

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

Benjamin N. Harris

Erik Lin-Greenberg

Department of Political Science
Massachusetts Institute of Technology

Abstract: As survey experiments have become increasingly common in political science, some scholars have questioned whether these experiments generate externally valid results. To answer 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 two 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 more broadly.

Keywords: survey methodology, experiment research, mass media and political communication, public opinion, international relations

¹ This project was preregistered with the OSF Registry.

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 how experiment design—particularly the content and structure of vignettes—affects experimental outcomes and external validity. Some studies have 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). 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 result in different levels of respondent attentiveness or produce different substantive outcomes than traditional plain text surveys?

As graphic design services and generative artificial intelligence software become more readily accessible, political scientists may increasingly incorporate realistic depictions of media or other documents into surveys. 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 explore whether realistic graphic design affects survey outcomes. Instead, existing research on graphics focuses on the effect of evocative imagery on 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. To make this intervention, we field two original survey experiments that vary whether a vignette is presented as plain text or as a contextually realistic graphic—either of a presidential tweet or of a leaked government document. 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 two specific contexts, limiting the generalizations we can draw from our findings. Consistency in findings across the two 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 voiced concern about whether survey experiments generate useful, externally valid findings (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 (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 formatting a vignette 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.

While past research has studied variations in vignette length (Kreps and Roblin 2019), level of vignette detail (Brutger et al. 2021), and vignette specificity (Dafoe, Zhang, and Caughey 2018), scholars have not explicitly focused 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 the identical information), there may be no difference in how respondents receive the treatment. 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 versus 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 contextual 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 research in the fields of education and psychology connects greater realism and immersion in learning

materials to improved memorization and retention (Joseph and Dwyer 1984; Vasu and Howe 1989). Although scholars of pedagogy and psychology have 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-intense 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 intense scenarios involving virtual reality, it is likely less of a risk given that our experiments include do not involve the degree of immersion associated with virtual or augmented reality.

METHOD

To assess whether vignette format affects outcomes, we turn to two 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.

Figure 1: Contextually Realistic Tweet Graphic Treatment



² Full survey instrument in Appendix A.

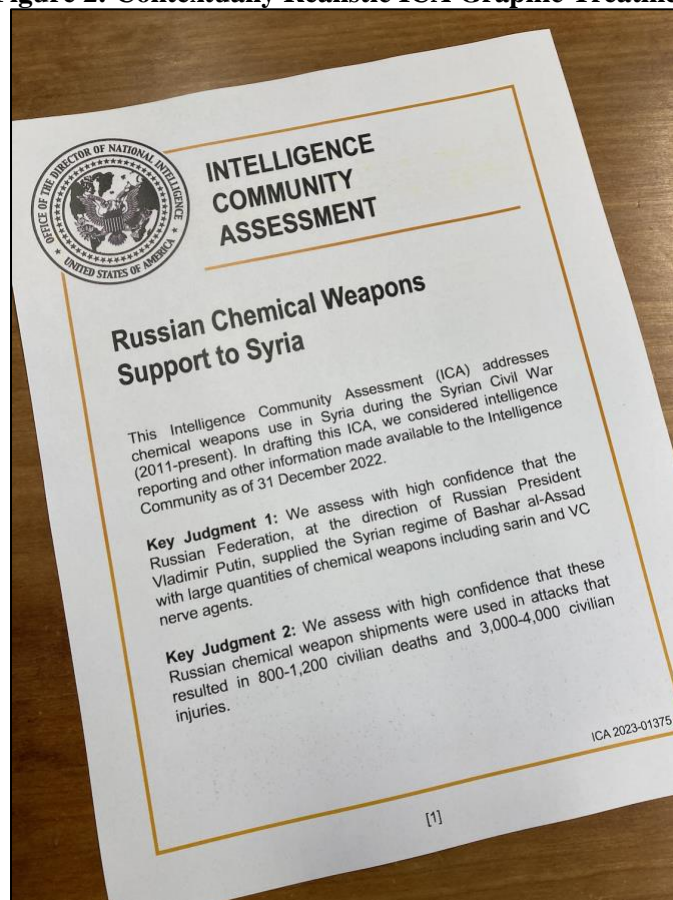
In the second experiment, we test the effect of graphical realism in a different context: a hypothetical leak of a U.S. Intelligence Community Assessment (ICA) regarding Russia and Syria.

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. The document’s first page is pictured below:

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 law governs their use.

Figure 2: Contextually Realistic ICA Graphic Treatment



We fielded the 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.³ Still, Lucid samples are more representative than other online convenience samples, like Amazon’s MechanicalTurk (Coppock and McClellan 2019).

FINDINGS

Experiment 1: Tweeting Threats

H1: Substantive Outcomes

In our Twitter experiment, we measure four substantive outcomes: respondent’s perceptions of 1) threat credibility, 2) support for the president’s handling of the threat, 3) the likelihood that Iranian officials believe the threat; and 4) crisis realism (i.e., whether a crisis involving threats made on Twitter could happen in the real world).⁴ We utilize 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?” and give them the option to choose between “very unlikely” (1) and “very likely” (5). The mean response of each of these substantive questions by treatment are presented in Table 1. We also run ordinary least squares regressions without (model 1) and with (model 2) demographic covariates. The average treatment effect for both model 1 and model 2 is presented in Figure 3.⁵

³ Full demographic data in Appendix B.

⁴ Question wording in Appendix A.

⁵ Regression results in Appendix B.

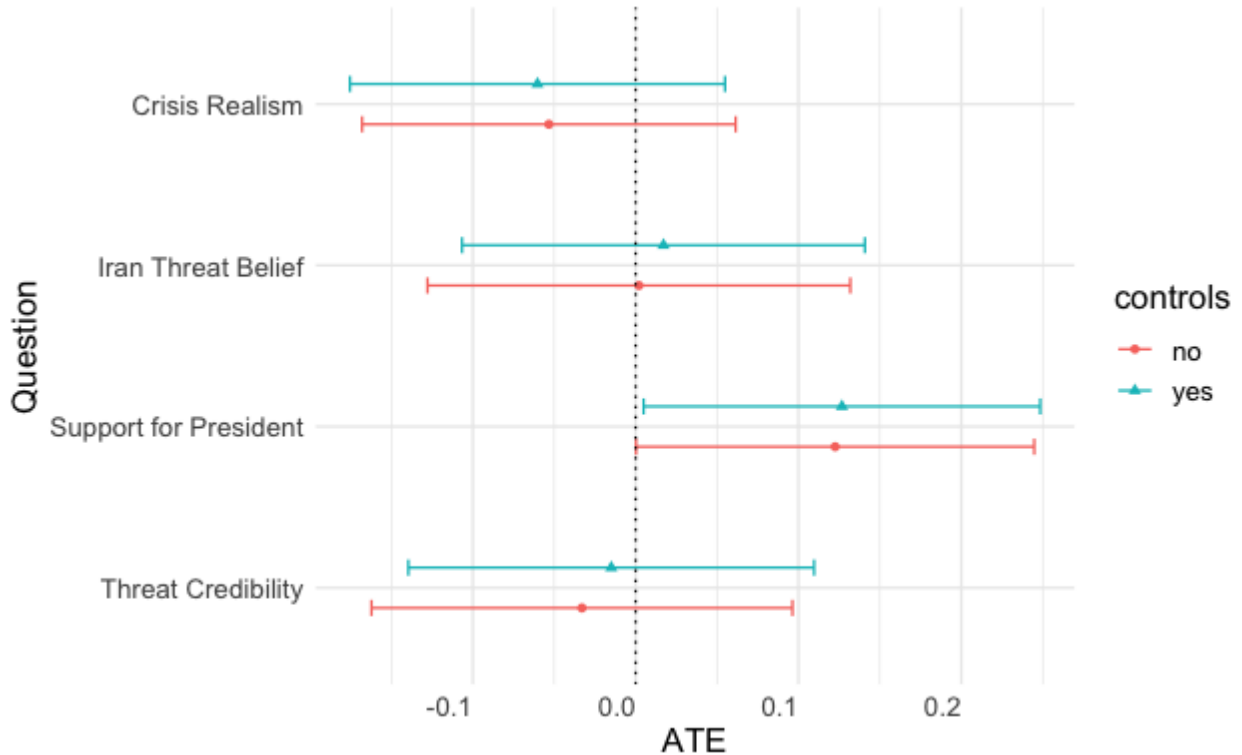
Table 1: Mean Response to Substantive Questions by Treatment

	Threat Credibility	Support for President	Iran Threat Belief	Crisis Realism
Tweet Graphic	3.08 (0.05) <i>n</i> = 760	3.33 (0.04) <i>n</i> = 760	2.83 (0.05) <i>n</i> = 760	3.67 (0.04) <i>n</i> = 760
Plain Text	3.12 (0.05) <i>n</i> = 751	3.21 (0.05) <i>n</i> = 751	2.83 (0.05) <i>n</i> = 751	3.72 (0.04) <i>n</i> = 751

Standard errors in parentheses.

We find general support for H_1 . Our results reveal no significant difference for three of our variables—threat credibility, the likelihood Iran believes 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.

Figure 3: Average Treatment Effect of Tweet Graphic (Substantive Questions)



Error bars represent 95% confidence intervals.

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 in 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. To examine this result in greater detail, we repeated the experiment on a different Lucid sample of 1,206 respondents. While the coefficient for presidential support remains positive, the effect size is smaller and not statistically significant ($p = 0.26$).⁶

H2: Information Retention

In addition to substantive outcomes, we examine whether contextually realistic graphics affect the attentiveness and information retention of respondents. To do this, our survey instrument includes three attention check (AC) questions that ask respondents to recall details of the crisis scenario: the target of the militia attacks (“target check”, the country supporting the militias (“support check”), and the sea in which the oil tanker was attacked (“sea check”). We also measure the time respondents spent reading the scenario vignette. Table 2 shows the correct response proportion to the attention check questions by treatment and the average time spent reading the treatment.

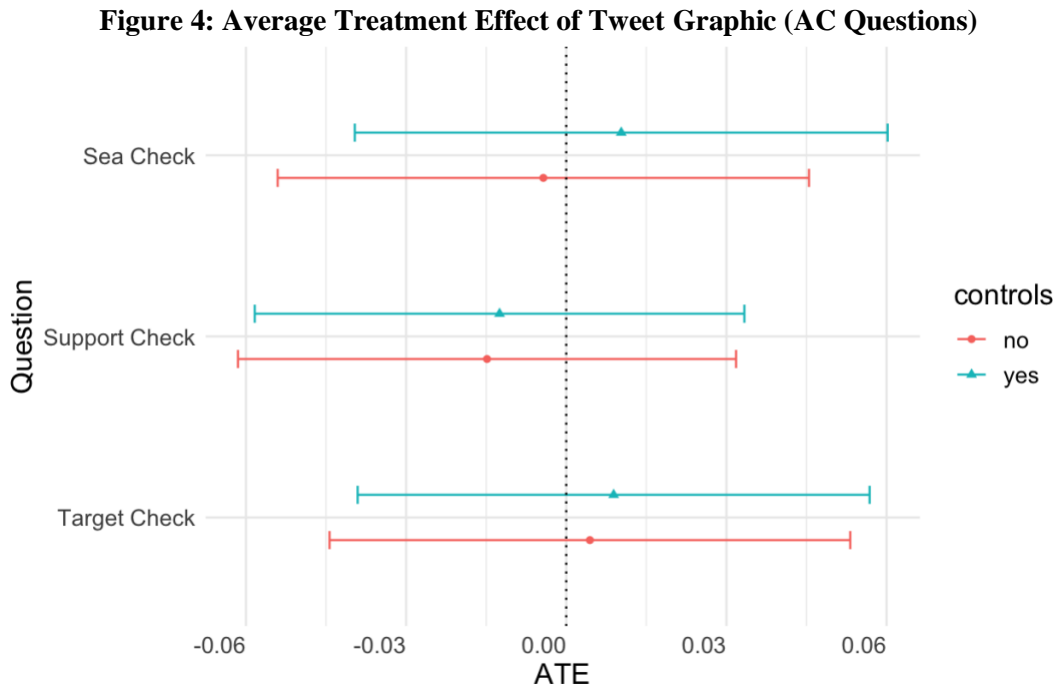
Table 2: Timer and Correct Responses to AC Questions by Treatment

	Time Spent on Treatment (in seconds)	Target Check	Support Check	Sea Check
Tweet Graphic	42.38 (3.76) <i>n</i> = 760	0.63 (0.02) <i>n</i> = 760	0.68 (0.02) <i>n</i> = 760	0.58 (0.02) <i>n</i> = 760
Plain Text	45.18 (3.78) <i>n</i> = 751	0.63 (0.02) <i>n</i> = 751	0.70 (0.02) <i>n</i> = 751	0.58 (0.02) <i>n</i> = 751

Standard errors in parentheses.

⁶ Regression results in Appendix C.

Surprisingly, we find no support for H_2 : there are no statistically significant differences between the contextually realistic tweet and the plain text treatment for our attention check questions and for the time spent reading the treatment.⁷ The average treatment effect for all AC questions is shown in Figure 4. Put differently, more realistic treatments do not appear to enhance information retention.



Error bars represent 95% confidence intervals.

Experiment 2: Leaked Intelligence

H1: Substantive Outcomes

We also measure four substantive outcomes in our leaked intelligence experiment: 1) credibility of the document contents (i.e., whether Russia supplied chemical weapons), 2) international perceptions (i.e., whether the international community will believe Russia has exported chemical weapons) 3) authenticity of the document, and 4) crisis realism (i.e., whether

⁷ Regression results in Appendix B.

a similar leak could happen in the real world).⁸ As before, we use a five-point Likert scale and run regression models. The mean response for the graphic and plain text treatments are presented in Table 3; Figure 5 displays the average treatment effect for model 1 and model 2.⁹

Table 3: Mean Response to Substantive Questions by Treatment

	Credibility	Int'l Perception	Authenticity	Crisis Realism
Leaked Document Graphic	3.86 (0.04) <i>n</i> = 754	3.96 (0.03) <i>n</i> = 754	3.61 (0.03) <i>n</i> = 754	4.14 (0.03) <i>n</i> = 754
Plain Text	3.94 (0.04) <i>n</i> = 757	4.03 (0.03) <i>n</i> = 757	3.77 (0.03) <i>n</i> = 757	4.20 (0.03) <i>n</i> = 757

Standard errors in parentheses.

We again find support for H_1 . We find no significant treatment effect on three of our variables: credibility, international perception, and crisis realism. However, we do find a strong *negative* correlation between the treatment and perceptions of the document's authenticity. Respondents in the graphic treatment were more likely to doubt the authenticity of the leaked ICA ($p < 0.05$).

We suspect that this negative correlation results from our decision to omit classification markings from the graphical depiction of the leaked document—something that appears to have hamstrung our efforts to make an authentic-looking document. We reran our experiment on a different Lucid sample of 1,206 respondents, altering only the country supplying the chemical weapons. As in our original experiment, we find a strong negative correlation between the treatment and authenticity ($p < 0.01$).¹⁰ The formatting of intelligence documents may have been particularly salient to respondents given that a large intelligence leak occurred just prior to survey

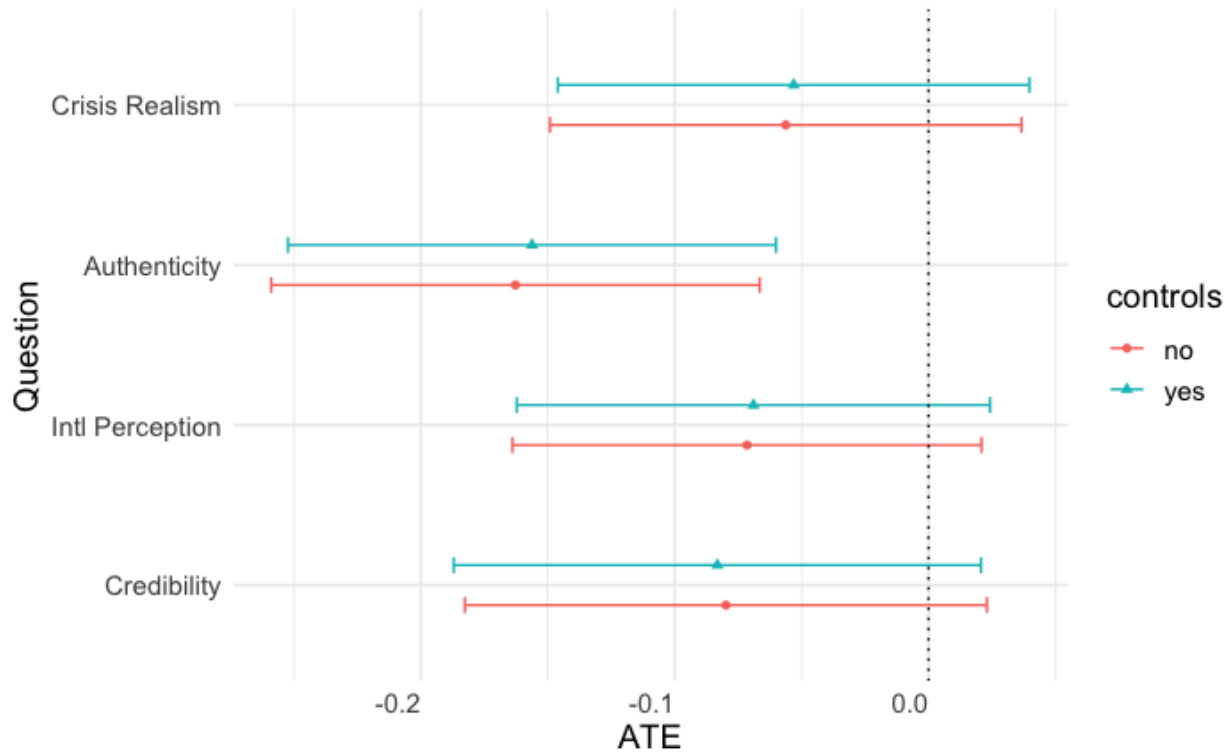
⁸ Question wording in Appendix A.

⁹ Regression results in Appendix B.

¹⁰ Regression results in Appendix C.

fielding. Importantly, this lack of perceived authenticity did not affect our other substantive dependent variables.

Figure 5: Average Treatment Effect of Intelligence Graphic (Substantive Questions)



Error bars represent 95% confidence intervals.

H2: Information Retention

As in the Twitter experiment, we asked respondents in the leaked document experiment three attention check questions, and we measured the time spent on the vignette.¹¹ Specifically, we ask: where the chemical weapons were used (“Used Check”), what country supplied the weapons (“Supplied Check”), and what entity wrote the intelligence report (“Intel Check”). Table 4 displays mean responses by treatment, and Figure 6 shows the average treatment effect.

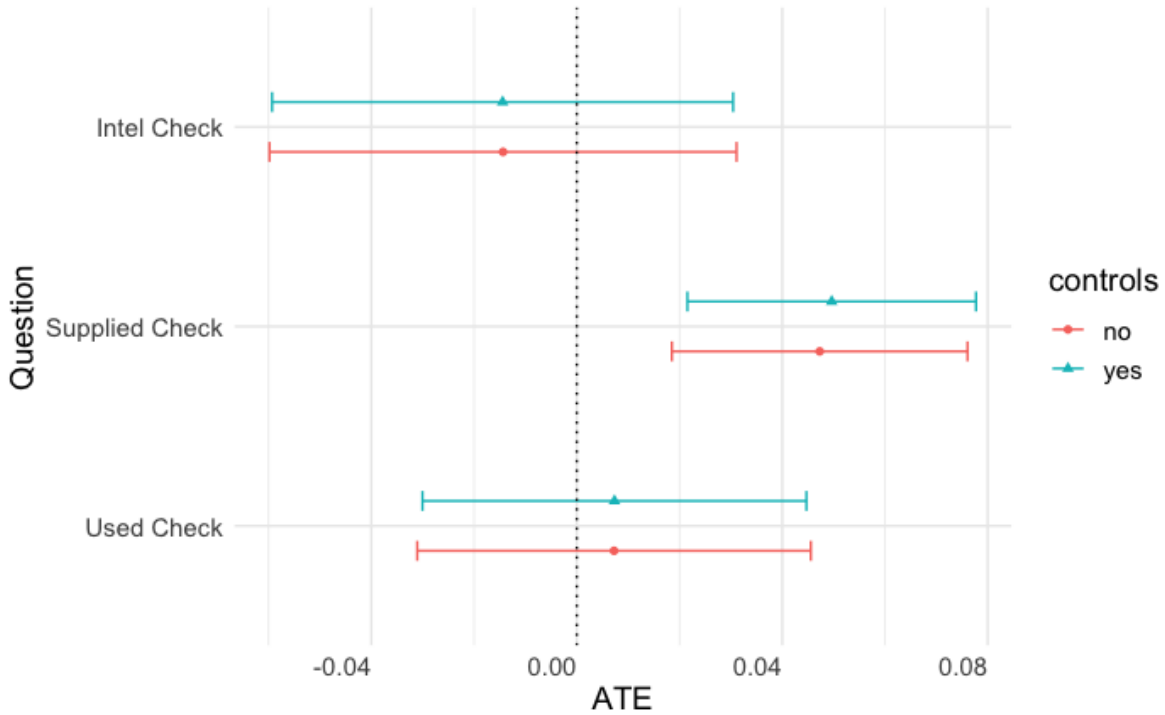
¹¹ Question wording in Appendix A.

Table 4: Timer and Correct Responses to AC Questions by Treatment

	Time Spent on Treatment (in seconds)	Used Check	Supplied Check	Intel Check
Leaked Document Graphic	46.31 (2.486) <i>n</i> = 754	0.83 (0.01) <i>n</i> = 754	0.93 (0.01) <i>n</i> = 754	0.71 (0.01) <i>n</i> = 754
Plain Text	47.17 (2.481) <i>n</i> = 757	0.82 (0.01) <i>n</i> = 757	0.89 (0.01) <i>n</i> = 757	0.72 (0.01) <i>n</i> = 757

Standard errors in parentheses.

Figure 6: Average Treatment Effect of Intelligence Graphic (AC Questions)



Error bars represent 95% confidence intervals.

We find little support for H_2 for most of our results; however, we do find a positive correlation between the treatment and the attention check regarding the nation who supplied chemical weapons.¹² We find that respondents in the graphical realism treatment are more likely to correctly identify Russia as the supplier of chemical weapons. One large difference between our

¹² Regression results in Appendix B.

leaked document graphic treatment and the plain text control is that “Russia” appears in large, bolded letters in the leaked document graphic. To examine this result in more depth, we repeated the experiment—as mentioned above—while changing the supplying country from Russia to North Korea. We were unable to replicate the result; in other words, the coefficient for the supplied attention check question was not statistically significant ($p = 0.12$).

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 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-on experiments. The one exception was that our graphically realistic ICA was consistently rated as less authentic by respondents than a plain text description of the ICA. These results suggest political scientists gain little from using contextually realistic graphics in lieu of plain text treatments. Creating realistic graphics requires time and resources that can be better spent elsewhere.

This research note does not represent the final word, but instead suggests further areas of inquiry. Future work could explore whether using contextually realistic graphics affect studies

outside of international relations. For example, scholars of comparative or American politics interested in framing effects might assess whether presenting information as plain text or as a mock newspaper article influences substantive outcomes. Additional work could also 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).

STATEMENTS

Funding Declaration

The authors declare none.

Competing Interests Declaration

The authors declare none.

Data Availability Statement

Upon acceptance, all experimental data will be made publicly available online.

REFERENCES

- Aguinis, Herman, and Kyle J. Bradley. 2014. "Best Practice Recommendations for Designing and Implementing Experimental Vignette Methodology Studies." *Organizational Research Methods* 17 (4): 351–71. <https://doi.org/10.1177/1094428114547952>.
- Alekseev, Aleksandr, Gary Charness, and Uri Gneezy. 2017. "Experimental Methods: When and Why Contextual Instructions Are Important." *Journal of Economic Behavior & Organization* 134 (February): 48–59. <https://doi.org/10.1016/j.jebo.2016.12.005>.
- Barabas, Jason, and Jennifer Jerit. 2010. "Are Survey Experiments Externally Valid?" *American Political Science Review* 104 (2): 226–42. <https://doi.org/10.1017/S0003055410000092>.
- Brutger, Ryan, Joshua Kertzer, Jonathan Renshon, Dustin Tingley, and Chagai Weiss. 2021. "Abstraction and Detail in Experimental Design." In .
- 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 (4): 529–38. <https://doi.org/10.1109/TVCG.2015.2391853>.
- 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 (3): 1874–77.
- 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 (1).
- Dafoe, Allan, Baobao Zhang, and Devin Caughey. 2018. "Information Equivalence in Survey Experiments." *Political Analysis* 26 (4): 399–416. <https://doi.org/10.1017/pan.2018.9>.
- 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 (6): 2006–33. <https://doi.org/10.1111/bjet.13030>.
- Dietrich, Simone, Heidi Hardt, and Haley J. Swedlund. 2021. "How to Make Elite Experiments Work in International Relations." *European Journal of International Relations*, February. <https://doi.org/10.1177/1354066120987891>.
- Egami, Naoki, and Erin Hartman. 2022. "Elements of External Validity: Framework, Design, and Analysis." *American Political Science Review*.
- Gadarian, Shana Kushner. 2014. "Scary Pictures: How Terrorism Imagery Affects Voter Evaluations." *Political Communication* 31 (2): 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*, February, 1–13. <https://doi.org/10.1017/XPS.2021.1>.
- 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): 1–32. <https://doi.org/10.1007/s40692-020-00169-2>.
- Hyde, Susan D. 2015. "Experiments in International Relations: Lab, Survey, and Field." *Annual Review of Political Science* 18 (1): 403–24.

- 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 (2): 110–21.
- Kertzer, Joshua D., and Jonathan Renshon. 2022. "Experiments and Surveys on Political Elites." *Annual Review of Political Science* 25.
- Kreps, Sarah, and Stephen Roblin. 2019. "Treatment Format and External Validity in International Relations Experiments." *International Interactions* 45 (3): 576–94. <https://doi.org/10.1080/03050629.2019.1569002>.
- McDermott, Rose. 2002. "Experimental Methodology in Political Science." *Political Analysis* 10 (4): 325–42.
- McDonald, Jared. 2020. "Avoiding the Hypothetical: Why 'Mirror Experiments' Are an Essential Part of Survey Research." *International Journal of Public Opinion Research* 32 (2): 266–83. <https://doi.org/10.1093/ijpor/edz027>.
- 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 (2): 20 Pages. <https://doi.org/10.12758/MDA.2020.06>.
- Schnotz, Wolfgang. 2001. "Towards an Integrated View of Learning From Text and Visual Displays." *Educational Psychology Review*.
- 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*, June, 0049124119852382. <https://doi.org/10.1177/0049124119852382>.
- 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 (2): 149–57. <https://doi.org/10.1002/hbe2.184>.
- 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 (2). <https://doi.org/10.2458/v7i2.20321>.
- Tomz, Michael. 2007. "Domestic Audience Costs in International Relations: An Experimental Approach." *International Organization* 61 (4): 821–40.
- Tomz, Michael, Jessica Weeks, and Keren Yarhi-Milo. 2020. "Public Opinion and Decisions About Military Force in Democracies." *International Organization* 74 (1): 119–43.
- 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 (5): 401–7.