survey experiment vignette presentation will have no significant effect on substantive outcomes.

While we predict graphical realism will have no effect on substantive outcomes, we theorize that it might affect the retention of details about the vignette. A long tradition of pedagogical and psychology research suggests greater realism and immersion in learning result in improved memorization and retention (Joseph and Dwyer 1984; Vasu and Howe 1989). Although these studies offer several explanations for why realism boosts information retention, there is widespread agreement that more realistic and immersive experiences engage participants more than less immersive, traditional approaches (Chittaro and Buttussi 2015; Di Natale et al. 2020; Hamilton et al. 2021). Moreover, the use of imagery rather than text alone is thought to engage multiple cognitive subsystems, potentially aiding with information recall (Schnotz 2001). In sum, respondents presented with contextually

realistic graphics are likely to spend more time examining and internalizing the information presented than respondents presented solely with text. Therefore, we predict that respondents who see a contextually realistic graphic are more likely to recall specific details of the vignette than those who see plain text.

H₂: Survey experiment respondents are likely to demonstrate greater retention of details from an experimental vignette when the vignette is realistically presented versus presented as plain text.

To be sure, some studies suggest that immersive or graphics-intensive experiences can lead to cognitive overload and stymie respondent performance (Skulmowski and Rey 2020). While this may be a risk with highly immersive and interactive experiences such as scenarios involving virtual reality, it is likely less of a risk given that our experiments do not involve the degree of immersion associated with virtual or augmented reality.

Method

To assess whether vignette format affects outcomes, we turn to three original survey experiments of hypothetical international crises that vary whether the survey instrument presents a contextually realistic graphic or plain text. We focus on crises because international relations scholars routinely use survey experiments to study public or elite preferences during international diplomatic or military confrontations (Tomz 2007; Tomz, Weeks, and Yarhi-Milo 2020), but we believe the experiments should yield insights applicable to a range of substantive topics.

Our first experiment presents respondents with a hypothetical, but plausible, crisis between the United States and Iran. All respondents are told:

Over the past several months, the Iranian government has provided funding, training, and weapons to militia groups that have launched several attacks on U.S. forces and partners throughout the Middle East. Earlier this week, Iranian-backed militias attacked two oil tankers in the Red Sea that were transporting fuel to the United States and fired rockets at the U.S. Embassy in Yemen. The attacks caused significant damage to the oil tankers and the embassy and killed eight people, including one American.

Respondents are then informed that “President Biden made his first statement about the situation” by issuing a tweet. We randomly assign respondents to one of two conditions in which respondents receive either a contextually realistic graphic of the tweet (Figure 1) or a plain text description that includes identical language. The simulated tweet replicates the layout and features of an actual tweet, including details such as the date, time, retweet information, and like and comment buttons.² In contrast, the plain text treatment includes only the president’s statement. We then ask a series of questions about perceived credibility, crisis realism, support for the president, and several attention checks.

²Full survey instrument in Appendix A, Section 2.1.



Figure 1. Contextually realistic tweet graphic treatment.

In the second experiment, we test the effect of graphical realism in a different context: a hypothetical leak of a U.S. Intelligence Community Assessment (ICA) regarding Russia and Syria. In this experiment, all respondents are told:

Last week, a document labeled as an official United States Intelligence Community Assessment began circulating on the internet. The U.S. government has neither confirmed nor denied the document’s authenticity, but the document highlights previously unknown Russian involvement in Syrian government chemical weapons attacks perpetrated against civilians in 2017.

As before, respondents are then randomly assigned to a contextually realistic graphic treatment (Figure 2) or a plain text treatment. The simulated graphic mimics key characteristics of an actual ICA such as the official seal and a serial number, while not including classification markings (e.g., “Top Secret”) as federal regulations govern their use.

For our third experiment, we replicate an experiment from Press et al.’s 2013 “Atomic Aversion” article (Press, Sagan, and Valentino 2013). In their original experiment, Press et al. presented respondents with vignettes styled like newspaper articles to test public attitudes toward nuclear weapons use. The experiment told respondents to consider a potential nuclear or conventional U.S. strike on an Al Qaeda nuclear lab and varied the relative effectiveness of the conventional or nuclear option. Press et al. (2013, 202) find that Americans “appear to weigh the consequences of using nuclear weapons in the narrow terms of immediate military effectiveness,” rejecting the idea that the public has internalized a nuclear taboo.

In our replication, we repeat Press et al.’s original variation and include additional experimental manipulations: whether respondents see a newspaper-styled article (as in the original experiment) or plain text.³ We present the nuclear advantage newspaper article in Figure 3. This design allows us to test the effect of graphical realism with difference-and-difference analysis by comparing the effect of

³We also vary the length of the treatment article, short versus long, for a $2 \times 2 \times 2$. For the complete survey instrument, see Appendix A, Section 2.3.

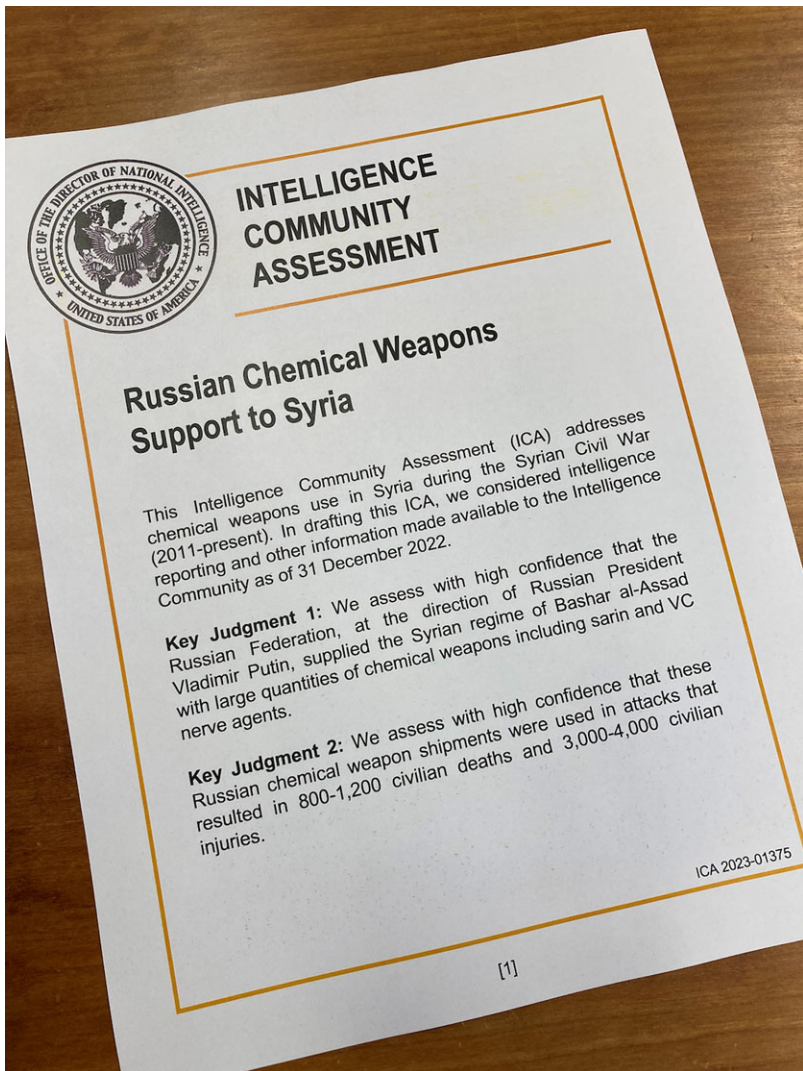


Figure 2. Contextually realistic ICA graphic treatment.

the newspaper treatments among respondents who were told nuclear weapons were more effective than conventional weapons with respondents who were told nuclear weapons were equally effective. In other words, this experiment allows us to explore whether contextually realistic graphics moderate substantive outcomes in a widely cited study.⁴ To assess factors that might moderate treatment, we ask respondents to directly assess cognitive load, enjoyment, and interest.

⁴In our preregistration for Experiment 3, we include a reframed version of H_1 related to this moderation effect: *Respondents in the contextually realistic graphic treatment will not differ in responses to substantive questions (including the treatment effect of nuclear advantage versus equal effectiveness) than respondents in the plain text treatment.*

Al Qaeda Building Atomic Bombs in Syria: Joint Chiefs Say U.S. Nuclear Option Offers Dramatically Increased Chances of Destroying Nuke Lab

Chiefs Conclude Nuclear Option Has 90% Chance of Success, Conventional Only 45%

The Associated Press

A report from General Charles Brown, Chairman of the Joint Chiefs of Staff, to the President concludes that nuclear weapons would be "dramatically more effective" than conventional strikes in destroying an al Qaeda nuclear weapons facility in Syria.

The report compares two American military options, a conventional strike using nearly one hundred conventionally-armed cruise missiles, and an attack using two small, nuclear-armed cruise missiles. The report estimates that that the conventional strike has a 45 percent chance of successfully destroying the atomic bomb lab while nuclear weapons increase the chances of success to approximately 90 percent.

The Joint Chiefs' assessment comes two weeks after Russian intelligence agents intercepted a shipment of centrifuges and low-enriched uranium which could be used to produce nuclear weapons.

The bomb-making equipment was being smuggled out of Russia to an Al Qaeda facility located near the remote town of As-Safih in northern Syria.

The suspects in the smuggling operation were employed at a Russian nuclear lab. The smugglers confirmed under questioning that other shipments of centrifuges and low-enriched uranium had already been delivered to the Al Qaeda base, where the centrifuges are being used to make fuel for a nuclear bomb. The smugglers stated that there will be enough bomb grade material produced for at least one weapon within two weeks. Syria has refused to allow international inspectors access to the facility.

"Nuclear weapons would be far more effective against this deeply buried target."

The Joint Chiefs' report to the President does not recommend a specific course of action. However, it concludes that "because the Al Qaeda facility is comprised of a series of deeply buried bunkers, nuclear weapons would be far more effective for destroying this

target."

The report was leaked to the Associated Press by a high-ranking administration official involved in planning the strike. According to the official, the centrifuges and nuclear materials are too large to be moved without detection. A US intelligence official stated that he has high confidence that al Qaeda is within two weeks of producing an operational bomb. After that, the official said, "all bets are off."

According to Dr. Robert Rust, a nuclear weapons expert at the Union of Concerned Scientists, an independent think-tank based in Washington, D.C., "If a bomb of this size exploded in New York City, it could easily kill 50,000 to

70,000 people."

The report states that the remote location of the Al Qaeda facility should limit Syrian civilian fatalities. Because many conventional weapons would be required to destroy the Al Qaeda base, the Joint Chiefs estimate that "the nuclear and conventional options would kill approximately the same number of Syrian civilians" - about 1,000, including immediate deaths and long term consequences of the conventional or nuclear strike. As both options will rely on cruise missiles launched from U.S. naval vessels, the report concludes that "no U.S. military personnel are at risk in either operation."

TARGET: AL QAEDA NUCLEAR WEAPONS LAB		
	U.S. NUCLEAR STRIKE	U.S. CONVENTIONAL STRIKE
PROBABILITY OF SUCCESS	90%	45%
ESTIMATED SYRIAN CIVILIAN DEATHS	1,000	1,000

IF U.S. STRIKE FAILS: 50,000 - 70,000 U.S. CIVILIAN FATALITIES

Chart from Joint Chiefs' report describing nuclear and conventional options for strike on Al Qaeda nuclear lab.

Figure 3. Long, nuclear advantage, contextually realistic treatment.

We fielded the first two experiments on a U.S. public sample of 1,511 respondents recruited using the online sampling service Lucid Theorem in June 2023. Lucid relies on quota sampling to recruit samples that align with U.S. Census demographics. Lucid samples, however, are not nationally representative across all dimensions. For instance, our sample underrepresents Hispanic Americans while overrepresenting college-educated Americans.⁵ Still, Lucid samples are more representative than other online convenience samples, like Amazon's Mechanical Turk (Coppock and McClellan 2019). We fielded the third experiment on a U.S. public sample of 1,793 respondents recruited on the Prolific platform in November 2023. Recent studies suggest Prolific samples offer higher data quality (Peer et al. 2022; Douglas, Ewell, and Brauer 2023). Still, Prolific samples are not perfectly representative. Our Prolific sample, for example, overrepresents Black Americans and underrepresents top wage earners.

⁵Full demographic data in Appendix B, Section 4.

Findings

Experiment 1: tweeting threats

H₁: substantive outcomes

In our Twitter experiment, we measure four substantive outcomes: respondent's perceptions of (1) crisis realism (i.e., whether a crisis involving threats made on Twitter could happen in the real world), (2) the perceived likelihood that Iranian officials will believe the threat, (3) support for the president's handling of the threat and (4) perceived credibility of the president's threat.⁶ We use a 5-point Likert scale to measure respondent perceptions. For example, to measure credibility, we ask respondents "In your opinion, how likely or unlikely is it that the president will follow through on his threat?" on a five-point scale between "very unlikely" (1) and "very likely" (5). We run ordinary least squares regressions without (model 1) and with (model 2) demographic covariates. The average treatment effects are presented in Figure 4.⁷

We find general support for H_1 .⁸ Our results reveal no causal effects for three of our variables – threat credibility, the likelihood Iranian officials believe the threat, and crisis realism. However, we do have an unexpected result: respondents in the Tweet graphic treatment are more likely to express support for the president's handling of the crisis than respondents in the plain text treatment.

Substantively, the effect size is small but noticeable given the relatively weak treatment manipulation. However, given that our analysis for this experiment includes regressions for four substantive variables and four attention check/timing variables, it would not be surprising to find at least one statistically significant result simply through chance.⁹ To assess whether this unexpected finding was replicable, we repeated the experiment on a different 1,206 respondent Lucid sample. While the coefficient for presidential support remains positive, we find no statistically significant causal effect ($p = 0.26$), suggesting the initial finding occurred by chance.¹⁰

H₂: information retention

In addition to substantive outcomes, we examine whether contextually realistic graphics affect respondents' attentiveness and information retention. To do this, our survey instrument tracks the time respondents spend reading the treatment and includes three attention check (AC) questions that ask respondents to recall details of the crisis scenario: the sea in which the oil tankers were attacked ("sea check"), the

⁶Survey instrument in Appendix A, Section 2.1.

⁷Regression results in Appendix B, Section 1.1.

⁸In addition to standard regressions, we run equivalence tests on all substantive variables for all experiments, which support the regression findings. See Appendix B, Section 3 and Appendix D, Section 6 for equivalence test results.

⁹If we employ a simple Bonferroni Correction—which decreases α , the acceptable false error rate, in proportion to the number of tests being run—our 95% confidence intervals would need to become 99.38% confidence intervals in order for the family-wise error rate to remain at the accepted 0.05 threshold. None of our p-values are significant with the Bonferroni Correction. See the full Bonferroni Correction results in Appendix B, Section 1.3.

¹⁰Regression results in Appendix C, Section 1.

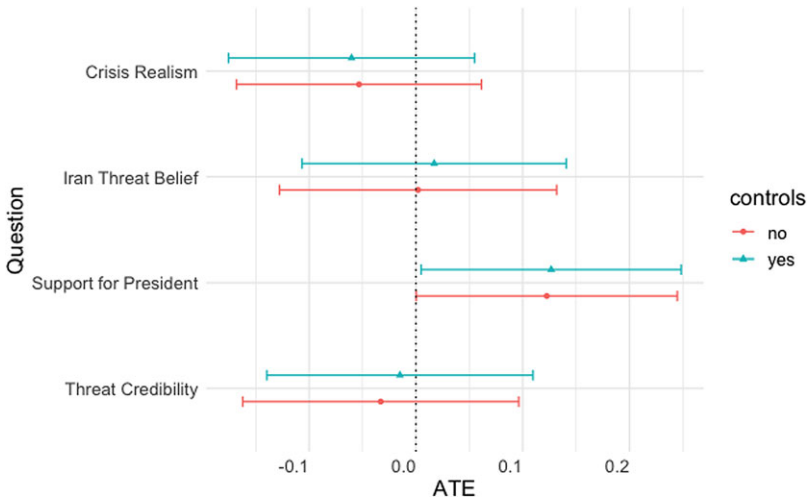


Figure 4. Average treatment effect of tweet graphic (substantive questions). Error bars represent 95% confidence intervals.

country supporting the militias (“support check”), and the target of the militia attacks (“target check”). Surprisingly, we find no support for H_2 : there are no statistically significant differences between the graphically realistic tweet and the plain text treatment for our attention check questions and for the time spent reading the treatment.¹¹ The average treatment effect for all AC questions is shown in Figure 5. Put differently, more realistic treatments do not appear to enhance information retention.

Experiment 2: leaked intelligence

H_1 : Substantive Outcomes

We measure four substantive outcomes in our leaked intelligence experiment: (1) crisis realism (i.e., whether a similar leak could happen in the real world), (2) document authenticity (i.e., whether the document is an actual intelligence document), (3) international perceptions (i.e., whether the international community will believe Russia exported chemical weapons), and (4) credibility of the document contents (i.e., whether Russia supplied chemical weapons).¹² As before, we use a five-point Likert scale and run regression models. Figure 6 displays the average treatment effects.¹³

In line with H_1 , we find no significant causal effects for three of our variables: crisis realism, international perception, and credibility. However, we surprisingly find a strong *negative* causal effect of the treatment on perceptions of the document’s authenticity. Respondents who received the graphic treatment were more likely to doubt the authenticity of the leaked report ($p < 0.05$). To assess this

¹¹Analysis in Appendix B, Section 1.2.

¹²Survey instrument in Appendix A, Section 2.1.

¹³Regression results in Appendix B, Section 2.1.

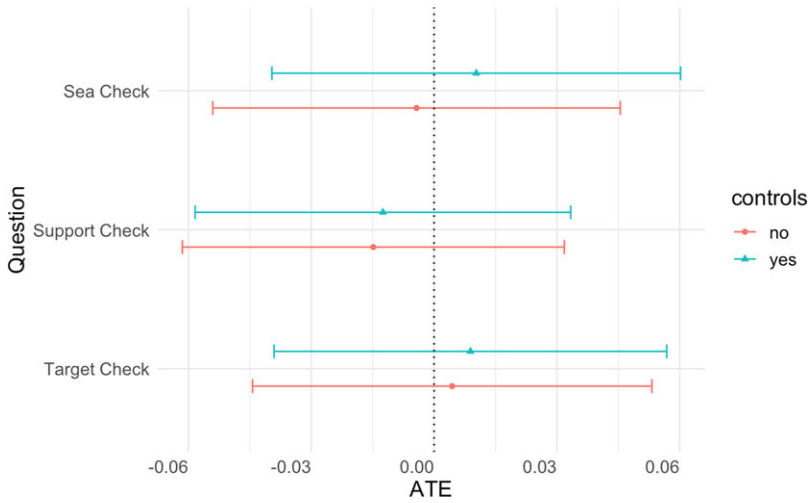


Figure 5. Average treatment effect of tweet graphic (AC questions). Error bars represent 95% confidence intervals.

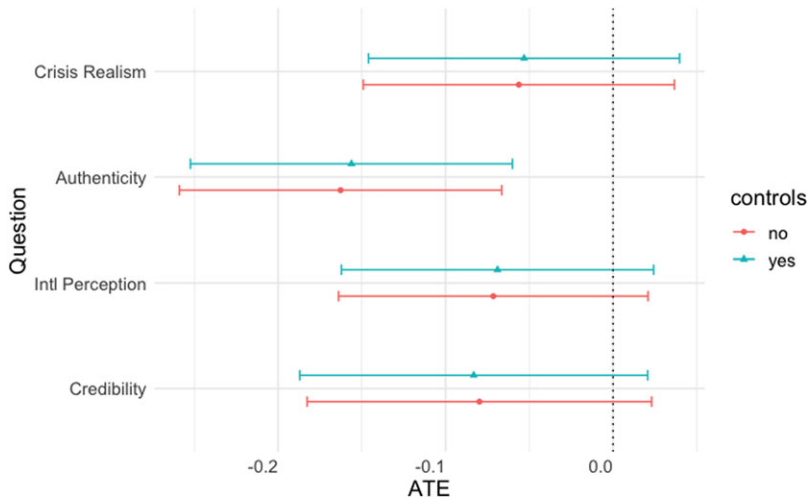


Figure 6. Average treatment effect of intelligence graphic (substantive questions). Error bars represent 95% confidence intervals.

finding, we reran a nearly identical experiment on a different Lucid sample.¹⁴ As in our original experiment, we find a strong negative relationship between the graphical treatment and perceived authenticity ($p < 0.01$).¹⁵

We suspect that this negative relationship might result from our decision to omit classification markings from the graphical depiction of the leaked document,

¹⁴The experiment replaced “Russian” with “North Korean” support to Syria. For full survey instrument, see Appendix A, Section 2.2.

¹⁵Regression results in Appendix C, Section 1.

potentially hindering our efforts to produce an authentic-looking document. The formatting of intelligence documents may have been particularly salient to respondents since a large real-world intelligence leak occurred just prior to our survey fielding.¹⁶ Regardless of the specific markings, participants might also simply be expressing skepticism that an internet survey would show them a picture of an actual leaked document. Importantly, this lack of perceived authenticity did not affect our other substantive dependent variables.

H₂: information retention

As in the Twitter experiment, we asked respondents three attention check questions and measured the time respondents spent reading the treatment.¹⁷ Specifically, we ask: what entity wrote the intelligence report (“Intel Check”), what country supplied the weapons (“Supplied Check”), and where the chemical weapons were used (“Used Check”).

Figure 7 shows the average treatment effect of the intelligence graphic on our three attention check questions. We find little support for H_2 for most of our results; however, we do find a positive effect of the graphical treatment on the attention check about the state supplying chemical weapons.¹⁸ We find that respondents in the graphical realism treatment are more likely to correctly identify Russia as the supplier of chemical weapons. One difference between our leaked document graphic treatment and the plain text control is that “Russia” appears in large, bolded letters in the leaked document graphic. To test whether the font size drives effects, we repeated the experiment on a different sample, but changing the supplying country from Russia to North Korea. We were unable to replicate a statistically significant result.¹⁹ We also found no statistically significant difference in time spent reading the plain text and graphical realism treatments.

Experiment 3: atomic aversion

H₁: substantive outcomes

We measure three substantive nuclear outcomes in the atomic aversion experiment, copied exactly from Press et al.’s original study: (1) preferred choice between the two options (a four-point scale ranging from strong conventional preference to strong nuclear preference); (2) approval for using conventional weapons (a six-point scale); and (3) approval for using nuclear weapons (a six-point scale).²⁰ We measure Press et al.’s original outcome variables in order to examine whether graphical realism moderates the findings from the original experiment.

We replicate Press et al.’s original findings. Greater military effectiveness for the nuclear strike option leads to sharply higher approval for nuclear use, declined

¹⁶See for example, (Nagourney 2023).

¹⁷Question wording in Appendix A, Section 2.1.

¹⁸As before, because we test a significant number of outcomes, it would be unsurprising to find a significant result through chance. However, unlike before, the p-value for the estimate of the graphical treatment on the supplied weapons attention check still meets the Bonferroni Correction’s stricter threshold for statistical significance. See Appendix B, Section 2.3.

¹⁹In the replication, $p = 0.12$. Regression results in Appendix C, Section 1.

²⁰Survey instrument in Appendix A, Section 2.2.

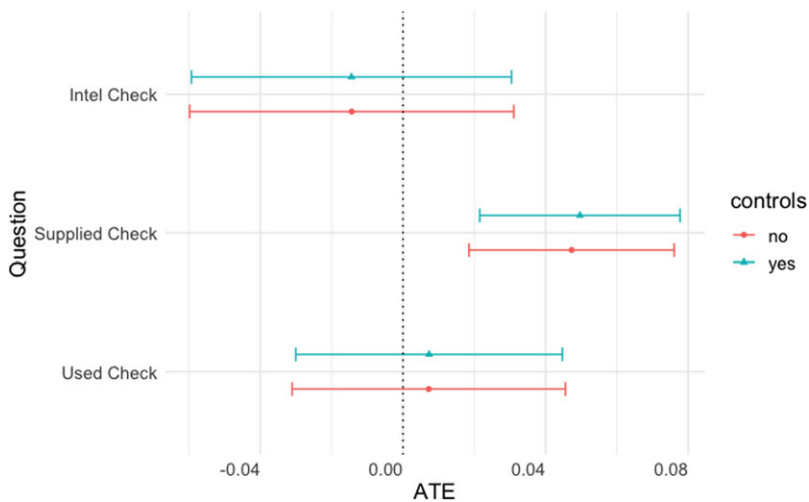


Figure 7. Average treatment effect of intelligence graphic (AC questions). Error bars represent 95% confidence intervals.

support for the conventional strike, and stronger preference for the nuclear option.²¹ Given our focus on graphical realism, we do not display these nuclear results graphically; we plot only the effect of graphical realism.

As before, we find no effect on substantive outcomes for the realistic graphic treatment. In Figure 8, we show the average treatment effect for the realistic graphic treatment across all three substantive outcomes.

In addition to the standard regression, we run a difference-in-difference (DiD) analysis. In our DiD analysis, we consider whether graphical realism moderates the effects of the nuclear advantage variable on the three outcome variables outlined above. This allows us to test whether Press et al.'s original experiment would have found different results if the researchers had used plain text treatments rather than mock newspaper articles.

We divide respondents into two groups, those in the graphical realism treatments and those in the newspaper treatments. This is the first difference. We then run regressions to determine the difference between treatments where nuclear weapons have an advantage and treatments where the nuclear and conventional options have equal effectiveness (the second difference). The DiD analysis finds a null result. Figure 9 displays the difference between the estimate of the nuclear advantage coefficient between the realistic graphic treatment groups and the plain text treatment groups for all three nuclear outcomes. A positive value means that respondents in the newspaper treatment groups were more influenced by the nuclear advantage variable. A negative value would mean the plain text treatment groups were more influenced by the nuclear advantage variable. The differences are small and not statistically significant.²²

²¹Regression results in Appendix D, Section 1.2.

²²We run a power analysis to determine the minimum detectable effect size for both the standard regressions and the difference-in-difference model. See Appendix D, Section 7.

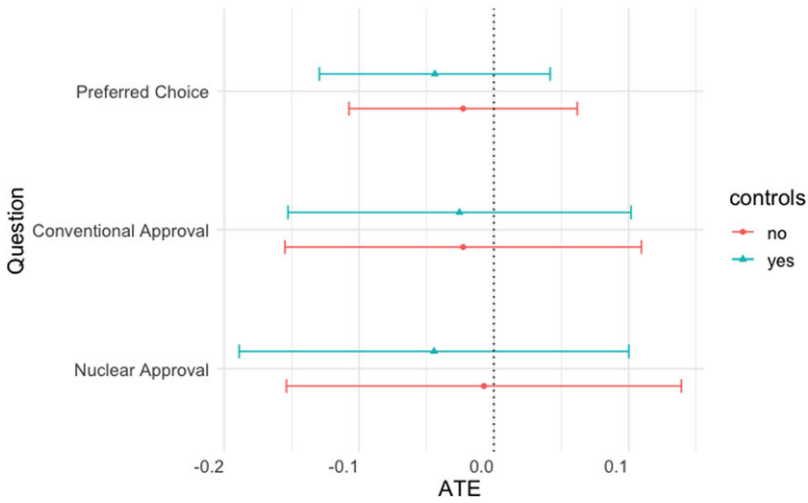


Figure 8. Average treatment effect of newspaper graphic (substantive questions). Error bars represent 95% confidence intervals.

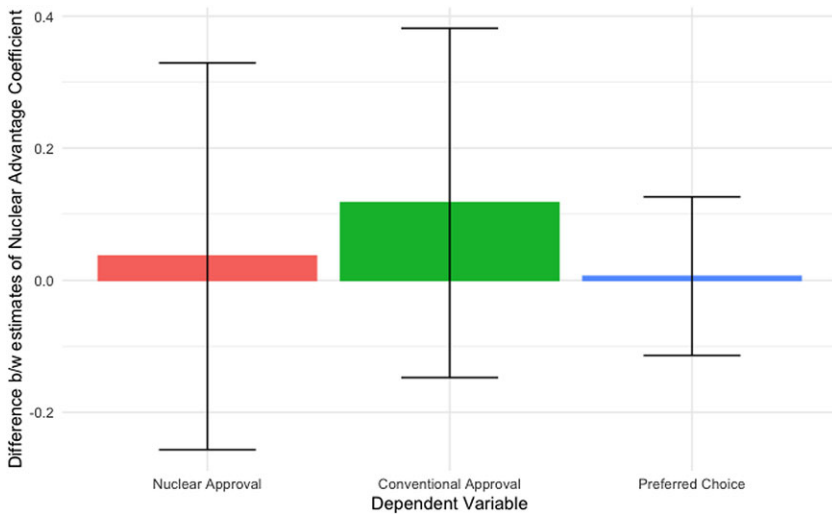


Figure 9. DiD – newspaper vs. plain text effect on nuclear advantage.

H₂: information retention

We include three attention check questions and, as before, measure the time respondents spent reading the treatments. We ask respondents to correctly identify the affiliation of the scientist quoted in the vignette (“scientist check”); the affiliation of the fictional report’s author (“author check”); and the country in which the terrorist lab is located (“country check”). As before, we find little support for *H₂* (Figure 10). We do, however, find that respondents shown the graphical realism

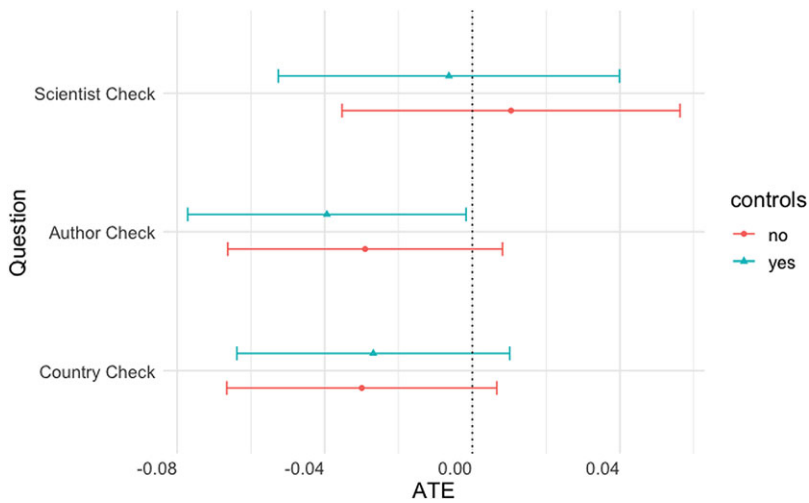


Figure 10. Average treatment effect of newspaper graphic (AC questions).

treatment spend far less time, about thirty-four seconds less time, reading the treatment than respondents in the plain text treatment.²³

To gain insight into whether graphical realism affects respondent engagement, we asked about their experience taking the survey experiment. We include three measures: cognitive load, enjoyment, and interest. To study cognitive load, we ask, “Compared to an average news article, how difficult or easy did you find the fictitious news article to read?” To collect data on enjoyment, we ask “How much did you like or dislike reading the fictitious news article?” Finally, to explore interest, we ask, “If you heard about this incident in real life, how likely or unlikely would you be to seek out more information about the proposed strike?” We find null results for graphical realism across all experiential variables, shown in Figure 11.²⁴

Implications and pathways for future research

Randomized control trials are the gold standard for generating internally valid results, leading to increased survey experiment use among political scientists. The dramatic reduction in cost and expanded reach of survey platforms is a boon for the profession, but one that calls for increased study of survey experiments themselves. Our findings contribute to that line of study while also suggesting avenues for future research to help develop best practices.

In this research note, we study whether contextually realistic graphics in survey experiment treatments yield results that differ from plain text vignettes. Our results

²³See Appendix D, Section 2 for full results.

²⁴While not shown here, we do find results for Press’ et al.’s nuclear advantage variable on the experiential variables. Respondents in the nuclear advantage treatment are much less likely to find the crisis scenario realistic and much more likely to report wanting to seek out additional information on the article, as might be expected. This result gives us greater confidence in the null results we find for graphical realism and is detailed in Appendix D, Section 3.

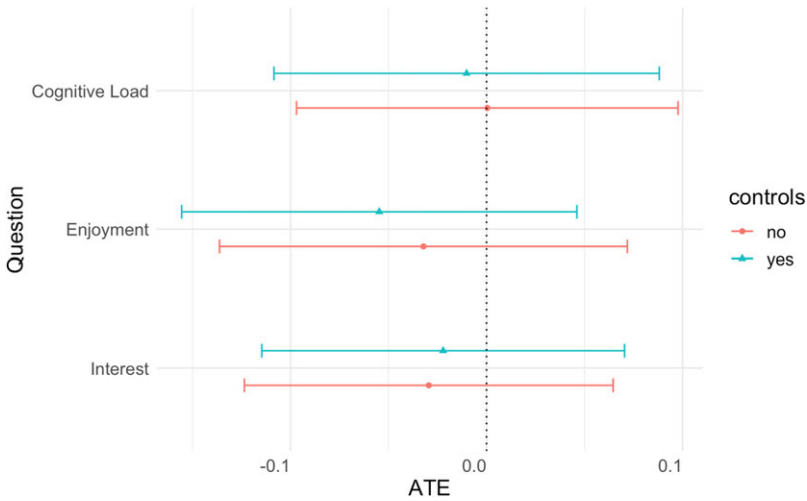


Figure 11. Average treatment effect of newspaper graphic (experiential questions).

show that varying between contextually realistic graphics and plain text has little effect on substantive outcomes or on respondents' information retention. Few of our dependent variables showed a statistically significant average treatment effect, and those that did were generally not replicable in follow-up experiments. Given the large number of tests we run, it is also unsurprising that at least a few of them would show significant results given the multiple comparison problem.

Our findings suggest political scientists gain little from using contextually realistic graphics in lieu of plain text treatments. If, however, political scientists decide to incorporate graphical realism, they should conduct thorough pretests to ensure their graphics are sufficiently realistic. Indeed, respondents consistently rated our graphically realistic intelligence document as less authentic than a plain text description. While we are unable to conclude with certainty why our leaked report was seen as less authentic, it seems plausible that it did not sufficiently resemble leaked documents that respondents had previously seen.

Future work could examine whether elites respond differently to treatments with contextually realistic graphics differently than members of the public, contributing to the burgeoning body of work on elite samples (Dietrich, Hardt, and Swedlund 2021; Chu and Recchia 2022; Kertzer and Renshon 2022). Future work could also explore whether using contextually realistic graphics affects studies outside of international relations. For example, scholars of comparative or American politics might assess whether presenting information as plain text or as a mock newspaper article or social media post influences substantive outcomes. In sum, this research note does not represent the final word, but instead suggests important areas for further inquiry.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/XPS.2024.10>

Data availability. The data, code, and any additional materials required to replicate all analyses in this article are available at the Journal of Experimental Political Science Dataverse within the Harvard Dataverse Network, at: <https://doi.org/10.7910/DVN/KTAJ9N>.

Funding. The authors declare none.

Competing interests. The authors declare none.

Ethics statement. The experimental research in this paper was approved by the MIT IRB (The Committee on the Use of Humans as Experimental Subjects), Protocol Number E-4634. The authors affirm that the research adheres to APSA's Principles and Guidance for Human Subjects Research. Appendix A details subject recruitment in more detail.

References

- Aguinis, Herman, and Kyle J. Bradley.** 2014. "Best Practice Recommendations for Designing and Implementing Experimental Vignette Methodology Studies." *Organizational Research Methods* 17: 351–371.
- Alekseev, Aleksandr, Gary Charness, and Uri Gneezy.** 2017. "Experimental Methods: When and Why Contextual Instructions Are Important." *Journal of Economic Behavior & Organization* 134: 48–59.
- Barabas, Jason, and Jennifer Jerit.** 2010. "Are Survey Experiments Externally Valid?" *American Political Science Review* 104: 226–242.
- Bode, Leticia, and Emily K. Vraga.** 2015. "In Related News, That Was Wrong: The Correction of Misinformation Through Related Stories Functionality in Social Media." *Journal of Communication* 65: 619–638.
- Brutger, Ryan, Joshua Kertzer, Jonathon Renshon, Dustin Tingley, and Chagai Weiss.** 2021. "Abstraction and Detail in Experimental Design." *American Journal of Political Science* 67: 979–995.
- Chittaro, Luca, and Fabio Buttussi.** 2015. "Assessing Knowledge Retention of an Immersive Serious Game vs. a Traditional Education Method in Aviation Safety." *IEEE Transactions on Visualization and Computer Graphics* 21: 529–538.
- Chu, Jonathan, and Stefano Recchia.** 2022. "Does Public Opinion Affect the Preferences of Foreign Policy Leaders? Experimental Evidence from the UK Parliament." *Journal of Politics* 84: 1874–1877.
- Coppock, Alexander, and Oliver A. McClellan.** 2019. "Validating the Demographic, Political, Psychological, and Experimental Results Obtained from a New Source of Online Survey Respondents." *Research & Politics* 6: 205316801882217.
- Dafoe, Allan, Baobao Zhang, and Devin Caughey.** 2018. "Information Equivalence in Survey Experiments." *Political Analysis* 26: 399–416.
- Di Natale, Anna Flavia, Claudia Repetto, Giuseppe Riva, and Daniela Villani.** 2020. "Immersive Virtual Reality in K-12 and Higher Education: A 10-Year Systematic Review of Empirical Research." *British Journal of Educational Technology* 51: 2006–2033.
- Dietrich, Simone, Heidi Hardt, and Haley J. Swedlund.** 2021. "How to Make Elite Experiments Work in International Relations." *European Journal of International Relations* 27: 596–621.
- Dill, Janina, Scott D. Sagan, and Benjamin A. Valentino.** 2022. "Kettles of Hawks: Public Opinion on the Nuclear Taboo and Noncombatant Immunity in the United States, United Kingdom, France, and Israel." *Security Studies* 31: 1–31.
- Douglas, Benjamin D., Patrick J. Ewell, and Markus Brauer.** 2023. "Data Quality in Online Human-Subjects Research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA." *PLOS ONE* 18: e0279720.
- Egami, Naoki, and Erin Hartman.** 2022. "Elements of External Validity: Framework, Design, and Analysis." *American Political Science Review* 117: 1070–1088.
- Gadarian, Shana Kushner.** 2014. "Scary Pictures: How Terrorism Imagery Affects Voter Evaluations." *Political Communication* 31: 282–302.
- Green-Riley, Naima, Dominika Kruszewska-Eduardo, and Ze Fu.** 2021. "Teargas and Selfie Cams: Foreign Protests and Media in the Digital Age." *Journal of Experimental Political Science* 9: 203–215.
- Hamilton, D., J. McKechnie, E. Edgerton, and C. Wilson.** 2021. "Immersive Virtual Reality as a Pedagogical Tool in Education: A Systematic Literature Review of Quantitative Learning Outcomes and Experimental Design." *Journal of Computers in Education* 8: 1–32.
- Harris, Benjamin.** 2024. "Replication Data for: Is a Picture Worth 280 Characters? Contextually Realistic Graphics vs. Plain Text in Survey Experiments", Harvard Dataverse. <https://doi.org/10.7910/DVN/KTAJ9N>.

- Hyde, Susan D.** 2015. "Experiments in International Relations: Lab, Survey, and Field." *Annual Review of Political Science* 18: 403–424.
- Joseph, John H., and Francis M. Dwyer.** 1984. "The Effects of Prior Knowledge, Presentation Mode, and Visual Realism on Student Achievement." *The Journal of Experimental Education* 52: 110–121.
- Kertzer, Joshua D., and Jonathan Renshon.** 2022. "Experiments and Surveys on Political Elites." *Annual Review of Political Science* 25: 529–550.
- Kreps, Sarah, and Stephen Roblin.** 2019. "Treatment Format and External Validity in International Relations Experiments." *International Interactions* 45: 576–594.
- McDermott, Rose.** 2002. "Experimental Methodology in Political Science." *Political Analysis* 10: 325–342.
- McDonald, Jared.** 2020. "Avoiding the Hypothetical: Why 'Mirror Experiments' Are an Essential Part of Survey Research." *International Journal of Public Opinion Research* 32: 266–283.
- Nagourney, Eric.** 2023. "A Quick Guide to What the Leaked U.S. Intelligence Documents Say." *The New York Times*, April 27, 2023. <https://www.nytimes.com/article/leaked-documents-ukraine-russia-war.html>.
- Peer, Eyal, David Rothschild, Andrew Gordon, Zak Evernden, and Ekaterina Damer.** 2022. "Data Quality of Platforms and Panels for Online Behavioral Research." *Behavior Research Methods* 54: 1643–1662.
- Press, Daryl G., Scott D. Sagan, and Benjamin A. Valentino.** 2013. "Atomic Aversion: Experimental Evidence on Taboos, Traditions, and the Non-Use of Nuclear Weapons." *American Political Science Review* 107: 188–206.
- Sauer, Carsten, Katrin Auspurg, and Thomas Hinz.** 2020. "Designing Multi-Factorial Survey Experiments: Effects of Presentation Style (Text or Table), Answering Scales, and Vignette Order." *Methods, Data, Analyses* 14: 20.
- Schnotz, Wolfgang.** 2001. "Towards an Integrated View of Learning From Text and Visual Displays." *Educational Psychology Review* 14: 101–120.
- Shamon, Hawal, Hermann Dülmer, and Adam Giza.** 2019. "The Factorial Survey: The Impact of the Presentation Format of Vignettes on Answer Behavior and Processing Time." *Sociological Methods & Research* 51: 396–438.
- Skulmowski, Alexander, and Günter Daniel Rey.** 2020. "Subjective Cognitive Load Surveys Lead to Divergent Results for Interactive Learning Media." *Human Behavior and Emerging Technologies* 2: 149–157.
- Smetana, Michal, Marek Vranka, and Ondrej Rosendorf.** 2023. "The 'Commitment Trap' Revisited: Experimental Evidence on Ambiguous Nuclear Threats." *Journal of Experimental Political Science* 11: 1–14.
- Steiner, Peter M., Christiane Atzmüller, and Dan Su.** 2017. "Designing Valid and Reliable Vignette Experiments for Survey Research: A Case Study on the Fair Gender Income Gap." *Journal of Methods and Measurement in the Social Sciences* 7:52–94.
- Tomz, Michael.** 2007. "Domestic Audience Costs in International Relations: An Experimental Approach." *International Organization* 61: 821–840.
- Tomz, Michael, Jessica Weeks, and Keren Yarhi-Milo.** 2020. "Public Opinion and Decisions About Military Force in Democracies." *International Organization* 74: 119–143.
- Vasu, Ellen, and Ann Howe.** 1989. "The Effect of Visual and Verbal Modes of Presentation on Children's Retention of Images and Words." *Journal of Research in Science Teaching* 26: 401–407.
- Young, Dannagal G., Kathleen Hall Jamieson, Shannon Poulsen, and Abigail Goldring.** 2018. "Fact-Checking Effectiveness as a Function of Format and Tone: Evaluating FactCheck. Org and FlackCheck. Org." *Journalism & Mass Communication Quarterly* 95: 49–75.

Cite this article: Harris BN and Lin-Greenberg E (2024). Is a Picture Worth 280 Characters?: Contextually Realistic Graphics vs. Plain Text in Survey Experiments¹. *Journal of Experimental Political Science*. <https://doi.org/10.1017/XPS.2024.10>