Jennifer M. Jensen (ISB # 9275)
Anne Henderson Haws (ISB #10412)
HOLLAND & HART LLP
800 W. Main Street, Suite 1750
Boise, ID 83702-7714
Telephone: 208.342.5000
jmjensen@hollandhart.com
ahhaws@hollandhart.com

Attorneys for Petitioner

IN THE SUPREME COURT OF THE STATE OF IDAHO

IDAHOANS UNITED FOR WOMEN AND FAMILIES,

Petitioner,

vs.

RAUL R. LABRADOR, in his official capacity as the Idaho Attorney General, PHIL MCGRANE, in his official capacity as the Idaho Secretary of State, LORI WOLFF, in her official capacity as Administrator of the Idaho Division of Financial Management, and IDAHO DIVISION OF FINANCIAL MANAGEMENT,

Respondents.

Docket No. **52636-2025**

DECLARATION OF DR. HILLARY C. SHULMAN, PH.D. IN SUPPORT OF PETITION FOR WRITS

I, Hillary C. Shulman, Ph.D., declare under penalty of perjury as follows:

1. I am an Associate Professor with the School of Communication at The Ohio State University. My area of academic focus includes Political Communication, but more specifically I study how word choice influences information processing and public engagement in the areas of politics, health, and science.

2. I was retained by Idahoans United for Women and Families ("Idahoans United") to analyze and provide my professional conclusions regarding the Fiscal Impact Statement, Short Title, and Long Title provided by the State of Idaho relating to the proposed citizen's ballot initiative titled "Reproductive Freedom and Privacy Act" ("Initiative").

3. A true and correct copy of my report ("Report") relating to the Initiative is attached hereto as **Exhibit A**.

4. A copy of my CV and list of publications is attached hereto as **Exhibit B**.

5. As described more fully in the Report, I can attest, with a reasonable degree of scientific certainty, to the four following conclusions regarding the Fiscal Impact Statement and two following conclusions related to the use of the uncommon term "fetus viability" in the Short Title and Long Title for the Initiative.

6. **Conclusion One (Fiscal Impact Statement) -** The Fiscal Impact Statement includes legal and technical terms in its explicit reference to two legal statutes (*Idaho Codes 20-237B and 56-255*). Voter's lack of knowledge of the meaning of these statutes, coupled with their inability to look these laws up in the voting booth, obfuscate the conclusions voters are able to draw about the fiscal impacts of this bill.

7. **Conclusion Two (Fiscal Impact Statement) -** The Fiscal Impact Statement contains reference to numbers and dollar amounts which are contradictory to the claim, "*The laws affected by the initiative would not impact income, sales, or product taxes. There is no revenue impact to the General Fund found.*" The subsequent inclusion of the numerical amounts of "$*20,000"* and "$*850 million in FY2024*" are likely to draw reader's attention and create the impression that the bill costs money. This is an inaccurate impression and thus misleading.

8. **Conclusion Three (Fiscal Impact Statement) -** There are references to two groups in the Fiscal Impact Statement that, based on strong scholarly precedent, will negatively bias voters' impressions of the bill. The Fiscal Impact Statement prejudices the initiative by including the terms "Medicaid" and "prisoners", a program and population negatively viewed my many, which are unrelated to the purpose of the initiative. When language that references discriminated against identities, or populations, is used, prejudicial attitudes are likely to become top of mind. This is a well-documented communication and psychological process referred to as a *framing effect*. In this instance, this framing effect will increase the likelihood that negative sentiment towards this legislation is produced, due to the evocation of these prejudices. This will negatively bias voters' impression of the legislation and will reduce the legislations likelihood of passage accordingly.

9. **Conclusion Four (Fiscal Impact Statement) -** The referenced issues with the Fiscal Impact Statement independently and collectively will confuse many voters and cause them to vote against the Initiative for reasons unrelated to the purpose and true fiscal impact of the Initiative.

10. **Conclusion One (Titles) -** The interchangeable use of the term "fetal" and "fetus" between the Short and Long Title is not semantically accurate. "Fetal" is a general term that relates to a fetus. Conversely, "fetus" is singular. Drawing upon the rich academic literature and studies in framing theory once again, references to general themes versus individual instances evoke different types of considerations. Specifically, singular references, known as "episodic" framing, is likely to increase attributions of individual responsibility and/or blame. The presence of these considerations will negatively bias attitudes towards this bill and negatively impact the

legislations' likelihood of passage. Notably, the phrase "fetal" does not conjure up individual responsibility considerations and would remedy this concern.

11. **Conclusion Two (Titles) -** Given that "fetal viability" is the commonly accepted medical term, there appears to be no good faith basis for the interjection of the uncommon term "fetus viability". The only purpose in using the uncommon term "fetus viability" is to bias voters against the Initiative.

12. The Report explains my conclusions in greater detail, identifies the materials that I reviewed, notes certain representations made my legal counsel for Idahoan United that I have accepted, and cites the bases for my conclusions, including references to my studies and other peer-reviewed academic literature.

13. My report contains hyperlinks to the key articles that I reference. In addition, an index and copies of the key articles cited in my Report are attached as **Exhibit C.**

I declare under penalty of perjury pursuant to the laws of the State of Idaho that the foregoing is true and correct.

DATED: January 30, 2025

/s/Hillary C. Shulman
Hillary C. Shulman, Ph.D.

CERTIFICATE OF SERVICE

I hereby certify that on this 30th day of January, 2025, I caused to be filed, via iCourt, and served a true and correct copy of the foregoing by the method indicated below, and addressed to the following:

Div. of Financial Management P.O. Box 83720 Boise, ID 83720-0032 info@dfm.idaho.gov	☑ U.S. Mail ☐ Hand Delivered ☐ Overnight Mail ☐ Email/iCourt/eServe:
Administrator Lori Wolff Idaho Division of Financial Management P.O. Box 83720 Boise, ID 83720-0032 info@dfm.idaho.gov	☑ U.S. Mail ☐ Hand Delivered ☐ Overnight Mail ☐ Email/iCourt/eServe:
Office of the Attorney General 700 W. Jefferson Street, Suite 210 P.O. Box 83720 Boise, ID 83720-0010 aglabrador@ag.idaho.gov	☑ U.S. Mail ☐ Hand Delivered ☐ Overnight Mail ☐ Email/iCourt/eServe:
Idaho Secretary of State P.O. Box 83720 Boise, ID 83720-0080 secretary@sos.idaho.gov	☑ U.S. Mail ☐ Hand Delivered ☐ Overnight Mail ☐ Email/iCourt/eServe:

/s/ Jennifer M. Jensen
Jennifer M. Jensen
FOR HOLLAND & HART LLP

EXHIBIT A

THE OHIO STATE UNIVERSITY

Expert Report
of
Dr. Hillary Shulman, PhD.

January 29, 2025

I. SUMMARY OF REPORT

I am an Associate Professor of Communication at The Ohio State University. My area of academic focus includes Political Communication, but more specifically I study how word choice influences information processing and public engagement in the areas of politics, health, and science. I was retained by Idahoans United for Women and Families ("Idahoans United") to analyze and provide my professional conclusions regarding the Fiscal Impact Statement, Short Title, and Long Title provided by the State of Idaho relating to the proposed citizen's ballot initiative titled "Reproductive Freedom and Privacy Act" ("Initiative").

Although I discuss these conclusions, and the evidence that supports my analysis, in the greater opinion below, my expert opinions, based on a reasonable scientific certainty, in this matter follow:

A. FISCAL IMPACT STATEMENT

The Fiscal Impact Statement is not written in clear and concise language and included terms which create a bias against the Initiative. I attest to a reasonable scientific certainty the lack of clarity and concision and the biased terms discussed below will confuse voters, bias voters, and adversely impact voter's understanding of the initiative.

1. The Fiscal Impact Statement includes legal and technical terms in its explicit reference to two legal statutes (*Idaho Codes 20-237B and 56-255*). Voter's lack of knowledge of

the meaning of these statutes, coupled with their inability to look these laws up in the voting booth, obfuscate the conclusions voters are able to draw about the fiscal impacts of this bill.

2. The Fiscal Impact Statement contains reference to numbers and dollar amounts which are contradictory to the claim, "*The laws affected by the initiative would not impact income, sales, or product taxes. There is no revenue impact to the General Fund found.*" The subsequent inclusion of the numerical amounts of "$*20,000"* and "$*850 million in FY2024*" are likely to draw reader's attention and create the impression that the bill costs money. This is an inaccurate impression and thus misleading.

3. There are references to two groups in the Fiscal Impact Statement that, based on strong scholarly precedent, will negatively bias voters' impressions of the bill.

4. The Fiscal Impact Statement prejudices the initiative by including the terms "Medicaid" and "prisoners", a program and population negatively viewed my many, which are unrelated to the purpose of the initiative. When language that references discriminated against identities, or populations, is used, prejudicial attitudes are likely to become top of mind. This is a well-documented communication and psychological process referred to as a *framing effect*. In this instance, this framing effect will increase the likelihood that negative sentiment towards this legislation is produced, due to the evocation of these prejudices. This will negatively bias voters' impression of the legislation and will reduce the legislations likelihood of passage accordingly.

5. The referenced issues with the Fiscal Impact Statement independently and collectively will confuse many voters and cause them to vote against the Initiative for reasons unrelated to the purpose and true fiscal impact of the Initiative.

B. USE OF UNCOMMON TERM "FETUS VIABILITY" IN THE SHORT AND LONG TITLE

1. The interchangeable use of the term "fetal" and "fetus" between the Short and Long Title is not semantically accurate. "Fetal" is a general term that relates to a fetus. Conversely, "fetus" is singular. Drawing upon the rich academic literature and studies in framing theory once again, references to general themes versus individual instances evoke different types of considerations. Specifically, singular references, known as "episodic" framing, is likely to increase attributions of individual responsibility and/or blame. The presence of these considerations will negatively bias attitudes towards this bill and negatively impact the legislations' likelihood of passage. Notably, the phrase "fetal" does not conjure up individual responsibility considerations and would remedy this concern.

2. Given that "fetal viability" is the commonly accepted medical term, there appears to be no good faith basis for the interjection of the uncommon term "fetus viability". The only purpose in using the uncommon term "fetus viability" is to bias voters against the Initiative.

II. BASIS FOR EXPERT QUALIFICATION

C. RELEVANT STUDY AND EXPERIENCE.

I have a Ph.D. in Communication and am an Associate Professor of Communication at The Ohio State University. My area of academic focus includes Political Communication. I study how wording on ballot initiatives impact the way voters' process information, and how this information processing can impact vote choice. I have conducted extensive research and published numerous peer-reviewed articles relating to political communication and how the wording on ballot initiatives predicts voter choice and election outcomes. In addition to my work on ballot language,

I also study how people attend to information in information rich, or highly complex, information environments.

Relevant examples of my accepted and published academic work include:

- Shulman, H. C., Holt, L. F., Riggs, E. E.*, Wade, R. B.* (conditional acceptance). The role of framing, race, and symbolic racism in policy support. Political Communication - This two-study experiment finds that when a ballot initiative includes language that implies that the measure will impact Black communities, people are less likely to support the ballot.

- Shulman, H. C., Sweitzer, M. D.*, Bullock, O. M.*, Coronel, J., Bond, R. M., & Poulsen, S.* (2022). Predicting vote choice and election outcomes from ballot wording: The role of processing fluency in low information direct democracy elections. Political Communication, 39(5), 652-673. https://doi.org/10.1080/10584609.2022.2092920. **Michael Pfau Outstanding Article Award Recipient** - This two-study experiment, testing over 75 real ballot initiatives, found that when ballots featured highly complicated language, people were less likely to support the ballot.

- Coronel, J., Bullock, O. M.*, **Shulman, H. C.,** Sweitzer, M. D.*, Bond, R. M., & Poulsen, S.* (2021). Eye movements predict large-scale voting decisions. *Psychological Science, 32*(6), 836-848. https://doi.org/10.1177/0956797621991142. This study brought registered voters into the lab and asked them to read ballot initiatives. We examined how people actually read ballots, using an eye tracker, and how reading patterns correspond with vote outcomes in real elections. Here again, more difficulties while reading reduced support for these ballots.

- **Shulman, H. C.,** Markowitz, D. M., & Rogers, T. (2024). Reading dies in complexity: Online news readers prefer simple writing. *Science Advances, 10*(23), 1-8. https://www.science.org/doi/epdf/10.1126/sciadv.adn2555. This series of studies investigated what types of information people attend to, or notice, in crowded information environments.

A copy of my CV and list of peer-reviewed articles is attached as **Exhibit B** to the Declaration of Dr. Hilary Shulman filed January 30, 2025.

D. MATERIALS CONSIDERED

In preparing this report, I have reviewed and considered the following documents and materials relating to the Initiative:

- Initiative;
- Short and Long title proposed by the Idahoans United for Women and Families
- Fiscal Impact Statement provided the Idaho Division of Financial Management ;
- Short and Long Title Provided by Office of the Idaho Attorney General;
- Idaho Code Section 34-1812 (requirements for fiscal impact statements);
- Idaho Code 20-237B (referenced IDFM's fiscal statement for initiative);
- Idaho Code 56-255 (referenced in IDFM's fiscal statement for Initiative); and
- Summary of Standards Applicable to Short and Long Titles (provided by legal counsel for Idahoans United). Those standards are as follows:

    - "Under I.C. § 34-1809, the fundamental inquiry is whether the short title is 'distinctive,' that is, whether the short title 'set[s] forth the characteristics which distinguish this proposed measure and expeditiously and accurately acquaint the prospective signer with what he or she is sponsoring.'" ACLU v. Echohawk, 124 Idaho 147, 151, 857 P.2d, 626, 630 (1993). This Court has defined "distinctive" as: "(1) referring primarily to that which marks or distinguishes one thing regarded in its relation to other things, (2) a mark or character indicating separation, (3) distinguishing from something diverse, or (4) serving or used to distinguish or discriminate." Buchin v. Lance, 128 Idaho at 270, 912 P.2d at, 638.
    - In addition to analyzing whether the short title comprehensively captures the distinctive characteristics of the initiative, the Court also evaluates whether the title is neutral, as opposed to "argumentative" or "prejudicial." *ACLU*, 124 Idaho at 149, 151, 857 P.2d at 628, 630; *see also Idahoans for Open Primaries*, 172 Idaho at 482-83, 533 P.3d at 1278-79 (holding short ballot title failed substantial compliance review in part because the term "nonparty blanket primary" was prejudicial and misleading). Neutrality is important because in preparing the ballot titles, the Attorney General is entrusted with a "quasi-judicial" task. *Buchin*, 128 Idaho at 270, at 638. Terminology posing a risk of confusing the voters is prejudicial. *See Idahoans for Open Primaries*, 172 Idaho at 486, 533 P.3d at 1282 ("At a

minimum, the statement is ambiguous; thus, it is likely to prejudice the Initiative.").

- "The plain and unambiguous language of section 34-1809(2)(d)(i) requires the Attorney General to ascertain how an initiative is commonly referred to or spoken of and incorporate that language into the short title." *Idahoans For Open Primaries*, 172 Idaho at 481, 533 P.3d at 1277. This "task necessarily requires the Attorney General to determine how Idahoans commonly refer to and speak of a measure[.]" *Id.* The Court has also held that, where the "the Attorney General is unable to identify the common language Idahoans use to refer to the measure," the Attorney General "may look outside of the state to determine whether common language can be found in other states." *Id.* The job of "[a]scertaining how *the public* refers to a measure is important because the short title must 'set forth the characteristics which distinguish [the] proposed measure and expeditiously and accurately acquaint the prospective signer with what he is sponsoring.'" *Id.* (quoting *In re Idaho State Fed'n of Labor*, 75 Idaho at 373, 272 P.2d at 710).
- "When ascertaining the language used to commonly refer to the measure, the Attorney General must remain mindful that the statute requires him to use language that is not 'intentionally an argument or likely to create prejudice either for or against the measure.'" Id. at 466, 481, 533 P.3d at 1262, 1277 (quoting Idaho Code § 34-1809(2)(e)).

In addition, I have assumed the accuracy of the following statements by counsel for Idahoans United:

- News articles published throughout the state of Idaho in the Idaho Statesman, the Coeur d'Alene Press, the Lewiston Morning Tribune, the Idaho Press Tribune, the Bonner County Daily Bee, the Idaho State Journal, the Post Register, the Moscow-Pullman Daily News, and the Twin Falls Times are replete with the use of the term "fetal viability" in reporting on the topic. Meanwhile, a search for "fetus viability" in these Idaho newspapers' online archives going back a decade turns up markedly blank—substantial evidence that the term "fetus viability" is not the common language Idahoans use.
- Outside of Idaho, the evidence of common usage is the much the same. A search of the online archives of the Wall Street Journal, the New York Times, the Salt Lake Tribune, and the Las Vegas Sun for the term "fetal viability" turned up blank. Only by searching for the word "abortion" in these papers' archives was any use found, once in the New York Times and once in the Wall Street Journal, in 1997 and 1996,

respectively. Casting an even wider net, by adding the Chicago Tribune, the Los Angeles Times, the Seattle Times, and the Washington Post to the search, conclusively shows that the term "fetal viability" is used far more frequently than "fetus viability" at the national level. Id. For instance, of the 488 search returns for the term "fetal viability" in the Washinton Post, only 19 returns showed for "fetus viability."

- Casting a different net yields substantially similar results. A Google search of the term "fetus viability" automatically returns results for "fetal viability" instead—both in its general "All" search category and when the search is filtered to the "News" category.

In relevant part, Idaho Code Section 34-1812 states:

> *"A fiscal impact statement must be written in clear and concise language and shall avoid legal and technical terms whenever possible. Where appropriate, a fiscal impact statement may include both estimated dollar amounts and a description placing the estimated dollar amounts into context."*

The fiscal impact statement created by Idaho's DFM states:

> The laws affected by the initiative would not impact income, sales, or product taxes. There is no revenue impact to the General Fund found.
>
> The initiative could change state expenditures in minor ways. Costs associated with the Medicaid and prisoner populations may occur; see Idaho Codes 20-237B and 56-255 and the Medicaid references from Health and Welfare.
>
> Passage of this initiative is likely to cost less than $20,000 per year. The Medicaid budget for providing services was about $850 million in FY2024. If passed, nominal costs in the context of the affected total budget are insignificant to the state.

In my professional opinion, the final two paragraphs of the Fiscal Impact Statement are not clear and concise, unnecessarily contain legal and technical terms, and interjects prejudicial references to Medicaid and Prisoners which are unrelated to the Initiative.

First, the second paragraph of the Fiscal Impact Statement provides an explicit reference to *Idaho Codes 20-237B and 56-255*. This language is not only both legal and technical, but disingenuous because a voter in the voting booth has no ability to access the meaning of these codes. Citing Idaho Code sections creates the impression that there is legalistic nuance to this statement and will lead people to infer that the bill will cost taxpayers money in unknown ways. To evidence this claim, work by Markowitz et al. (2021) found that when companies used highly technical and confusing language in their values statement, an idea known as linguistic obfuscation (p. 278), people perceived the company as trying to be intentionally dishonest and untrustworthy. Further, my own research (Shulman et al., 2022) found, across 64 ballot measures, ballots that contained more complex and legalistic language were more opposed by participants than ballots written with simpler and more colloquial language. In short, based on established research I can only conclude that that references to legal codes can turn off voters and make them feel as though there are unknown costs, beyond their understanding, associated with this piece of legislation (see also, Bullock et al., 2020; Coronel et al., 2021; Martínez et al., 2023)

Second, the inclusion of the numbers *"$20,000"* and "$*850 million in FY2024*" is unclear and will lead to misinterpretation. When people read ballots, they do not read them left to right and up and down. In other words, they do not read the text in order. Instead, our eye tracking data on how people actually read ballots shows that people scan the document and bounce around to salient parts (Coronel et al., 2021). Here, the presence of numbers, in the financial impact section, will clearly draw reader's attention. For those who do not read closely, a trait of most readers in information dense environments (Shulman et al., 2024), the presence of these numbers in this section will indicate that this act will cost money. Although Idaho Code does stipulate that, *"Where*

appropriate, a fiscal impact statement may include both estimated dollar amounts and a description placing the estimated dollar amounts into context" in this instance it is not appropriate, and misleading, to include these numbers considering the very first paragraph of this section. If the first paragraph reads, *"The laws affected by the initiative would not impact income, sales, or product taxes. There is no revenue impact to the General Fund found,"* then including dollar amounts (especially a number like $850 million) in subsequent paragraphs will confuse voters rather than inform them. This is antithetical to the intent stated in the Idaho Code referenced above.

Finally, the reference to the "*Medicaid*" and "*prisoner*" populations referred to in the second paragraph of the impact section confuses and prejudices the Initiative. Referencing groups that are highly prejudiced against will create a negative bias in voters' mind (Mendelberg, 2001). Given that this initiative is not about these groups, referencing these two - discriminated against – populations can only be read as duplicitous. For instance, my experimental research (Shulman et al., included in the supplement) finds that when ballot initiatives include language that signals the ballot will be enacted in a community of color (e.g., "urban" versus "suburban), people are less likely to vote in support of this ballot. The idea that people are less likely to support policy when discriminated groups are referenced is a well-established finding in the social sciences in an area of research that looks at implicit racial appeals (see Gilens, 2009) and dog whistle politics (Haney-López, 2014).[1] For instance, work by Hurwitz and Peffly (2005) found that when White respondents were asked about whether they supported spending money on prisons, people were more supportive of doing so when racially coded language (violent "inner city" criminals), as

[1] The three books referenced in this paragraph (Gilens, 2009; Haney-López, 2014; Mendelberg, 2001) have been cumulatively cited 7,812 times according to Google Scholar (as of 1.30.2025).

opposed to non-racially coded language ("violent criminals), was used. Given this evidence, it is my strong belief that the inclusion of this language, with no context, is meant to exploit prejudices that undermine a fair consideration of this initiative.

In summary, as a researcher who studies word choice and information processing in the context of ballot initiatives, I can conclude to a reasonable scientific certainty that the financial impact statement, as written, does *not* sufficiently adhere to the standards of Idaho Code 34-1812. Direct democracy requires fair, accessible, and balanced language to let the democratic process function as intended. Violations to this standard directly undermine this process, reduce voter trust in elections, and create a precedent that allows for the manipulation of voters through word choice.

III. PROFESSIONAL ASSESSMENT OF THE SHORT AND LONG TITLE

From a semantic perspective, the terms "fetal" and "fetus" are not the same, despite being used interchangeably between the Short and Long Title. In my expert opinion, and based on reasonable scientific certainty, the use of the term "fetus" will negatively bias attitudes towards this initiative. The term "fetus" is singular. By contrast "fetal" is a more general term that references, or reflects, this stage of development. Work in framing theory (e.g., Gross, 2008; Iyengar, 1996; Spring & Harwood, 2015) has found that policy support is impacted by the use of general (i.e., thematic) versus individualized (i.e., episodic) terms. Policy attitudes are impacted by this language choice because while attributions of individual responsibility and/or blame are associated with individualized terms, this is not the case for more general language (Iyengar, 1996). In the context of this initiative, the concern is that the episodic frame referenced in the Long Title ("fetus") will make attributions of responsibility and/or blame towards individuals salient. And when this initiative is personalized, through the use of "fetus", policy support will be

negatively impacted by this attributional process. Work by Springer and Harwood (2015) illustrates this point. In this experiment, researchers found that an episodic frame, compared to a thematic frame, increased endorsement of individual responsibility for retirement planning. By making individual responsibility salient, participants reported more negative attitudes towards social security. Applying these dynamics to the current case, the use of the singular "fetus" is likely to make attributions of blame salient, and these attributions will negatively impact attitudes towards the legislation (see also Iyengar, 1996). In addition to attributions of responsibility and blame, episodic framing also impacts the publics' emotional response to information (Gross, 2008). Specifically, Gross (2008) contends that people respond more emotionally to episodic framing, whereas they respond more cognitively to thematic framing. To obviate the interjection of blame attributions and emotional processing in response to word choice on this initiative, I recommend consistent usage of "fetal" throughout the Long and Short Titles, as attributions of blame and/or responsibility and emotionality have not been associated with language of this kind.

34116736_v3

EXHIBIT B

Hillary C. Shulman, Ph.D.

Curriculum Vitae

Associate Professor
School of Communication
The Ohio State University
Shulman.36@osu.edu, (847) 772 – 8949

Education

- **Ph.D. Communication**
  Michigan State University, August 2011
  Specializations: Political, Interpersonal, and Group Communication
  Advisor: Dr. Timothy Levine
  Committee: Drs. Dan Bergan, William Jacoby (Political Science), and Maria Lapinski

- **M.A. Communication**
  Michigan State University, May 2007
  Advisor: Dr. Timothy Levine
  Committee: Drs. Chuck Atkin and Stan Kaplowitz (Sociology)

- **B.S. Communication Science**
  University of Wisconsin at Madison, May 2004

Specialized Academic Training

- **Summer institute in political psychology, Stanford University**- Completed 90 hours of academic training in political psychology, July 2008.

Professional Experience

- *Associate Professor*, School of Communication, The Ohio State University
  Columbus, Ohio, May 2022 - Present

- *Assistant Professor*, School of Communication, The Ohio State University
  Columbus, Ohio, August 2015- May 2022

- *Visiting Assistant Professor*, School of Communication, The Ohio State University
  Columbus, Ohio, August 2014- 2015

- *Assistant Professor*, Department of Speech Communication, North Central College
  Naperville, Illinois, September 2011 – 2014

- *Research Assistant*, Department of Media and Information Studies, Michigan State

University, and the Department of Energy, Labor, and Economic Growth, State of Michigan, Lansing, Michigan, August 2010 – August 2011

- *Teaching and Research Graduate Assistant*, Department of Communication, Michigan State University, East Lansing, Michigan, May 2007 – August 2010

- *Communication Consultant Graduate Assistant*, Department of Accounting, Eli Broad College of Business, Michigan State University, East Lansing, Michigan, 2005 – 2007

Peer-Reviewed Publications
***Indicates student as co-author at the time of data collection**

41. **Shulman, H. C.,** Holt, L. F., Riggs, E. E.*, Wade, R. B.* (conditional acceptance). The role of framing, race, and symbolic racism in policy support. *Political Communication*

40. Wade, R. B.,* Ryu, J.,*, **Shulman, H. C.,** & Hovick, S. (2024). Improving processing fluency to encourage family health information seeking: The mediating role of communication efficacy. Journal of Health Psychology, https://doi.org/10.1177/13591053241294116 [Advance online publication]

39. Fort, K. S.*, Lopez, R., **Shulman, H. C.,** Riggs, E. E.*, Cruz Ibarra, J.* (2024). The impacts of code-mixing in a cross-cultural narrative: How processing fluency impacts narrative engagement and racial attitudes. *Communication Research*, https://doi.org/10.1177/00936502241287334 [Advance online publication]

38. Bashian, S.*, Wade, R. B.*, Lerner, B.*, & **Shulman, H. C.** (2024). When fears come true: An experimental approximation of patient comprehension during initial cancer diagnoses. *Health Communication*, https://doi.org/10.1080/10410236.2024.2400819 [Advance online publication]

37. Fort. K. S.* & **Shulman. H. C.** (2024). Using a signal detection approach to understand the impacts of processing fluency and efficacy on accuracy in misinformation detection. *Frontiers in Psychology*, *15*, 1417910. doi: 10.3389/fpsyg.2024.1417910

36. Lerner, B.*, & **Shulman, H.C.** (2024). Science terms elicit ideological differences in message processing. *Communication Research Reports*, *41*(3), 793-198. http://dx.doi.org/10.1080/08824096.2024.2382743.

35. **Shulman, H. C.,** Markowitz, D. M., & Rogers, T. (2024). Reading dies in complexity: Online news readers prefer simple writing. *Science Advances*, *10*(23), 1-8. https://www.science.org/doi/epdf/10.1126/sciadv.adn2555
Note: First and second author shared first authorship.

34. Bergan, D. E., **Shulman, H. C.,** & Carnahan, D. (2024). Discounting constituent attitudes: Motivated reasoning, ambiguity, and policymaker perceptions of constituent characteristics. *Human Communication Research, 50,* 53-65. https://doi.org/10.1093/hcr/hqad047

33. **Shulman, H. C.,** Sweitzer, M. D.*, Bullock, O. M.*, Coronel, J., Bond, R. M., & Poulsen, S.* (2022). Predicting vote choice and election outcomes from ballot wording: The role of processing fluency in low information direct democracy elections. *Political Communication, 39*(5), 652-673. https://doi.org/10.1080/10584609.2022.2092920
Michael Pfau Outstanding Article Award Recipient

32. Riggs, E. E.*, **Shulman, H. C.,** & Lopez, R.* (2022). Using infographics to reduce the negative effects of jargon on intentions to vaccinate against COVID-19. *Public Understanding of Science, 31*(6), 751-765. https://doi.org/10.1177/09636625221077385

31. **Shulman, H. C.,** Bullock, O. M.*, & Riggs, E. E.* (2021). The interplay between jargon, motivation, and fatigue while processing COVID-19 crisis communication over time. *Journal of Language and Social Psychology, 40*(5-6), 546-573. https://doi.org/10.1177/0261927X211043100

30. Bullock, O. M.*, **Shulman, H. C.**., & Huskey, R. (2021). Narratives are persuasive because they are easier to understand: Examining processing fluency as a mechanism of narrative persuasion. *Frontiers in Communication, 6,* 1-12. https://doi.org/10.3389/fcomm.2021.719615

29. Coronel, J., Bullock, O. M.*, **Shulman, H. C.,** Sweitzer, M. D.*, Bond, R. M., & Poulsen, S.* (2021). Eye movements predict large-scale voting decisions. *Psychological Science, 32*(6), 836-848. https://doi.org/10.1177/0956797621991142

28. Bullock, O. M.* & **Shulman, H. C.** (2021). Utilizing framing theory to design more effective health messages. *Communication Studies, 72*(3), 319-332. https://doi.org/10.1080/10510974.2021.1899007

27. Markowitz, D. M., & **Shulman, H. C.** (2021). The predictive utility of word familiarity for online engagements and funding. *Proceedings of the National Academy of Sciences of the United States of America, 118*(18). 1-7. https://doi.org/10.1073/pnas.2026045118

26. **Shulman, H. C.,** & Bullock, O. M.* (2020). Don't dumb it down: The effects of jargon in COVID-19 crisis communication. *PLOS ONE, 15*(10). 1-10. https://doi.org/10.1371/journal.pone.0239524

25. **Shulman, H. C.,** Dixon, G. N., Bullock, O. M.*, & Colón Amill, D.* (2020). The effects of

jargon on processing fluency, self-perceptions, and scientific engagement. *Journal of Language and Social Psychology, 39*(5-6), 579-597. https://doi.org/10.1177/0261927X20902177

24. Rhodes, N., **Shulman, H. C.,** & McClaren, N.* (2020). Changing norms: A meta-analytic integration of research on social norms appeals. *Human Communication Research*, 46(2-3), 161-191. https://doi.org/10.1093/hcr/hqz023

23. Bullock, O. M.*, Colón Amill, D.*, **Shulman, H. C.,** Dixon, G. N. (2019). Jargon as a barrier to effective science communication: Guidance from metacognition. *Public Understanding of Science, 28*(7), 845-853. https://doi.org/10.1177/0963662519865687

22. **Shulman, H. C.,** & Bullock, O. M.* (2019). Using metacognitive cues to amplify message content: A new direction in strategic communication. *Annals of the International Communication Association, 43*(1), 24-39. https://doi.org/10.1080/23808985.2019.1570472
Top Article Award Recipient

21. Sweitzer, M. D.* & **Shulman, H. C.** (2018). The effects of metacognition in survey research: Experimental, cross-sectional, and content-analytic evidence. *Public Opinion Quarterly, 82*(4), 745-768. https://doi.org/10.1093/poq/nfy034

20. **Shulman, H. C.,** & Sweitzer, M. D.* (2018). Advancing framing theory: Designing an equivalency frame to improve political information processing. *Human Communication Research, 44*(2), 155-175. https://doi.org/10.1093/hcr/hqx006

19. Bond, R. M., **Shulman, H. C.,** Gilbert, M.* (2018). Does having a political discussion help or hurt intergroup perceptions?: Drawing guidance from social identity theory and the contact hypothesis. *International Journal of Communication, 12*, 1-21. https://ijoc.org/index.php/ijoc/article/view/9033

18. **Shulman, H. C.,** & Sweitzer, M. D.* (2018). Varying metacognition through public opinion questions: How language can affect political engagement. *Journal of Language and Social Psychology, 37*(2), 224-237. https://doi.org/10.1177/0261927X17707557

17. **Shulman, H. C.,** Rhodes, N., Davidson, E.*, Ralston, R.*, Borghetti, L.*, & Morr, L.* (2017). The state of the field of social norms research. *International Journal of Communication, 11*, 1-21. https://ijoc.org/index.php/ijoc/article/view/6055

16. **Shulman, H. C.,** & DeAndrea, D. C. (2014). Predicting success: Revisiting assumptions about family political socialization. *Communication Monographs, 81*(3), 386-406. https://doi.org/10.1080/03637751.2014.936478

15. **Shulman, H. C.**, & Boster, F. J. (2014). The effect of test-taking venue and test format on political knowledge test performance. *Communication Methods and Measures, 8*(3), 177-189. https://doi.org/10.1080/19312458.2014.937526

14. Roozen, B.* & **Shulman, H. C.** (2014). Tuning in to the RTLM: Tracking the evolution of language alongside the Rwandan Genocide using social identity theory. *Journal of Language and Social Psychology, 33*(2), 167-184. https://doi.org/10.1177/0261927X13513765

13. **Shulman, H. C.,** & Wittenbaum, G. M. (2013). Group discussion that promotes positive political experiences. *Human Communication, 16*(3), 121-132. [available upon request]

12. Levine, T. R., **Shulman, H. C.**, Carpenter, C., & DeAndrea, D. C. (2013). The impact of accusatory, non-accusatory, bait, and false evidence questioning in deception. *Communication Research Reports, 30*(2), 169-174. https://doi.org/10.1080/08824096.2012.762905

11. Lapinski, M. K., Maloney, E. K., Braz, M. E., & **Shulman, H. C.** (2013). Testing the effects of social norms and behavioral privacy on hand-washing: A field experiment. *Human Communication Research, 39*(1), 21 – 46. https://doi.org/10.1111/j.1468-2958.2012.01441.x

10. **Shulman, H. C.,** & Levine, T. R. (2012). Exploring social norms as a group-level phenomenon: Do political participation norms exist and influence political participation on college campuses? *Journal of Communication, 62*(3), 532-552. https://doi.org/10.1111/j.1460-2466.2012.01642.x

9. Banas, J., Turner, M. M., & **Shulman, H. C.** (2012). A test of competing hypotheses of the effect of mood on persuasion. *Communication Quarterly, 60*(2), 143-164. https://doi.org/10.1080/01463373.2012.668845

8. Levine, T. R., Serota, K. B., **Shulman, H. C.,** Clare, D., Park, H. S., Shaw, A. S., et al. (2011). Sender demeanor: Individual differences in sender believability have a powerful impact on deception detection judgments. *Human Communication Research, 37*(3), 377-403. https://doi.org/10.1111/j.1468-2958.2011.01407.x
Distinguished Article Award Recipient

7. Levine, T. R., Shaw, A. S., & **Shulman, H. C. (**2010). Increasing deception detection accuracy with strategic questioning. *Human Communication Research, 36*(2), 216-231. https://doi.org/10.1111/j.1468-2958.2010.01374.x

6. Levine, T. R., Serota, K. B., **Shulman, H. C. (**2010). The impact of *Lie to Me* on viewers' actual ability to detect deception. *Communication Research, 37*(6), 847-856. https://doi.org/10.1177/0093650210362686

5. Levine, T. R., Shaw, A. S., & **Shulman, H. C.** (2010). Assessing deception detection accuracy with dichotomous truth-lie judgments and continuous scaling: Are people really more accurate when honesty is scaled? *Communication Research Reports, 27*(2), 112-122.

https://doi.org/10.1080/08824090903526638

4. Wittenbaum, G. M, **Shulman, H. C.,** Braz, M. E. (2010) Social ostracism in task groups: The effects of group composition. *Small Group Research, 41*(3), 330-353. https://doi.org/10.1177/1046496410363914

3 DeAndrea, D. C., Carpenter, C. J., **Shulman, H. C.,** & Levine, T. R. (2009). The relationship between cheating behavior and sensation-seeking. *Personality and Individual Differences, 47* (8), 944-947. https://doi.org/10.1016/j.paid.2009.07.021

2. Rimal, R.N., Lapinski, M.K., Klein, K.A., & **Shulman, H.C**. (2009). Risk perceptions of people living with HIV/AIDS: How similarity affects optimistic bias. *Journal of Health Psychology, 14*(2), 251-257. https://doi.org/10.1177/1359105308100209

1. Walther, J. B., Van Der Heide, B., Hamel, L. M., **Shulman, H. C.,** (2009) Self-generated versus other-generated statements and impressions in computer-mediated communication: A Test of warranting theory using Facebook. *Communication Research, 36*(2), 229-253. https://doi.org/10.1177/0093650208330251

Research Grants

Proposals Funded

2. Bergan, D., & **Shulman, H. C.** (2019). Local policymaker perceptions of the opioid crisis and the efficacy of extension communications. Funded by the North Central Regional Center for Rural Development ($20,500).

1. **Shulman, H. C.,** Boster, F. J., & Carpenter, C. (2010) Do data collection procedures influence political knowledge test performance? *Time-Sharing Experiments for the Social Sciences.*

Proposals Submitted

1. Coronel, J., **Shulman, H. C.**, & Bond, R. M. Determining comprehension of direct democracy initiatives and predicting large-scale voting decisions through facial expressions and eye movements. *National Science Foundation* ($205,186.00). Status: Not Funded

Invited Publications

4. Carpenter, C. J. & **Shulman, H. C.** (in press). Persuasion. In T. Reimer, L. Von Swol, and A. Florack (Eds.), *The Routledge handbook of communication and social cognition.* Routledge/Taylor and Francis.

3. Riggs, E. E.*, Coronel, J., & **Shulman, H. C.** (in press). Eye-tracking as a powerful tool for investigating language processing in messages. In T. Reimer, L. Von Swol, and A. Florack (Eds.), *The Routledge handbook of communication and social cognition.* Routledge/Taylor and Francis.

2. Bullock, O. M.* & **Shulman, H. C.** (2020). Framing. In J. Van den Bulck (Ed.), *The international encyclopedia of media psychology.* Hoboken, NJ: Wiley-Blackwell. https://doi.org/10.1002/9781119011071.iemp0268

1. **Shulman, H. C.** (2015). Rethinking the way we communicate about politics with millennials. In S. M. Chod, S. M. Caliendo & W. Muck (Eds.), *Technology and civic engagement in the college classroom: Engaging the unengaged.* New York: Palgrave Macmillan.

Under Review

4. Kalny, C.*, Walter, N., Lapinski, M. K., **Shulman, H. C.**, Demetriades, S. (under review). Descriptive norms ≠ injunctive norms? A meta-analytic review. *Journal of Communication*

3. Lerner, B.*, Hubner, A., **Shulman, H. C.** (under review). Who are the science populists in the United States and which experts do they trust? *PLOS One*

2. Bullock, O. M.*, **Shulman, H. C.,** Dixon, G. (revise & resubmit). Improving message engagement and persuasion for dissonant political ads: The role of processing fluency and motivated resistance to persuasion. *Communication Research Reports*

1. Riggs, E. E.*, **Shulman, H. C.,** Huskey, R., Lynch, T., & Fisher, J. T. (first revision under review). What can cognitive load and processing fluency tell us about difficult processing? *Media Psychology*

Awards and Fellowships

College-Level Distinction

- Recipient of the College of Arts and Sciences **Early-Career Faculty Excellence Award**, recognizing outstanding performance in all three areas of research, teaching, and service, Ohio State University (2022).

College-Level Grant Recipient

- Recipient of the **Arts and Sciences Curriculum Committee Service-Learning Grant**, College of Arts and Sciences, Ohio State University (2018-2019)

Research and Scholarly Awards

- Recipient of the **Michael Pfau Outstanding Article Award** for the article entitled "Predicting vote choice and election outcomes from ballot wording: The role of processing fluency in low information direct democracy elections." Political Communication Division, National Communication Association (Fall 2024)

- Recipient of a **Time-Sharing Experiments in the School of Communication (TESoC)** award for a project entitled, "Unpacking scientific expertise," with Blue Lerner and Austin Hubner (2024, $2,760)

- Recipient of the **Miller Award**, for a project entitled "Race, resignation, and activism," with Rachel B. Wade and Monique M. Turner (2023, $10,000)

- Recipient of the **Miller Award**, for a project entitled "An examination of how foreign languages are processed in narratives: An initial investigation", with Rachel Lopez, Kara Fort, Jorge Cruz Ibarra, and Elizabeth E. Riggs (2023, $2,080)

- Recipient of the **Top Article Award** for the manuscript entitled "Using metacognitive cues to amplify message content: A new direction in strategic communication." Communication and Social Cognition Division, National Communication Association (Fall 2021)

- Recipient of the **Miller Small Grant Program** award, for a project entitled "Examining the efficacy of emergency communication", with Olivia Bullock (2020, $3,024.00)

- Co-recipient of the **Miller Award,** for a project entitled "Using eye movements to determine when findings from the lab can be generalized to naturalistic settings: Linguistic features of messages and real-world voting behaviors", with Drs. Jason Coronel and Robert Bond (2018, $17, 288.87)

- Recipient of the **Faculty Professional Development Grant Award** (Summer 2012, 2014, North Central College)

- Recipient of the **Distinguished Article Award** for the manuscript entitled "Sender demeanor: Individual differences in sender believability have a powerful impact on deception detection judgments" Communication and Social Cognition Division, National Communication Association (Fall 2011)

Top Papers

- Recipient of a **Top Paper** award in the Communication and Social Cognition Division at the annual meeting of the National Communication Association (2024)

- Recipient of a **Top Paper** award in the Communication Science and Biology Division at the annual meeting of the International Communication Association (2023)

- Recipient of a **Top Paper** award in the Communication Science and Biology Division at the annual meeting of the International Communication Association (2022)

- Awarded **Top Three Paper** in the Communication and Social Cognition Division at the annual meeting of the National Communication Association (2021)

- Awarded **Top Three Paper** in the Communication and Social Cognition Division at the annual meeting of the National Communication Association (2018)

- Awarded **Top Five Paper** for the Intergroup Interest Group at the annual meeting of the International Communication Association (2013)

Invited Speaker

- Invited speaker at **West Virginia University** in the Department of Communication in Morgantown, WV, Autumn 2024

- Invited (virtual) speaker at the **Language, Public Engagement, and New (Quantum) Technology** research symposium at Vrije Universiteit Amsterdam, Autumn 2023

- Invited speaker for the **Political Communication Working Group** hosted by the Department of Communication and Media at the University of Michigan, Spring 2023

- Invited speaker at the **Midwest Association of Core Directors Annual Meeting** in Columbus, Ohio, Autumn 2022

- Invited speaker at **Michigan State University's Charles K. Atkin Distinguished Speaker Series** hosted by the Department of Communication, Spring 2022

Teaching Acknowledgements

- **Faculty Mentor of the Year** award, School of Communication at Ohio State University (2024)

- Nominee for the **Carmen Common Sense Award** for the course COMM 3620 – Introduction to Interpersonal Communication (AU19), an award granted to instructors with the best Carmen course page at Ohio State University (2019-2020)

Select Graduate School Honors

- Recipient of the **G.R. Miller Scholar Award**, awarded to the PhD student with the most promising research agenda, by faculty within the Department of Communication (2010, Michigan State University)

- Awarded **Top Three Paper** in the Group Communication Division at the annual meeting of the National Communication Association (2010)

- Awarded the University wide, **Excellence in Teaching Citation**, Office of the Provost, Michigan State University (Fall 2009)

- Awarded **Top Four Paper** in the Student Division at the annual meeting of the National Communication Association (2008)

Advising

Ohio State University

- **Ph.D. Advisor**
  Kara Fort (Co-Advisor; Communication) 2023 - present
  Blue Lerner (Communication) 2023 – present
  Rachel Barry Wade (Communication) 2022 – present
  Elizabeth E. Riggs (Communication) Ph.D. 2024
  Olivia M. Bullock (Communication) Ph.D. 2022

- **Committee Member**
  Samuel Bashian (Communication)
  Ji Youn (Jessica) Ryu (Communication)
  Jorge Cruz-Ibarra (Communication)
  Michael A. Gilbert (Communication)
  Nina Freiberger (Communication)
  Kristina Medero (Communication)
  Tim O'Neil (Communication)
  Erin Drouin (Communication)
  Shannon Poulsen (Communication)
  Victoria Abou-Ghalioum (Environment and Natural Resources)
  Matthew D. Sweitzer (Communication)
  Min Seon Jeong (Communication)
  David Clemenson (Communication)

- **M.A. Advisor**
  Emily Schutz (Communication)

- **Undergraduate Honor's Thesis Advisor**
  Travis Filiky (Majors: Communication, Political Science)

North Central College

- **Second Reader (Honor's Thesis Committee Member)**
  Aleksandra Ruseva (Major: Political Science) 2013 - 2014
  Christine Badowski (Major: Marketing) 2013 - 2014
  Brittnea Roozen (Major: Political Science) 2011- 2012

Select Media Coverage

- Thomas, B. (2024 October, 23). Does ballot language effect how you will vote. *55KRC THE Talk Station*.
- Markowitz, D., **Shulman, H. C.**, & Rogers, T. (2024 August, 20). Readers Prefer to Click on a Clear, Simple Headline – Like This One. *The Conversation.*
- Merrefield, C. (2024 June, 27). Readers of Online News Prefer Simple Headlines, Research Suggests. Journalists? Not So Much. *The Journalist's Resource*.
- Britt, R. R. (2024 June, 5). The Best Headlines: Short and Simple: Writers generally suck at headline writing, according to new research that reveals what works best. *Medium*.
- Grabmeier, J. (2024 June, 5). Simple Headlines Attract More Online News Readers: Study Examined Real-World Examples from Washington Post. *Ohio State News*.
- Caldwell, E. (2021 April, 14). Telling Sunbathers What They Don't Want to Hear: Tanning is Bad. *Ohio State News*.
- Kornei, K. (2021 April, 9). Are You Confused by Scientific Jargon? So are Scientists. *New York Times*
- LaPlante, M. (2020 April, 10). Undisciplined: The Devil's Jargon. *Utah Public Radio*.
- Woolston, C. (2020 February, 27). Words Matter: Jargon Alienates Readers. *Nature*.
- Feder, T. (202 March, 3). Speak Plainly to Attract People to STEM, Study Suggests. *Physics Today*.
- Yoder, K. (2020 February, 26). Want People to Care About Climate Change? Skip the Jargon. *Grist*.
- Todd, M. (2020 February, 24). Jargon May Be Even Worse for Communicating Complexity Than You Thought. *Social Science Space*.
- Off, C. (2020 February, 12). Scientists: Lose the Jargon, or You'll Lose Readers, Says Communications Prof. *CBC's As It Happens*.
- Grabmeier, J. (2020 February, 12). The Use of Jargon Kills People's Interest in Science and Politics. *Ohio State News*.
- Beck, L. (2013 January, 28). Like Children, Men Need to be Told to Wash Their Filthy Hands. *Jezebel*.

Teaching Experience

Ohio State University

COMM7998 – Directed Research

COMM7790 – Statistical Applications in Communication II
COMM4998 – Directed Undergraduate Research
COMM4635 – Communication Dynamics in Teams
COMM4337 – Public Communication Campaigns
COMM3624 – Communication in Personal Relationships
COMM3620 – Introduction to Interpersonal Communication
COMM2850 – Media and Citizenship
COMM2321 – Writing for Strategic Communication
COMM1100 – Communication and Society

North Central College

SPC392 – Introduction to Public Relations
SPC367 – Persuasion Theories
SPC295 – Research Practicum
SPC260 – Introduction to New Media
SPC230 – Business & Professional Communication
SPC214 – Group Processes
SPC200 – Interpersonal Communication
SPC100 – Introduction to Public Speaking
Veranda Course - Writing for Popular Culture
Veranda Course - Community Action in Theory and Practice

Michigan State University

- **Sole Instructor**
  COM 325 – Interpersonal Influence and Conflict
  COM 340 – Leadership and Group Communication
  COM 399 – Special Topics: Political Communication
  COM 100 – Human Communication and Public Speaking
- **Teaching Assistant**
  COM 200 – Methods of Communication Inquiry
  COM 475 – Communication Campaign Design & Analysis
  COM 100 – Human Communication and Public Speaking
  COM 875 – Communication Leadership Skills
  COM 325 – Interpersonal Influence and Conflict
  COM 340 – Leadership and Group Communication

Conference Presentations
***Indicates student as a co-author at the time of data collection**

68. Kalny, C.*, Walter, N., Lapinski, M. K., **Shulman, H. C.,** & Demetriades, S. (2025, June). *Normative influence, revisited: A meta-analytic assessment of perceived descriptive and injunctive norms in health.* Paper to be presented at the annual meeting of the International Communication Association in Denver, CO.

67. Hubner, A., Lerner, B.*, & **Shulman, H. C.** (2025, June). *Public understanding of scientists: A descriptive examination of how the public evaluates various sources of scientific information.* Paper to be presented at the annual meeting of the International Communication Association in Denver, CO.

66. Fort, K. S.*, Lopez, R., **Shulman, H. C.,** Riggs, E. E.*, Cruz Ibarra, J.* (2024, November). *The impacts of code-mixing in a cross-cultural narrative: How processing fluency impacts narrative engagement and racial attitudes.* Paper presented at the annual meeting of the National Communication Association in New Orleans, LA. **Top Paper Award in Communication and Social Cognition Division**

65. Kalny, C., Walter, N., Lapinski-LaFaive, M., **Shulman, H. C.,** & Demetriades, S. (2024, April 4-6). *Descriptive norms ≠ injunctive norms? A meta-analysis.* Kentucky Health Communication Conference (KCHC) 18th Biennial Meeting, Lexington, KY.

64. Wade, R. B.*, Ryu, J.*, **Shulman, H. C.,** & Hovick, S. (November, 2023). *Identifying the causes of uncertainty and uncertainty management: A metacognitive approach.* Paper presented at the annual meeting of the National Communication Association in National Harbor, MD.

63. Harvill, B.* & **Shulman, H. C.** (November, 2023). *A method to madness: Processing scientific method jargon to evaluate credibility.* Paper presented at the annual meeting of the National Communication Association in National Harbor, MD.

62. Riggs, E. E.*, Huskey, R., **Shulman, H. C.,** Lynch, T., Fisher, J., & Mutialu, S.* (May, 2023). *The impact of cognitive load on recognition for health narrative information.* Pre-registered report presented at the annual meeting of the International Communication Association in Toronto, CA. **Top Paper Award in Communication Science and Biology Division**

61. Bergan, D., Carnahan, D., & **Shulman, H. C.** (January, 2023). *Discounting constituent attitudes: Motivated reasoning, ambiguity, and policymaker perception of constituent characteristics.* Paper to be presented at the annual meeting of the Southern Political Science Association in St. Pete Beach, FL.

60. Barry, R. C.* & **Shulman, H. C.** (November, 2022). *An experimental test of the accessibility-applicability model using three different types of frames.* Paper accepted for presentation at the annual meeting of the National Communication Association in New Orleans, LA.
59. Riggs, E. E.*, Huskey, R., Bullock, O. M.*, & **Shulman. H. C.** (November, 2022). *The impact of load on message recognition.* Paper accepted for presentation at the annual meeting of the National Communication Association in New Orleans, LA.

58. Lopez, R.*, **Shulman, H. C.,** Riggs, E. E.*, & Barry, R. C.* (November, 2022). *An experimental comparison of equivalency, emphasis, and metacognitive frames.* Paper accepted for presentation at the annual meeting of the National Communication Association in New Orleans, LA.

57. **Shulman, H. C.**, Barry, R. C.*, Riggs, E. E.*, & Holt, L. F. (November, 2022). *The role of framing, race, and symbolic racism in policy support: The case of police reform.* Paper accepted for presentation at the annual meeting of the National Communication Association in New Orleans, LA.

56. Riggs, E. E.,* Lopez, R.,* Bullock, O. M.,* & **Shulman, H. C.** (May, 2022). *An examination of the measurement of accessibility: Is an easy experience always faster?* Paper presented at the annual meeting of the International Communication Association in Paris, France. **Top Paper Award in Communication Science and Biology Division**

55. **Shulman, H. C.,** Riggs, E. E.,* Lopez, R.,* Bullock, O. M.,* & Barry, R. C.* (May, 2022). *An experimental paradigm designed to explain (and synthesize) frames and framing effects: Some answers, more questions.* Paper presented at the annual meeting of the International Communication Association in Paris, France.

54. Riggs, E. E.,* **Shulman, H. C.**, & Lopez, R.* (November, 2021). *Using infographics to reduce the negative effects of jargon on predicting intentions to vaccinate against COVID-19.* Paper presented at the annual meeting of the National Communication Association in Seattle, WA. **Top Three Papers in Communication and Social Cognition Division**

53. **Shulman, H. C.**, Bullock, O. M.,* & Riggs, E. E.* (November, 2021). *The influence of jargon, motivation, and fatigue while processing information about COVID-19 over time.* Paper presented at the annual meeting of the National Communication Association in Seattle, WA.

52. Bullock, O. M.* & **Shulman, H. C.,** Dixon, G. N. (November, 2020). *Improving message engagement and persuasion for dissonant information: Processing fluency reduces motivated resistance to persuasion.* Paper presented virtually at the annual meeting of the National Communication Association in Indianapolis, IN.

51. **Shulman, H. C.,** Sweitzer, M. D.*, Bullock, O. M.*, Coronel, J., Bond, R. M., & Poulsen, S.* (May, 2020). *Explaining how people vote on ballot initiatives with language difficulty and metacognition: Results from two ecological experiments.* Paper presented virtually at the annual meeting of the International Communication Association in Gold Coast, Australia.

50. Bullock, O. M.,* **Shulman, H. C.,** & Huskey, R. (May, 2020). *Enhancing our understanding of when and why narrative persuasion is successful: A test of processing fluency and identification.* Paper presented virtually at the annual meeting of the International Communication Association in Gold Coast, Australia.

49. Poulsen, S.*, Coronel, J., Sweitzer, M. D.*, Bullock, O. M.*, **Shulman, H. C.,** & Bond, R. M. (May, 2020). *Thinking reflectively or intuitively: How cognitive reflection moderates the effect of language complexity on abstention.* Paper presented virtually at the annual meeting of the
International Communication Association in Gold Coast, Australia.

48. **Shulman, H. C.** (November, 2019). *Open science practices for communication research.* Panelist for the National Communication Association pre-conference entitled "Getting Results That Survive: Improving Communication Science" in Baltimore, MD.

47. **Shulman, H. C.,** Dixon, G. N., Bullock, O. M.*, Colón Amill, D.* (November, 2019). *The effects of jargon on processing fluency and self-perceptions: Strengthening framing theory with metacognition.* Paper presented at the annual meeting of the National Communication Association in Baltimore, MD.

46. Bullock, O. M.*, Colón Amill, D.*, **Shulman, H. C.,** & Dixon, G. N. (November, 2019). *Jargon as a barrier to effective science communication: Evidence from metacognition.* Paper presented at the annual meeting of the National Communication Association in Baltimore, MD.

45. **Shulman, H. C.,** & Bullock, O. M.* (May, 2019). *How the joint consideration of primary and
secondary cognitions in message design should improve the effectiveness of strategic messages.* Paper presented at the annual meeting of the International Communication Association in Washington DC.

44. Coronel, J., Bullock, O. M.*, **Shulman, H. C.**, Sweitzer, M. D.*, Bond, R. M., & Poulsen, S.* (May, 2019). *Using eye movements to determine when laboratory findings can be generalized to naturalistic settings: Linguistic features of messages and real-world voting behaviors.* Paper presented at the annual meeting of the International Communication Association in Washington DC.

43. **Shulman, H. C.,** & Bullock, O. M.* (November, 2018). *Pairing a gain-loss frame with a metacognitive frame to explain health and risk perceptions and the cognitive processes associated with framing effects.* Paper presented at the annual meeting of the National Communication Association in Salt Lake City, UT. **Top Three Papers in Communication and Social Cognition Division**

42. Bond, R. M., **Shulman, H. C.,** Gilbert, M.* (November, 2018). *Does having a political discussion help or hurt intergroup perceptions?: Drawing guidance from social identity theory*

and the contact hypothesis. Paper presented at the annual meeting of the National Communication Association in Salt Lake City, UT.

41. Sweitzer, M. D.,* & **Shulman, H. C.** (November, 2017). *Survey of surveys: A content analysis of the language complexity of public opinion polls*. Paper presented at the annual meeting of the National Communication Association in Dallas, TX.

40. **Shulman, H. C.** (November, 2016). *Applying metacognition to communication research: Improving political efficacy and interest through word choice*. Paper presented at the annual meeting of the National Communication Association in Philadelphia, PA.

39. **Shulman, H. C.** & Sweitzer, M. D.* (November, 2016). *Advancing framing theory: Using frames to improve public opinion via metacognition*. Paper presented at the annual meeting of the National Communication Association in Philadelphia, PA.

38. **Shulman, H. C.,** Rhodes, N., Davidson, E.*, Ralston, R.*, Borghetti, L.*, & Morr, L.* (November, 2016). *The state of the field of social norms research*. Paper presented at the annual meeting of the National Communication Association in Philadelphia, PA.

37. **Shulman, H. C.,** & Chod, S. M., (May, 2015). *A closer look at the relationship between institutions, political participation, and interpersonal political discussions*. Paper presented at the annual meeting of the International Communication Association in San Juan, Puerto Rico.

36. **Shulman, H. C.,** & Chod, S. M. (April, 2015). *The university structure and political networks: Clues into why college campuses affect political participation*. Paper presented at the annual meeting the annual meeting of the Midwest Political Science Association in Chicago, IL.

35. **Shulman, H. C.,** Bushman, K.*, Huizenga, E.*, Ward, M.*, & Wresinski, K.*(November, 2014). *Can group discussions be used to facilitate political interest and efficacy in college students?: A longitudinal study*. Paper presented at the annual meeting of the National Communication Association in Chicago, IL.

34. **Shulman, H. C.** Valleskey, K.*, Solus, J.*, & Bray, A.* (November, 2013). *Can public opinion survey wording affect internal political efficacy?: An experiment*. Poster presented at the annual meeting of the National Communication Association in Washington, D.C.

33. Roozen, B.* & **Shulman, H. C.** (June, 2013). *Tuning in to the RTLM: Tracking the evolution of language alongside the Rwandan Genocide using social identity theory*. Paper presented at the annual meeting of the International Communication Association in London, UK. **Top Five Paper in Intergroup Interest Group**

32. Boster, F. J., & **Shulman, H. C.** (June, 2013). *Political knowledge test performance as a function of venue, time pressure, and performance norms.* Paper presented at the annual meeting of the International Communication Association in London, UK.

31. **Shulman, H. C.,** & DeAndrea, D. C. (November, 2012). *Revisiting assumptions about political socialization processes in the family.* Paper presented at the annual meeting of the National Communication Association in Orlando, FL.

30. **Shulman, H. C.**, Boster, F. J., Carpenter, C., & Shaw, A. S. (November, 2011). *Why do students completing a political knowledge test score higher online than in the classroom? A series of studies.* Paper presented at the annual meeting of the National Communication Association in New Orleans, LA.

29. **Shulman, H. C.,** Carpenter, C., & Boster, F. J. (2011, April). *Do data collection procedures influence political knowledge test performance?* Paper presented at the annual meeting of the Midwest Political Science Association, Chicago, IL.

28. **Shulman, H. C.,** & Levine, T. R. (2010, November). *Exploring social norms as a group-level phenomenon: Do political participation norms exist and influence political participation on college campuses?* Paper presented at the annual meeting of the National Communication Association in San Francisco, CA.

27. **Shulman, H. C.,** Boster, F. J., Carpenter, C. (2010, November). *The effect of test-taking venue and test format on political knowledge test performance.* Paper presented at the annual meeting of the National Communication Association in San Francisco, CA.

26. **Shulman, H. C.,** & Wittenbaum, G. M. (2010, November). *Promoting perceived deliberative success through group communication.* Paper presented at the annual meeting of the National Communication Association in San Francisco, CA. **Top Three Paper in Group Communication Division**

25. Boster, F. J., Carpenter, C., **Shulman, H. C.**, DeAngelis, B., Manata, B., & Shaw, A. (2010, November). *In search of the elusive boomerang.* Paper presented at the annual meeting of the National Communication Association in San Francisco, CA.

24. Serota, K., Boster, F. J., Carpenter, C., & **Shulman, H. C.** (2010, November). *Political influentials: Validation of the political superdiffuser scale.* Paper presented at the annual meeting of the National Communication Association in San Francisco, CA.

23. Levine, T. R., **Shulman, H. C.**, Carpenter, C., DeAndrea, D. C., & Blair, P. (2010, November). *The impact of accusatory, non-accusatory, bait, and false evidence questioning in deception.* Paper presented at the annual meeting of the National Communication Association in San Francisco, CA.

22. Levine, T. R., Serota, K. , **Shulman, H. C.**, Clare, D., Park, H., Shaw, A., et al. (2010, November). *Sender demeanor: Does individual differences in sender believability have a power impact on deception detection judgments.* Paper presented at the annual meeting of the National Communication Association in San Francisco, CA.

21. Levine, T. R., Shaw, A. S., **Shulman, H. C.** (2010, June). *Increasing deception detection accuracy with strategic direct questioning.* Paper presented at the annual meeting of the International Communication Association in Singapore.

20. Levine, T. R., Serota, K. B., **Shulman, H. C. (**2010, June). *The impact of Lie to Me on viewer's actual ability to detect deception.* Paper presented at the annual meeting of the International Communication Association in Singapore.

19. **Shulman, H. C.,** & Wittenbaum, G. M. (2010, April). *Do beliefs about the political orientations of others affect political discussions?: Arguments for political labeling as an inhibitor of successful deliberation.* Panel accepted for presentation at the meeting of the Central States Communication Association in Cincinnati, OH.

18. **Shulman, H. C.,** & Neuberger, L. (2009, November). *Attitude constraint or successful framing?: Testing a political constraint measurement model and identifying its methodological implications.* Paper presented at the annual meeting of the National Communication Association in Chicago, IL.

17. Wittenbaum, G. M., & **Shulman, H. C.** (2009, November). *Social ostracism in task groups: The effects of group composition.* Paper presented at the annual meeting of the National Communication Association in Chicago, IL.

16. Bergan, D., Neuberger, L., **Shulman, H. C.**, & Risner, G. (2009, November). *Unenlightened self interest, knowledge and partisanship.* Paper presented at the annual meeting of the National Communication Association in Chicago, IL

15. Levine, T. R., Shaw, A. S., & **Shulman, H. C.** (2009, November). *Assessing deception detection accuracy with dichotomous truth-lie judgments and continuous scaling: Are people really more accurate when honesty is scaled?* Paper presented at the annual meeting of the National Communication Association in Chicago, IL

14. Maloney, E. K., Lapinski, M. K., Braz, M. E., & **Shulman, H. C**. (2009, November). *Modifying perceptions of descriptive norms and behaviors through messaging: A field experiment.* Paper presented at the annual meeting of the National Communication Association in Chicago, IL.

13. Wittenbaum, G. M., & **Shulman, H. C.** (2009, July). *Social ostracism in task groups.* Paper presented at the annual meeting of the International Network of Group Researchers in Colorado Springs, CO

12. **Shulman, H. C.,** Neuberger, L., & DeAndrea, D. (2009, May). *Old or experienced?: Exploring age related cognitions in US elections.* Paper presented at the annual meeting for the International Communication Association in Chicago, IL

11. Bergan, D., **Shulman, H. C.,** & Neuberger, L. (2009, May). *Frames v. evidence: Investigating the influence of question wording on policy support and political beliefs.* Paper presented at the annual meeting for the International Communication Association in Chicago, Illinois.

10. **Shulman, H. C.** (2008, November). *The role of integrative complexity and involvement on political message perceptions: How we make sense of political contradiction.* Paper presented at the annual meeting for the National Communication Association in San Diego, California. **Top Four Paper in Student Division**

9. Wittenbaum, G. M, **Shulman, H. C.,** Braz, M. E., & Skowronek, J. (2008, November). *Consequences of being ignored in group conversation.* Paper presented at the annual meeting for the National Communication Association in San Diego, California.

8. Lapinski, M. K., Maloney, E. K., Kim, S.-Y., Braz, M. E., **Shulman, H. C.**, & Klein, K. (2008, November). *Injunctive norms: Designing messages to modify perceived social sanctions.* Paper presented at the annual meeting of the National Communication Association in San
Diego, California.

7. Bergan, D. E., **Shulman, H. C.**, & Neuberger, L. (2008, August). *The effects of frames and evidence on political attitudes.* Paper presented at the annual meeting for the American Political Science Association in Boston, MA, August 2008.

6. Neuberger, L., **Shulman, H. C.,** & Maginnis, J. A. (2008, August). *Candidate websites and local newspaper coverage in the 2006 senatorial elections, implications for 2008 and beyond.* Paper presented at the political communication pre-conference for the American Political Science Association annual meeting in Cambridge, MA.

5. Walther, J. B., Van Der Heide, B., Hamel, L. M., **Shulman, H. C.** (2007, May). *Self-generated versus other-generated statements and impressions in computer-mediated communication: A test of warranting theory using Facebook.* Paper presented at the annual meeting of the International Communication Association in Montreal, Canada.

4. Lapinski, M.K., Rimal, R.N., Klein, K.A., **Shulman, H.C.,** & Hepler, A.L. (2006, November). *Perceived similarity, locus of control, and optimistic bias among people living with HIV/AIDS.* Paper presented at the annual meeting of the National Communication Association in San Antonio, TX.

3. Ellis, J.B., Van Der Heide, B., **Shulman, H**. (2006, November). *Can we quantify success?: A multilevel approach to assessment analysis in communication centers.* Paper presented at the annual meeting of the National Communication Association in San Antonio, TX.

2. **Shulman, H. C**., Ellis, J.B., & Van Der Heide, B. (2006, April). *The 7-minute communication workshop: An exercise in educational efficiency.* Paper presented at the annual meeting of the National Association of Communication Centers in Omaha, NE.

1. Van Der Heide, B., **Shulman, H.C**., & Ellis, J.B. (2006, April). *On developing a basis for ethical operations in communication centers.* Paper presented at the annual meeting of the National Association of Communication Centers in Omaha, NE.

Service

College of Arts and Sciences

Ohio State University

- 2023 – present — Member, Advisory Committee, ASC Office of Diversity, Equity, Inclusion, and Justice

Departmental Service

Ohio State University

- Spring 2025 — Chair, Undergraduate Journalism Program Committee
- 2023 – 2024 — Member, Undergraduate Journalism Program Committee
- 2022 – 2023 — Chair, Diversity, Equity, and Inclusion Committee
- 2017 – 2022 — Member, Undergraduate Journalism Program Committee
- 2021 – 2022 — Member, Search committee
- 2021 – 2022 — Assistant Professor Representative, Executive Committee
- 2019 – 2020 — Member, Search committee
- 2017 – 2018 — Assistant Professor Representative, Executive Committee
- 2017 – 2018 — Member, Search committee
- Spring 2017 — Member, Committee to assess stat training for grad program
- 2015 – 2016 — Guest speaker at PRSSA job training event

North Central College

- Winter 2014 — Worked with SPC and International Programs to create a course equivalency guide for study abroad programs
- Fall 2013 — Assisted with the SPC Department's Program Review
- 2012 – 2014 — Assessment Coordinator for the Department

Institutional Service

North Central College

- 2013 – 2014 Secretary, Academic Programs and Policies Committee
- 2013 – 2014 Search Committee Outside Member, Department of Psychology
- 2013 – 2016 Social Science representative, Research Ethics Committee
- 2013 – 2014 Faculty Advisor, NCC Ultimate Frisbee Club
- 2012 – 2013 Participant in a new faculty mentoring group to help acclimate first year faculty members.
- December 2011 Instructed a faculty workshop on incorporating social media in the classroom
- October 2011 Panelist on a faculty forum entitled "A Constitution in Crisis: The State of American Democracy"

Professional Service

National Communication Association

- 2024 – 2025 Vice chair, Communication and Social Cognition Division
- 2023 – 2024 Vice chair elect, Communication and Social Cognition Division
- 2022 – 2023 Research chair, Communication and Social Cognition Division
- 2021 – 2022 Research chair elect, Communication and Social Cognition Division

Editorial Board Member

- Journal of Language and Social Psychology (2021-present)
- Journal of Communication (2018-present)

Invited Reviewer

- Applied Cognitive Psychology
- Basic and Applied Social Psychology
- Communication Monographs
- Communication Quarterly
- Communication Research
- Communication Research Reports
- Communication Studies
- Communication Theory
- Frontiers in Psychology
- Health Communication
- Human Communication Research
- International Communication Association (2007 to present)
- International Journal of Communication
- International Network of Group Researchers
- Journal of Applied Social Psychology
- Journal of Communication
- Journal of Computer Mediated Communication
- Journal of Language and Social Psychology

- Journal of Media Psychology
- Journal of Public Deliberation
- Journal of Youth Studies
- Management Communication Quarterly
- Media and Communication
- Media Psychology
- National Communication Association (2007 to present)
- National Science Foundation
- Nature Human Behaviour
- Personal Relationships
- PLOS ONE
- Political Behavior
- Political Communication
- Public Opinion Quarterly
- Public Understanding of Science
- Science Communication
- Social Influence
- Social Science Computer Review
- Western Journal of Communication
- Zeitschrift für Psychologie

Grant Experience and Consulting

Reviewer

- National Science Foundation Proposal Reviewer, March 2017

Research Assistant

- Research assistant for a project with INgage media, the State of Michigan, and Michigan State University to develop a social networking site to improve energy efficiency practices in local units of government, Fall 2010 – Spring 2011
- Worked as a research assistant for a series of deception studies funded by the *National Science Foundation*, Fall 2008 – Spring 2009

Project Coordinator

- Kolt Communication and Lansing Regional Airport, Dec. 2007, 2008, 2009, 2010
- Atkin Group: Coordinator for public opinion polls measuring citizens' level of awareness and concerns with local issues (2007 – 2008)

Professional Associations

- National Communication Association, Life Member
- International Communication Association, Life Member

EXHIBIT C

Index

Research note

P U S

Public Understanding of Science
1–9

DOI: 10.1177/0963662519865687
journals.sagepub.com/home/pus
SAGE

Jargon as a barrier to effective science communication: Evidence from metacognition

Olivia M. Bullock, Daniel Colón Amill, Hillary C. Shulman and Graham N. Dixon
Ohio State University, USA

Abstract

In this experiment ($N = 650$), we examine the negative consequences of jargon on individuals' perceptions of emerging scientific technology and aim to explain these effects. We find that the presence of jargon impairs people's ability to process scientific information, and that this impairment leads to greater motivated resistance to persuasion, increased risk perceptions, and lower support for technology adoption. These findings suggest that the use of jargon undermines efforts to inform and persuade the public through the cognitive mechanism of metacognition.

Keywords

metacognition, persuasion, processing fluency, science communication

In order for the public to support scientific endeavors, it is important that research findings are effectively communicated to lay audiences. However, there is growing concern that scientific communities and the public may not be successful at engaging with one another, resulting in an uptick of practical and scholarly work aimed at clarifying science communication. From these works, a common recommendation is to reduce jargon (e.g. Baron, 2010; Dean, 2009; Sharon and Baram-Tsabari, 2014). Although this recommendation stems from a desire to "speak the same language" as the target audience, little research has examined the mechanisms that underlie this recommendation.

Thus, the purpose of this study is to investigate and explain the ramifications of jargon use in emerging science and technology contexts. Guided by research in metacognition (Petty et al., 2007; Schwarz, 2015), we demonstrate that jargon impairs people's ability to easily process scientific information, and that this impairment leads to greater motivated resistance to persuasion (MRP), increased risk perceptions, and lower support for technology adoption. Taken together, these theoretically guided findings offer practical implications for science

Corresponding author:
Olivia M. Bullock, School of Communication, Ohio State University, 154 N Oval Mall, Derby Hall, Columbus, OH 43210-1132, USA.
Email: bullock.181@osu.edu

communicators who aim to engage, inspire, and persuade the public of their important, yet often complicated, pursuits.

1. Jargon and processing fluency

Jargon refers to specialized, technical vocabulary terms that are associated with a situational context or purpose and are rarely used outside of these particular circumstances (Sharon and Baram-Tsabari, 2014). Jargon is often used to demonstrate expertise, convey idiosyncratic knowledge, or reference highly particularized ideas (Grupp and Heider, 1975). In addition to being technical, jargon is also used primarily by members of a particular group or trade, such as scientists, lawyers, or medical professionals, and is less frequently used or understood by individuals who fall outside of these groups (Sharon and Baram-Tsabari, 2014). Research examining the problematic impact of jargon (Grupp and Heider, 1975; Sharon and Baram-Tsabari, 2014) theorizes that negative effects are observed because non-experts are unable to fully comprehend jargon-laden information due to their lack of understanding. Here, we offer and test an additional explanation guided by metacognition and the feelings associated with information processing. We argue that in addition to jargon impairing people's ability to comprehend information, the presence of jargon may also affect the difficulty with which people process information. By understanding *why* message features produce undesirable outcomes, future efforts can utilize this information to reduce communication barriers to scientific engagement (e.g. Mellor, 2018).

Research from social psychology in metacognition theorizes about how one's subjective experience with information processing can affect judgments and decision-making (Petty et al., 2007; Schwarz, 2015). *Metacognition* can be defined as people's perceptions of, or experiences with, their own thought processes (Schwarz, 2010). The specific type of metacognitive experience studied here, called *processing fluency*, refers to the ease or difficulty with which new information is processed (Schwarz, 2010). Processing fluency is associated with feelings of ease, speed, and familiarity during information processing and is hedonically marked such that an easy processing experience is associated with positive feelings (Schwarz, 2006), while a difficult processing experience is associated with negative feelings (Schwarz, 2010). Here, we test whether the presence or absence of jargon produces variance in how easily people are able to process complex scientific information and whether this variance affects perceptions of new scientific technologies.

Prior research has found that language difficulty can influence processing fluency (Shulman and Sweitzer, 2018a, 2018b). Specifically, the use of more challenging words significantly impairs processing fluency relative to easier language. This study extends this work by testing this notion with jargon. Namely, we expect that scientific information that includes jargon should be more difficult to process than ordinary terminology. Moreover, if this difference is attributable to processing fluency, as opposed to comprehension alone, then this difference should persist even when jargon words are defined. If this is the case, then we expect the following result:

H1: Participants in the jargon condition will report lower levels of processing fluency than participants in the no-jargon frame condition, even when jargon definitions are provided.

2. Processing fluency and resistance to persuasion

The expected negative relationship between jargon and processing fluency suggests that *support*, an outcome pertinent to scientists, may be impacted by metacognition as well. Research suggests that processing fluency is hedonically marked such that when fluency is experienced as easy,

positive affective responses occur, such as feelings of knowing (Schwartz and Metcalfe, 1994), safety (Song and Schwarz, 2009), liking (Dragojevic and Giles, 2016), interest, and efficacy (Shulman and Sweitzer, 2018a, 2018b). These positive responses evoke the *naïve theory* (Schwarz, 2010) that if something feels good, it must be safe and familiar. Taken together, under conditions of easier processing, individuals are less motivated to seek out or consider additional information in order to be persuaded (Briñol et al., 2013). A difficult processing experience, however, is associated with unfamiliarity, which leads to negative outcomes such as uncertainty (Nelson et al., 1998), risk (Song and Schwarz, 2009), and a lack of confidence, liking, and knowledge perceptions (Shulman and Sweitzer, 2018a, 2018b). As such, metacognitive experiences of difficulty produce scrutiny (Briñol and Petty, 2004) as individuals feel a greater need to seek out more information in order to render a valid judgment (Briñol et al., 2013).

The skepticism and scrutiny associated with disfluent processing has implications for why individuals may resist scientific information. MRP refers to a person's motivation to oppose, or resist, perceived efforts to change existing attitudes (Nisbet et al., 2015). MRP is conceptualized as a combination of two experiences: (1) counterarguing, which reflects the generation of thoughts that undermine a message's persuasiveness and credibility, and (2) reactance, which refers to an oppositional response that arises from a message that is perceived to be threatening (Moyer-Gusé and Nabi, 2010). This experiment uniquely integrates processing fluency with MRP to consider whether the heightened scrutiny that extends from a disfluent experience will lead people to resist the scientific information presented. Guided by research in metacognition (e.g. Schwarz, 2010), we expect that participants will misattribute negative affect from their difficult processing experience toward the subject under investigation. If this is the case, participants should be more likely to discredit scientific information following a disfluent processing experience evoked by jargon. This leads to the second hypothesis:

H2: Processing fluency will mediate the relationship between exposure to jargon and motivated resistance to persuasion.

3. Risk perceptions and support

The notion that jargon will compel a difficult processing experience and increase MRP suggests that people's endorsement of the scientific technologies presented should be affected by these processes as well. When new technologies are introduced to the public, two outcomes become important for public acceptance: (1) the risk posed by these new technologies and (2) support for adopting these technologies. When people encounter something for the first time, a natural response is skepticism stemming from unfamiliarity (Song and Schwarz, 2009). Thus, scientists who need to communicate new findings must overcome a well-established cognitive obstacle—things that are new *feel* unsafe (Song and Schwarz, 2009). Here, we extend this idea to test whether communication strategies, such as the inclusion or exclusion of jargon and the processing fluency evoked by this manipulation, can improve or degrade people's responses to new information via MRP. If this is the case, then it stands to reason that variance in MRP should affect risk perceptions such that higher message resistance should lead to higher risk perceptions. This logic is reflected in the third hypothesis:

H3: Jargon will indirectly influence perceptions of risk through multiple mediators of processing fluency and motivated resistance to persuasion.

The second persuasion-related outcome, support for, or willingness to adopt, these technologies, should also be affected by participants' response to the scientific information presented. If the MRP scale functions as intended, then those who report higher scores on this scale should also be less likely to support, or adopt, the technologies in question. This claim contributes to prior research by stating that this relationship is expected based on the presence or absence of jargon and the subsequent information processing experience induced from this manipulation. This leads to our final hypothesis:

H4: Jargon condition will indirectly influence support through multiple mediators of processing fluency and motivated resistance to persuasion.

4. Method

Participants

Participants were recruited from Qualtrics' online general population panel in the United States ($N = 650$).[1] The sample was 62% female, and participants ranged in age from 18 to 80 ($M = 44.04$; $SD = 16.19$) years. The racial breakdown of the sample was 74.2% White; 12.6% African American or African; 7.1% Latino; 2.8% Asian; 1.8% American Indian or Alaska Native; 0.3% Native Hawaiian or Pacific Islander; and 0.9% mixed.

Procedure

Participants were randomly assigned to condition in a 2 (jargon vs no-jargon) × 2 (definitions vs no-definitions) between-subjects experimental design. All participants read three paragraphs about three different emerging scientific technologies: self-driving cars, surgical robots, and three-dimensional bioprinting. Three topics were chosen based on a message sampling approach, which ensures that findings are not unique to specific messages and are therefore more generalizable to other contexts (Jackson and Jacobs, 1983). For each of the three paragraphs, presentation order and condition assignment were held constant. Topic paragraphs were held on-screen for at least 4 seconds in an effort to ensure that individuals read the information presented. Processing fluency and risk were assessed after each message in order to capture participants' immediate information processing experience and risk perceptions. This sequence was repeated for the second and third topics. After exposure to all three paragraphs, participants responded to scales measuring MRP and support. The survey took about 20 minutes to complete ($M = 21.45$, $SD = 17.41$), and participants were paid through Qualtrics.

Stimuli

Before creating each experimental condition, information about the selected topics was obtained from credible science and technology sources (for details, see Supplementary Materials). This information was used to create three-sentence paragraphs about each scientific technology, where the first sentence provided context, the second described how it worked, and the third described possible risks (Supplementary Appendix A). In the jargon condition ($n = 328$), 10 jargon terms were included in each paragraph. In the no-jargon condition ($n = 312$), jargon was replaced by short explanations using simpler synonyms. Jargon was operationalized through terms that were technical or scientific, including descriptions of technologies, minerals, or

chemicals, as well as acronyms. Acronyms were replaced with their full form in the no-jargon condition.

To control for comprehension, participants were randomly assigned to a definitions condition ($n = 323$) or a no-definitions condition ($n = 317$). Definitions were provided using a mouseover text feature. In this condition, participants were told they could scroll over underlined terms (jargon) to receive their definition. The definition provided was identical to the language in the no-jargon condition. Word count was held constant across topic and condition.

Measures

All items were measured using seven-point Likert-type scales wherein higher scores reflect stronger agreement with the concept (full scales available in the Supplementary Materials).

Processing fluency. After exposure to each paragraph, participants responded to a five-item measure assessing *processing fluency* (Shulman and Sweitzer, 2018a, 2018b). The scale included items such as "A lot of the terms felt familiar to me." To account for fluency across topics, the five items were averaged across the three topics to form a 15-item scale, with higher scores reflective of an easier processing experience ($M = 4.92$, $SD = 1.07$, $\alpha = .90$).

MRP. MRP was measured using an eight-item scale (Nisbet et al., 2015). Items included "The scientific messages tried to pressure me to think a certain way" and "The scientific messages were not very credible" ($M = 2.96$, $SD = 0.95$, $\alpha = .84$).

Risk. Risk was measured following exposure to each topic paragraph. Three-scale items were presented after each topic for a total of nine measures ($M = 3.52$, $SD = 1.26$, $\alpha = .89$). An example item includes "[self-driving cars/surgical robots/3-D bioprinting] pose a serious threat to human safety" (Kahan et al., 2012).

Support. Support was measured using a 15-item scale that assessed support for adopting each technology. A sample item includes "Self-driving cars can solve transportation problems" ($M = 4.25$, $SD = 1.09$, $\alpha = .91$).

5. Results

Hypothesis 1 predicted that jargon condition assignment would affect reports of processing fluency independent of definition condition. To test this hypothesis, a two-way analysis of variance (ANOVA) was conducted. As predicted, there was a significant main effect for jargon, $F(1, 636) = 76.03$, $p < .001$, $\eta^2 = .11$, such that those in the jargon condition ($M = 4.57$, $SD = 1.11$) reported significantly lower processing fluency than those in the no-jargon condition ($M = 5.27$, $SD = 0.90$). In addition, consistent with expectations, there was not a significant main effect for definition condition, $F(1, 636) = 0.37$, $p = .543$, $\eta^2 = .0005$, nor a significant interaction effect, $F(1, 636) = 0.17$, $p = .678$, $\eta^2 = .0002$. Although the manipulations of jargon use and definitions appear to be operating independently, to isolate the effect of jargon, the definition condition was used as a covariate for all remaining analyses.

Hypothesis 2 predicted that processing fluency would mediate the relationship between jargon condition and MRP. This hypothesis was tested using the mediation model from Hayes' (2013) macro PROCESS (Model 4, 95% bias-corrected bootstrap confidence intervals (CIs) based on 10,000 resamples). As expected, significant indirect effects were obtained in the predicted

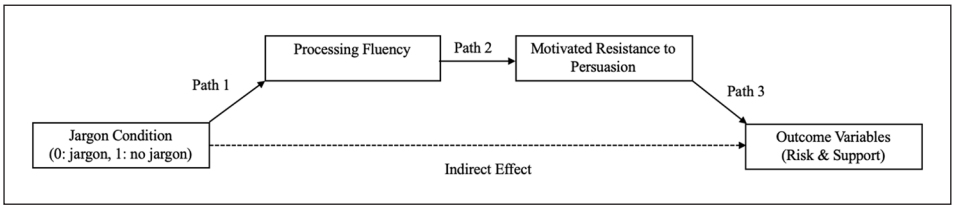


Figure 1. Model 6 from Hayes' (2013) PROCESS along with path labels that correspond with Table 1, wherein indirect effects are calculated as the product of paths 1, 2, and 3.

direction, $B = -.21$, $SE = .03$, 95% CI = [−.28, −.15], such that the no-jargon condition was associated with greater processing fluency, $B = .70$, $SE = .08$, $t = 8.71$, $p < .001$, which, in turn, reduced MRP, $B = -.29$, $SE = .04$, $t = 8.45$, $p < .001$. In total, this model explained 10% of the variance, indicative of a medium-to-large effect (Cohen, 1992). Thus, H2 was supported, even when controlling for the effect of definitions on MRP, $B = -.10$, $SE = .07$, $t = -1.41$, $p = .159$.

Hypothesis 3 predicted that the presence of jargon would indirectly influence perceptions of risk through the multiple mediators of processing fluency and MRP. This hypothesis was tested using Hayes' (2013) serial mediation model with two mediators, Model 6, 95% bias-corrected bootstrap CIs based on 10,000 resamples. Figure 1 represents this model and includes labels that correspond with each of the paths estimated in Table 1. In support of H3, the indirect effect was significant, $B = -.11$, $SE = .02$, 95% CI = [−.15, −.07], and explained 20% of the variance in risk, which is a large effect (Cohen, 1992). Once again, the covariate of definition condition never reached statistical significance ($-1.34 < t\text{'s} < -0.86$).

Finally, H4 predicted that the presence of jargon would indirectly influence support for emerging science technologies through the multiple mediators of processing fluency and MRP. The same serial mediation model from H3 (Figure 1) was used to test H4, with support as the outcome measure (see Table 1). As expected, the indirect effect of jargon on support through processing fluency and MRP was significant, $B = .08$, $SE = .02$, 95% CI = [.06, .12]. Unlike other models, definition condition was found to be a significant predictor of support, $B = .22$, $SE = .08$, $t = 2.79$, $p < .05$. Nevertheless, despite this finding, all relationships consistent with H4 were supported and explained 21% of the variance in technology support, which is a large effect (Cohen, 1992).

6. Discussion

This study examined the effect of jargon and processing fluency on individuals' resistance to persuasion, perceptions of risk, and willingness to support three different science technologies. Understanding how jargon impacts audiences has become particularly important amid concerns about a growing communication gap between scientific communities and the public. Here, we find support for the extant practical and scholarly recommendation that scientists reduce their jargon use but build on what is already known in several ways. First, we extend existing literature that recognizes how easy language can evoke engagement with science information (Scharrer et al., 2017) by offering processing fluency as another mechanism that explains these effects. Second, we believe that these findings generalize to other contexts where language difficulty has been found to alter judgments and decision-making, including politics and policy preferences (Carpenter and Boster, 2013; Goldberg and Carmichael, 2017; Sweitzer and Shulman, 2018).

We find that using jargon significantly disrupts processing fluency, in addition to and separate of comprehension. Furthermore, this reduction in processing fluency increases MRP,

Table 1. Results from the serial mediation analyses for hypotheses 3 and 4.

Outcomes	Path 1 *B* (*SE*)	Path 2 *B* (*SE*)	Path 3 *B* (*SE*)	R^2	Indirect effect *B* (*SE*)	95% CI [LL, UL]
H3						
Risk perceptions	.73 (.08)***	−.30 (.04)***	.51 (.05)***	.20	−.11 (.02)	[−.15, −.07]
H4						
Support	.70 (.08)***	−.31 (.04)***	−.39 (.04)***	.21	.08 (.02)	[.06, .12]

CI: confidence interval; LL: lower limit; UL: upper limit.
Path 1 denotes the path coefficient between the jargon condition (0: jargon, 1: no-jargon) and processing fluency. Path 2 denotes the relationship between processing fluency (higher scores = easier experience) and motivated resistance to persuasion. Path 3 indicates the relationship between motivated resistance to persuasion and outcomes (Figure 1). All models were run using Model 6 (Hayes' (2013) 95% bias-corrected bootstrap CIs based on 10,000 resamples), with definition condition as a covariate. Non-zero indirect effects indicate support for the serial mediation model hypothesized.
*$p < .05$; **$p < .01$; ***$p < .001$.

risk perceptions, and reduces overall support. Because science communication often serves to introduce scientific advancements to non-scientific audiences, these results suggest that initial messaging should strive to facilitate an easy processing experience and eliminate jargon where possible. In addition to this recommendation, the insight offered here extends to other communication techniques that also might impair processing fluency. This could include complex graphs, branding that includes acronyms, the offering of unintuitive data, or highly technical evidence, to provide just a few examples (see Shulman and Bullock, 2019). More broadly, we recommend that scholars not only consider information and comprehension in their communication to the public but also think about how message presentation may inadvertently impair information processing.

Despite these findings, there were methodological limitations of this study. First, we used an online experiment with a non-representative sample, thus limiting the generalizability of our findings. Second, these messages were free of any images, source cues, or context. The absence of these features hampers the ecological validity of our results, even though the information presented was obtained from real science communication sources. Finally, we asked participants to view three messages, rather than one. We chose to do this to increase the generalizability of our findings beyond any one science topic but recognize that these messages may have been differentially effective.

Theoretically, it is important to acknowledge that we did not directly measure comprehension. Our goal was to hold comprehension constant by including the same information across conditions. Furthermore, the manner in which we held information constant—through the use of mouseover text—introduced the behavior of searching for definitions if desired. Additional research should consider alternative strategies for capturing changes in comprehension without altering information presentation or adding a behavioral, and possibly affective, component.

Finally, we presented a serial mediation model despite using cross-sectional data. Because the measure of our dependent variables lacked a temporal element, we cannot be sure that we find causal effects between processing fluency, MRP, risk, and support. Nonetheless, we believe that the model presented has strong theoretical support and practical implications for science communicators.

In sum, this experiment provides evidence for the negative effects of jargon use on lay audiences. Our results imply that minimizing jargon within science communication should reduce resistance to persuasion and risk perceptions, and ultimately increase support. Future research

should explore the effects of jargon, or other forms of language that may affect processing fluency, with the hopes of ultimately enabling communicators to craft more effective appeals.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Olivia M. Bullock https://orcid.org/0000-0001-5403-7149

Supplemental material

Supplemental material for this article is available online.

Note

1. This dataset is used in another paper that also considers the effects of jargon on metacognition. However, that paper examines this topic with a different theoretical framework and outcome measures that are not reported here.

References

Baron N (2010) *Escape from the Ivory Tower: A Guide to Making Your Science Matter*. Washington, DC: Island Press.

Briñol P and Petty RE (2004) Self-validation processes: The role of thought confidence in persuasion. In: Haddock G and Maio G (eds) *Contemporary Perspectives on the Psychology of Attitudes*. London: Psychology Press, pp. 205–226.

Briñol P, Tormala Z and Petty R (2013) Ease and persuasion: Multiple processes, meanings, and effects. In: Unkelbach C and Greifeneder R (eds) *The Experience of Thinking: How the Fluency of Mental Processes Influences Cognition and Behavior*. London: Psychology Press, pp. 101–118.

Carpenter CJ and Boster FJ (2013) Modeling the effects of processing effort and ability in response to persuasive message arguments. *Communication Quarterly* 61: 413–430.

Cohen J (1992) A power primer. *Psychological Bulletin* 112: 155–159.

Dean C (2009) *Am I Making Myself Clear? A Scientist's Guide to Talking to the Public*. Cambridge, MA: Harvard University Press.

Dragojevic M and Giles H (2016) I don't like you because you're hard to understand: The role of processing fluency in the language attitudes process. *Human Communication Research* 42: 396–420.

Goldberg MH and Carmichael CL (2017) Language complexity, belief-consistency, and the evaluation of policies. *Comprehensive Results in Social Psychology* 2: 1–17.

Grupp G and Heider M (1975) Non-overlapping disciplinary vocabularies. In: Day SB (ed.) *Communication of Scientific Information*. Basel: Karger, pp. 28–36.

Hayes A (2013) *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. New York, NY: Guilford Press.

Jackson S and Jacobs S (1983) Generalizing about messages: Suggestions for design and analysis of experiments. *Human Communication Research* 9: 169–181.

Kahan D, Peters E, Wittlin M, Slovic P, Ouellette LL, Braman D, et al. (2012) The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nature Climate Change* 2: 732–735.

Mellor F (2018) Book Review: Kathleen Hall Jamieson, Dan Kahan and Dietram A. Scheufele (eds), The Oxford Handbook of The Science of Science Communication. *Public Understanding of Science* 27: 750–752.

Moyer-Gusé E and Nabi R (2010) Explaining the effects of narrative in an entertainment television program: Overcoming resistance to persuasion. *Human Communication Research* 36: 26–52.

Nelson TO, Kruglanski AW and Jost J (1998) Knowing thyself and others: Progress in metacognitive social psychology. In: Lories G and Dardenne B (eds) *Metacognition: Cognitive and Social Dimensions*. London: SAGE, pp. 69–89.

Nisbet EC, Cooper KE and Garrett RK (2015) The partisan brain: How dissonant science messages lead conservatives and liberals to (dis)trust science. *The ANNALS of the American Academy of Political and Social Science* 658: 36–66.

Petty R, Briñol P, Tormala Z and Wegener DT (2007) The role of metacognition in social judgments. In: Kruglanski AW and Higgins ET (eds) *Social Psychology: Handbook of Basic Principles*. New York, NY: Guilford Press, pp. 254–284.

Scharrer L, Rupieper Y, Stadtler M and Bromme R (2017) When science becomes too easy: Science popularization inclines laypeople to underrate their dependence on experts. *Public Understanding of Science* 26: 1003–1018.

Schwartz BL and Metcalfe J (1994) Methodological problems and pitfalls in the study of human metacognition. In: Shimamura AP (ed.) *Metacognition: Knowing about Knowing*. Cambridge, MA: MIT Press, pp. 137–156.

Schwarz N (2006) Feelings, fit, and funny effects: A situated cognition perspective. *Journal of Marketing Research* 43: 20–23.

Schwarz N (2010) Meaning in context: Metacognitive experiences. In: Mesquita B, Barrett LF and Smith ER (eds) *The Mind in Context*. New York, NY: Guilford Press, pp. 105–125.

Schwarz N (2015) Metacognition. In: Mikulincer M, Shaver PR, Borgida E and Bargh JA (eds) *APA Handbook of Personality and Social Psychology: Attitudes and Social Cognition*. Washington, DC: APA, pp. 203–229.

Sharon A and Baram-Tsabari A (2014) Measuring mumbo jumbo: A preliminary quantification of the use of jargon in science communication. *Public Understanding of Science* 23: 528–546.

Shulman HC and Bullock OM (2019) Using metacognitive cues to amplify message content: A new direction in strategic communication. *Annals of the International Communication Association* 43: 24–39.

Shulman HC and Sweitzer MD (2018a) Advancing framing theory: Designing an equivalency frame to improve political information processing. *Human Communication Research* 44: 155–175.

Shulman HC and Sweitzer MD (2018b) Varying metacognition through public opinion questions: How language can affect political engagement. *Journal of Language and Social Psychology* 37: 224–237.

Song H and Schwarz N (2009) If it's difficult to pronounce, it must be risky: Fluency, familiarity, and risk perception. *Psychological Science* 20: 135–138.

Sweitzer MD and Shulman HC (2018) The effects of metacognition in survey research: Experimental, cross-sectional, and content-analytic evidence. *Public Opinion Quarterly* 82: 745–768.

Author biographies

Olivia M. Bullock (BA, American University) is a PhD student in the School of Communication at the Ohio State University. Her research focuses on message design strategies that can reduce ideologically motivated processing of science, health, and political information.

Daniel Colón Amill (BA, University of Puerto Rico) is a Master's student in the School of Communication at the Ohio State University. His research interests include understanding how people process political information, particularly in online spaces.

Hillary C. Shulman (PhD, Michigan State University) is an assistant professor in the School of Communication at the Ohio State University. Her research interests include understanding how message design can improve information processing in the areas of politics, health, and science.

Graham Dixon (PhD, Cornell University) is an assistant professor in the School of Communication at the Ohio State University. His research interests include science and risk communication.

Research Article

Psychological Science
2021, Vol. 32(6) 836–848

DOI: 10.1177/0956797621991142
www.psychologicalscience.org/PS

Eye Movements Predict Large-Scale Voting Decisions

Jason C. Coronel, Olivia M. Bullock, Hillary C. Shulman, Matthew D. Sweitzer, Robert M. Bond, and Shannon Poulsen
School of Communication, The Ohio State University

Abstract

More than 100 countries allow people to vote directly on policies in direct democracy elections (e.g., 2016 Brexit referendum). Politicians are often responsible for writing ballot language, and voters frequently encounter ballot measures that are difficult to understand. We examined whether eye movements from a small group of individuals can predict the consequences of ballot language on large-scale voting decisions. Across two preregistered studies (Study 1: $N = 120$ registered voters, Study 2: $N = 120$ registered voters), we monitored laboratory participants' eye movements as they read real ballot measures. We found that eye-movement responses associated with difficulties in language comprehension predicted aggregate voting decisions to abstain from voting and vote against ballot measures in U.S. elections (total number of votes cast = 137,661,232). Eye movements predicted voting decisions beyond what was accounted for by widely used measures of language difficulty. This finding demonstrates a new way of linking eye movements to out-of-sample aggregate-level behaviors.

Keywords

eye movements, decision making, voting, direct democracy, open data, open materials, preregistered

Received 1/9/20; Revision accepted 10/20/20

On Election Day, millions of voters make important policy decisions on a wide range of issues, including repealing the death penalty, setting tax rates, and legalizing marijuana, by voting directly on ballot initiatives and statewide referendums. Such direct democracy elections have become more common around the world as more than 100 countries allow people to vote directly on laws and policies in their communities (Kaufmann & Mathews, 2018). The 2014 Scottish independence and 2016 Brexit referendums are prominent recent examples of direct democracy elections.

However, there is growing concern among social scientists and the general public that voters often encounter ballot measures that use language that is difficult to understand (e.g., legalistic or unfamiliar words), which can influence people's voting decisions (Quesenbery & Chisnell, 2016; Reilly, 2010; Reilly & Richey, 2011; Shockley & Fairdosi, 2015). The question of whether ballot language influences voting decisions has important implications for democratic societies. Politicians and special-interest groups are often responsible for writing ballot language (Reilly, 2010) and may unintentionally or deliberately influence election outcomes. Particularly concerning is the possibility that strategically minded political actors can craft language to influence the outcome of an election by either obscuring issues or causing certain groups to abstain from voting.

Thus, it is important to examine the consequences of ballot language on voting decisions and identify the psychological mechanisms underlying its effects. To this end, our study makes two key contributions. First, we used eye movements to measure the difficulties in comprehension that people experience while reading ballot measures. Eye-movement measures are useful because they can provide a moment-by-moment record of the comprehension challenges that people experience

Corresponding Author:
Jason C. Coronel, The Ohio State University, School of Communication
E-mail: coronel.4@osu.edu

while reading text. For example, a vast literature on eye movements and reading suggests that people are more likely to look longer at, or direct their gaze toward, words that they are unfamiliar with or are difficult to understand (Hyönä & Olson, 1995; Rayner, 1998, 2009; Rayner et al., 1989, 2006). Surprisingly, no study has used eye movements to examine the influence of ballot language on voting decisions.

Second, we devised an empirical strategy for linking eye movements in response to ballot measures obtained from a small group of individuals to aggregate voting behaviors of large groups of people. Social scientists frequently study the psychological processes underlying voting decisions using a small group of individuals (often in the context of the lab; Lau & Redlawsk, 2006; Lodge et al., 1995; Shockley & Fairdosi, 2015), and it is important to determine the extent to which phenomena observed from such contexts can be generalized to much larger groups of voters in naturalistic settings.

Therefore, the central and novel question we asked in our study is this: Can the eye movements made by a small group of individuals as they read ballot measures predict the voting decisions of a separate and much larger group (i.e., millions) of voters during actual elections? The finding that eye-movement responses from a small group of individuals predict aggregate-level behaviors advances work in eye movements and political psychology. It is currently unknown whether eye movements from a small group can forecast decisions aggregated at the level of societal units (e.g., states, countries). This is unknown because work in psychology has primarily focused on whether eye movements can predict decisions (e.g., economic and moral choices) within the same individuals (Krajbich et al., 2010; Pärnamets et al., 2015).

Additionally, there is growing recognition that the generalizability of psychological processes and behaviors is moderated by context and individual differences (Cartwright & Hardie, 2012; Henrich et al., 2010). Thus, it is important to know the conditions under which findings from one group of individuals generalize across other groups and settings. There is evidence that eye-movement responses to linguistic features (e.g., word frequency) generalize across individuals and different languages (Li et al., 2014; Tiffin-Richards & Schroeder, 2015; Whitford & Titone, 2017). This suggests that language-comprehension processes indexed by eye movements for one group of individuals can be extrapolated to a different group of individuals.

Finally, because eye movements provide a continuous record of reading performance, they can potentially reveal whether the challenges in understanding ballot language occur at the level of words, phrases, sentences, paragraphs, or the entire text. The ability of eye movements to provide information at different levels is unique and difficult to obtain using other measures of language difficulty. Ultimately, the information provided by eye movements may aid researchers and policymakers in crafting ballot language that is comprehensible to a large group of voters.

Statement of Relevance

More than half of the world's nations employ direct democracy elections, in which policy choices are made directly by the public. Using eye-tracking technology, we found that as ballot language becomes more difficult to understand, voters are more likely to abstain from voting or vote against ballot measures. These findings expose the concerns of direct democracy elections because politicians and special-interest groups may inadvertently or deliberately influence election outcomes by crafting difficult-to-understand ballot language. However, our study also lays the groundwork for how these concerns can be addressed through the use of eye-movement monitoring. Because eye movements provide a continuous measure of reading performance, they can potentially reveal whether the challenges in understanding ballot language occur at the level of specific words, sentences, or the entire text. Eye movements may be able to assist researchers and policymakers in crafting ballot language that is comprehensible to a larger group of voters.

Our study takes the first critical steps toward this long-term goal. We expected language that is difficult to understand would influence people's voting decisions in two ways. First, ballot measures that are difficult to comprehend might increase rates of abstention (Reilly, 2010; Reilly & Richey, 2011). This is because voters are unable to translate how ballot measures relate to their own political preferences and, thus, decide not to cast a vote.

Second, ballot measures that are hard to understand could lead voters to vote against the proposed policies. This prediction is based on the notion that voters have a general aversion to risk and uncertainty (Bowler & Donovan, 1998). Voters may feel uncertain about ballot measures they do not understand because they are unable to ascertain the potential consequences of the proposed policy. In contrast, voters will likely know more about the status quo, or the current state of affairs that may change by passage of a proposed law (Burnett, 2019; Lupia, 1992). For a voter who does not understand a proposed policy, changing the status quo can

be perceived as involving greater risk and uncertainty than maintenance of the status quo. As a consequence, voters may be more likely to prefer the status quo as ballot measures become harder to understand. Given how ballot choices are often structured (i.e., a "no" vote corresponds to not implementing a policy), a vote against a proposed policy is effectively a vote for maintaining the status quo.

We predicted that as ballot measures became more difficult to understand, as indicated by eye-movement responses from a group of voters, the rate of aggregate decisions to (a) abstain from voting or (b) vote against the measure in actual elections would also increase. Before data collection, we preregistered our hypotheses, research design, and analysis plan (https://osf.io/hdc7x and https://osf.io/srxyu). Across two studies, we recruited participants to come into the laboratory and read a set of real ballot measures while their eye movements were tracked. Then, we collected voting data elicited by these ballot measures (i.e., rates of abstention, rates of support and opposition) in actual elections. Our critical analyses involved whether eye-movement responses to ballot measures in the lab predicted aggregate voting decisions in actual elections.

Method

Participants

We analyzed data from 120 registered voters from the state of Ohio for Study 1 (60 women; age: $M = 34.99$ years, $SD = 16.19$, range = 18–79; race: White = 97, Black = 9, Latinx/Hispanic = 3, Asian = 4, mixed = 6, other = 1; partisan affiliation: Democrat = 71, independent = 37, Republican = 12) and another 120 registered voters from Ohio for Study 2 (60 women; age: $M = 33.98$ years, $SD = 18$, range = 18–73; race: White = 103, Black = 9, Asian = 4, mixed = 3, other = 1; partisan affiliation: Democrat = 61, independent = 29, Republican = 30; for recruitment information, see https://osf.io/65gjf/).

We collected data from July 17, 2018, to November 3, 2018, prior to the U.S. midterm elections on November 6, 2018. We identified our target population as voters in the United States. We therefore checked voter-registration files to ensure that participants who took part in our study were registered voters in the state of Ohio. This increased the likelihood that our sample consisted of individuals who have voted or will vote in elections.

Materials

For this study, we examined the effects of ballot language on voting decisions at the level of words. We used real ballot measures that appeared in U.S. elections as our stimuli. We selected measures that varied in the number of familiar and unfamiliar words they used (see Tables S30 and S31 at https://osf.io/65gjf/) given that the presence of unfamiliar words (e.g., ad valorem taxes) is one feature that could make ballot language difficult to understand (Quesenbery & Chisnell, 2016; Reilly, 2010; Shockley & Fairdosi, 2015).

We estimated the word frequency of each word for a given ballot measure using the SUBTLEXUS corpus (Brysbaert & New, 2009). The SUBTLEXUS corpus comprises words from subtitles in films and television series in the United States and has been shown to be a valid estimate of everyday language exposure (Brysbaert & New, 2009). Words that appear more frequently in the English language are more likely to be familiar to most people than low-frequency words (Rayner, 1998). We calculated the median word frequency for each ballot measure and selected ballot measures that were high (which should be relatively easy to understand) or low (which should be relatively hard to understand) in median word frequency (see https://osf.io/65gjf/).

It was necessary for us to use real ballot measures because we sought to examine whether eye-movement responses to the ballot measures in the lab predict aggregate voting decisions in elections. However, the trade-off with using real ballot measures is that we had less control over their characteristics, raising the possibility of confounding factors. We used two approaches in our research design to address this issue.

First, we intentionally sampled ballot measures that satisfied specific criteria to ensure that certain factors were not confounded with the frequency of unfamiliar words across the ballot measures (see https://osf.io/65gjf/). For example, we selected ballot measures about which voters would likely possess low levels of familiarity and that were generally nonpartisan. Specifically, none of the ballot measures covered issues such as abortion, the death penalty, the legalization of marijuana, or gun control. Further, no expenditures for campaign advertisements had been made on any of the ballot measures at the time they were selected. We employed this selection rule to increase the likelihood that voters in both the lab and actual elections had little knowledge of the ballot measures. This reflects real-world situations because voters are often unfamiliar with the ballot measures they encounter (Barth et al., 2020). In addition, we selected ballot measures that were not from the state of Ohio to increase the likelihood that the lab participants were unfamiliar with them.

Second, in our empirical analysis, we employed covariate adjustment in our regression analyses to account for other potential confounds (see https://osf.io/65gjf/).

We preregistered several covariates that included ballot-measure properties such as number of words and individual differences in our lab participants (e.g., age, level of political knowledge). In addition, we had a separate group of participants rate the ballot measures on the extent to which they perceived them as important, familiar, and interesting. We used these preregistered normative ratings as covariates to account for differences across the policy issues covered by the ballot measures.

The resulting 64 ballot measures we used (Study 1 = 40, Study 2 = 24) generally covered political issues often encountered by voters during the 2012, 2013, 2014, and 2018 U.S. elections (505 ballot measures; see https://osf.io/65gjf/). Specifically, the four most common issues during this time period appeared on approximately 53% of all ballot measures and consisted of issues pertaining to taxation, state and local government, infrastructure projects, and state budgets. These issues are important because they involve, for example, allowing people to determine how public education is financed, whether major infrastructure projects (e.g., public transport, waterworks) are carried out, and what powers are given to state governments. These issues were also common in our stimuli: 85% of ballot measures in Study 1 and 62.5% of ballot measures in Study 2 pertained to these issues. The percentage of these issues in our stimuli was higher than in the full set of ballot measures likely because of our selection procedure. Finally, some of the high-salience issues that we intentionally excluded from our stimuli (e.g., abortion, immigration) formed a small minority of all the ballot measures (see https://osf.io/65gjf/).

Procedure

For both Study 1 and Study 2, we tested participants individually in a quiet room, where they were seated 100 cm away from a computer monitor (resolution = 1,920 × 1,080 pixels; refresh rate = 60 Hz). Before the start of the experiment, we used a desktop-mounted EyeLink 1000 eye tracker (SR Research, Kanata, Ontario, Canada) that was fitted and calibrated for each participant with a 9-point calibration system. We employed a rigid mount to keep the chin and forehead from moving. Recordings were taken from the right eye, except for instances in which reflection off the participant's glasses or contact lenses necessitated recording from the left eye.

We informed participants at the start of the study that they would be reading about real ballot measures in Ohio. We instructed them to imagine that they were in the voting booth, to read each ballot measure carefully, and to vote on it. Each trial began with a drift-check target in the form of a dot in the middle of the screen. Participants controlled the time spent on this screen by fixating on the dot while pressing the advance button on the left side of the handheld controller. Participants were then presented with the proposed ballot measure. Participants controlled the time spent on this screen and could advance to the next part of the trial by pressing the advance button. Participants were then instructed to report, via a button press, whether they supported or opposed the proposed law or whether they would like to abstain from voting. The location of the text indicating "support," "oppose," or "abstain" on the computer screen was counterbalanced across participants. After the participant made a voting decision (i.e., pressed a button), the participant advanced to the next trial. We randomized the presentation order of the trials.

Postelection design of Study 1. The two studies differed in important ways. In Study 1, we used 40 ballot measures that appeared in the 2012 and 2014 elections spanning 21 states in the United States (total votes cast = 63,211,324; see Table S30 at https://osf.io/65gjf/). An important feature of Study 1 is that data in the lab were collected after the ballot measures already appeared in actual elections. The advantage of using ballot measures from previous elections was that information we used in our selection criteria could not change. For example, no expenditures for campaign advertisements could be made for selected ballot measures during the course of the study because the elections were over.

Preelection design of Study 2. In Study 2, we selected 24 measures that appeared in the 2018 midterm elections in 11 states (total votes cast = 74,449,908; see Table S31 at https://osf.io/65gjf/). Importantly, we collected data from the lab before the ballot measures were voted on in the 2018 midterm elections. The advantage of Study 2 was that participants could not be influenced by knowledge of the ballot measures' election outcomes because the outcomes were not yet known. However, a limitation of Study 2 was that information we used in our selection criteria could change prior to the election. For example, although no expenditures had been made for any of the ballot measures at the time they were selected (months prior to the election), funds for campaign advertising were allocated for several of them over the course of lab data collection. In addition, local media focused extensively on some ballot measures, whereas additional text was added to others between the time when we selected the stimuli and the time when they appeared in actual elections.

We did not foresee these circumstances prior to writing our preregistration protocol. To account for these unexpected issues and the possibility of omitted-variable

bias, we conducted statistical analyses that used covariate adjustment (e.g., using covariates for expenditure, number of newspaper editorials, additional text) in addition to our preregistered analyses (see https://osf.io/65gjf/).

Eye-movement measures. Our key independent variables were six distinct eye-movement measures. We used multiple measures for two reasons. First, the six measures index different processes involved in text comprehension. Second, we examined whether our results were robust and reliable, displaying consistent patterns in the direction of the associations between multiple eye-movement metrics and aggregate voting decisions.

Our eye-movement measures consisted of different types of fixations and fixation durations. While reading a passage of text, people's eyes make a series of rapid ballistic jumps separated by discrete pauses. The pauses are called *fixations*, and one of their functions is to place information in the environment, such as a word, within the part of the eye called the fovea, where visual acuity is the highest (Rayner, 1998). Fixation duration corresponds to the amount of time that the fovea is directed at a specific location in the visual environment.

The six measures can be categorized as early- and late-stage measures given that they index different processes in text comprehension. Early-stage measures are thought to reflect initial processing of word information, such as accessing the meaning of the word in long-term memory (Rayner, 1998, 2009). Early-stage measures were *first fixation duration*, *first-pass fixations*, and *first-pass fixation duration* (see https://osf.io/65gjf/).

In contrast, late-stage measures are thought to reflect higher order processes such as integrating the meaning of a word to the sentence context (Rayner, 1998, 2009). Late-stage measures were *regression fixations*, *total fixations*, and *total fixation duration* (see https://osf.io/65gjf/).

Analytic strategy

Critically, previous work has shown that an increase in the number of fixations or fixation durations for both early- and late-stage measures is associated with greater levels of difficulty in text comprehension (Rayner, 1998, 2009; Rayner et al., 1989, 2006). For each ballot measure, we calculated the average number of fixations or fixation durations elicited by each word that composed the ballot measure across the six eye-movement metrics. If eye-movement responses predict aggregate voting decisions, then we would expect that as the average number of fixations or fixation durations increased for each of the six eye-movement measures (indicating greater difficulties in real-time text comprehension), the rate of aggregate decisions to (a) abstain from voting or (b) vote against the ballot measure in actual elections would also increase.

In our analyses, we estimated separate linear regression models for each eye-movement measure; robust standard errors were clustered on the participants. For the analyses involving abstention rates, we used each of the eye-movement measures as the independent variable and the natural log of the proportion of abstentions during the actual election as the dependent variable. Following prior work, we used the natural log of the abstention rate, given that its distribution is skewed (Reilly & Richey, 2011). For analyses involving the opposition rate, we also used each of the eye-movement measures as the independent variable and the proportion of votes against the measure in the actual election as the dependent variable (see https://osf.io/65gjf/). Evidence consistent with our hypotheses would be positive coefficient estimates for the eye-movement measures in both the abstention- and opposition-rate analyses.

Results

Study 1

In Study 1, we first examined whether eye movements were associated with the rate of abstention for the ballot measures in actual elections. Our first set of analyses used our preregistered covariates. Figure 1a presents the relationship between eye movements and abstention rates for all six eye-movement measures. Across all six eye-movement measures, as predicted, an increase in the average number of fixations or fixation durations (in milliseconds) was associated with a positive and statistically significant increase in the election-abstention rate (first fixation duration: b = 0.00071, SE = 0.00016, $p < .001$; first-pass fixations: b = 0.13, SE = 0.031, $p < .001$; first-pass fixation duration: b = 0.00074, SE = 0.00013, $p < .001$; regression fixations: b = 0.043, SE = 0.0089, $p < .001$; total fixations: b = 0.037, SE = 0.0070, $p < .001$; total fixation duration: b = 0.00018, SE = 0.000033, $p < .001$; see Fig. 1a and Table S1 at https://osf.io/65gjf/).

Next, we examined the extent to which eye-movement measures were associated with an increased rate of opposition toward the ballot measures in actual elections. As expected, and as seen in Figure 1b, the coefficients for both early- and late-stage eye-movement measures were consistently positive. The associations appeared to be stronger for the early-stage measures; increases in the average first fixation duration (b = 0.00013, SE = 0.000061, p = .03) and average first-pass fixations (b = 0.033, SE = 0.012, p = .007) were associated with a positive and statistically significant increase

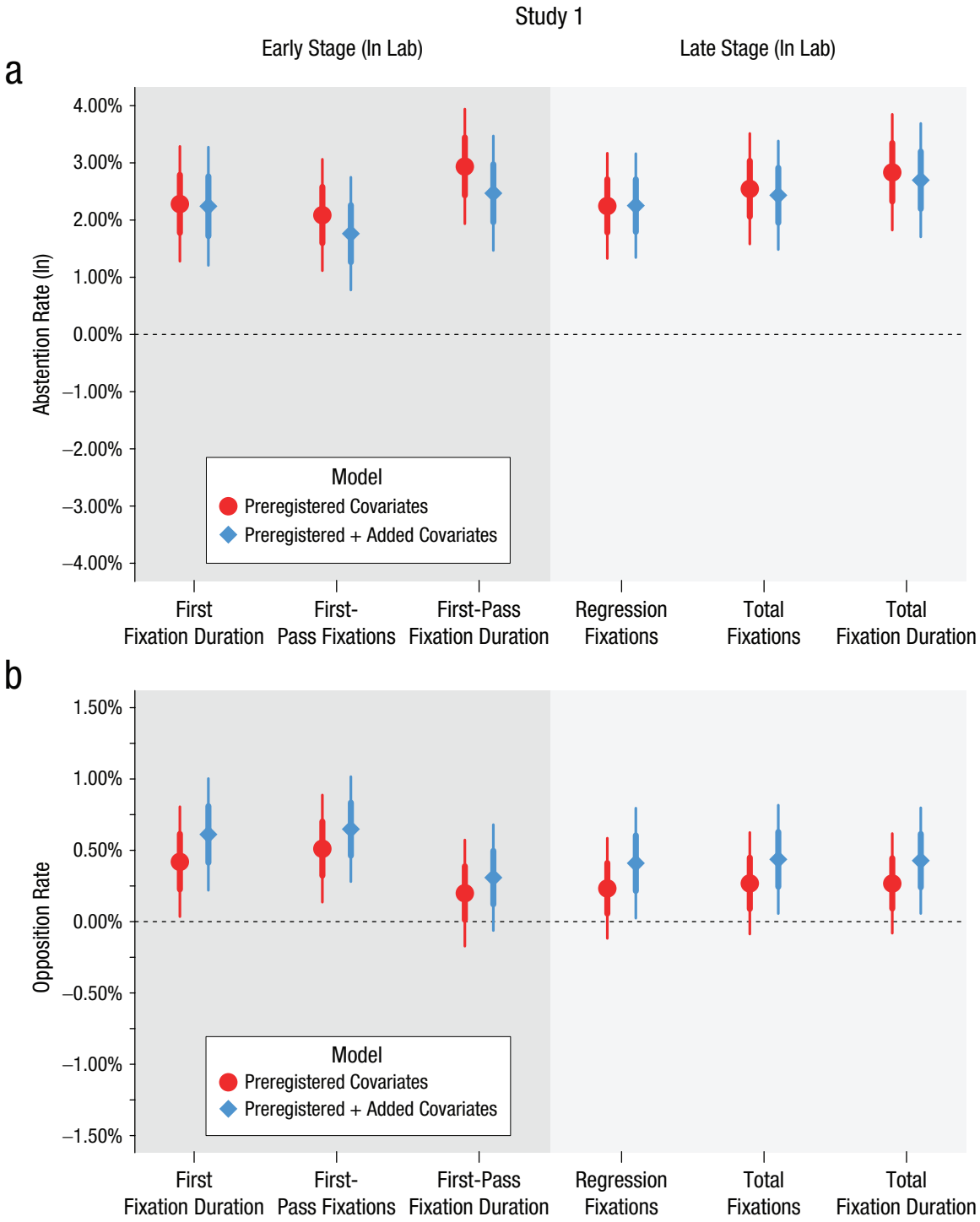


Fig. 1. Effect of a 1-*SD* increase on aggregate voting decisions in actual elections (Study 1; lab data were collected after actual voting occurred). The graphs show (a) the average abstention rate and (b) the average opposition rate for both the preregistered and alternative models, separately for each of the six-eye movement measures across the early and late stages. Point estimates are shown for both the preregistered and alternative models. Thicker lines represent ±1 cluster-robust standard errors, and thinner lines are 95% confidence intervals.

in the election-opposition rate (see Table S2 at https://osf.io/65gjf/). Although positive, the coefficients for the remaining cases did not reach conventional levels of statistical significance (all *p*s > .10).

We conducted an additional set of analyses that allowed us to examine the robustness of our results to alternative-model specifications. First, three of the ballot measures included additional text in the form of a

ballot explainer or fiscal-impact statement. Voters in the actual elections were exposed to this additional text, but this was not shown to lab participants. Second, in our preregistration plan, we had no analytical procedure to account for instances in which voters in actual elections could learn about the ballot measures beyond campaign advertisements. Newspaper coverage has been shown to be an important source of information from which voters can learn about ballot measures (Nicholson, 2003). To account for these two issues in our analyses, we included a dummy variable that indicated whether additional text had been added to a ballot measure in the actual election and another variable that indicated the total number of newspaper editorials written about each ballot measure.

As seen in Figure 1, inclusion of these two additional variables did not change our substantive results. In terms of abstentions, the coefficients for all six eye-movement measures remained positive and statistically significant (ps < .001; Fig. 1a; see also Table S3 at https://osf.io/65gjf/). In terms of the opposition rate, the coefficients for all six eye-movement measures remained positive and five were statistically significant (ps < .05; Fig. 1b; see also Table S4 at https://osf.io/65gjf/).

We also estimated bivariate models in which each model contained only one independent variable from each of the six eye-movement measures. The results of the bivariate models were consistent overall with the results of the multivariate models (see Tables S24 and S25 at https://osf.io/65gjf/).

Finally, we examined the size of the effect of language-comprehension difficulties (as measured by eye movements) on aggregate voting decisions. To do so, we examined the effect of a 1-*SD* increase of the independent variable (i.e., each of the eye-movement measures) on the dependent variable (i.e., aggregate voting decisions). We used a 1-*SD* increase because it represents a plausible counterfactual shift in the independent variable.

As can be seen in Figure 1, the effect sizes were small. For example, in the preregistered analyses, a 1-*SD* (156.50 ms) increase in average total fixation duration was associated with a 0.38% increase in the rate of abstention (95% confidence interval [CI] = [0.23%, 0.54%]).[1] The average total fixation duration was 285.30 ms. For the opposition analyses, a 1-*SD* (31.79 ms) increase in average first fixation duration was associated with a 0.42% increase in the rate of opposition (95% CI = [0.035%, 0.81%]). The average first fixation duration was 141.50 ms.

Although these effects are modest, it is important to note that even small effects can influence electoral outcomes. In competitive elections, for example, ballot measures can win by a razor-thin margin (see the Discussion section).

In summary, the results for Study 1 show evidence consistent with the hypotheses. Specifically, as the average number of fixations or fixation durations increased for each of the six eye-movement measures, the rate of aggregate decisions in actual elections to either abstain from voting or to vote against the ballot measure also increased.

Study 2

In Study 2, we first estimated models using the set of covariates that were preregistered. We also conducted additional analyses to account for unexpected issues that we encountered given our research design for Study 2. The Supplemental Material (https://osf.io/65gjf/) includes a full accounting of the five unexpected issues. In the alternative-model specifications, we added five covariates to account for the possibility of omitted-variable bias in our regression analyses.

In terms of abstentions, although coefficients for all six eye-movement measures were positive for the preregistered analysis, none reached conventional levels of statistical significance (see Table S5 at https://osf.io/65gjf/). In terms of the opposition analysis, the early-stage measures (similar to the findings of Study 1) showed the most robust associations; all three were positive and statistically significant (first fixation duration: b = 0.00018, SE = 0.000035, p < .001; first-pass fixation: b = 0.057, SE = 0.0087, p < .001; first-pass fixation duration: b = 0.00019, SE = 0.000032, p < .001; see Table S6 at https://osf.io/65gjf/).

Next, we estimated models that accounted for the unexpected issues we encountered in Study 2. For the abstention analysis, the coefficients for all six eye-movement measures were positive, as predicted. This consistent pattern can be observed in the alternative models in Figure 2a. The late-stage measures had the strongest effects: Increases in regression fixations (b = 0.031, SE = 0.0096, p = .002), total fixations (b = 0.020, SE = 0.0066, p = .003), and total fixation duration (b = 0.000083, SE = 0.000027, p = .003) were associated with positive and statistically significant increases in the abstention rate (see Table S7 at https://osf.io/65gjf/).

For the opposition analysis, five of the coefficients were in the predicted direction with a positive sign (see Fig. 2b). The early-stage measures demonstrated the strongest associations: All three were positive and statistically significant (first fixation duration: b = 0.000073, SE = 0.000025, p = .004; first-pass fixations: b = 0.036, SE = 0.0067, p < .001; first-pass fixation duration: b = 0.00012, SE = 0.000024, p < .001; see Table S8 at https://osf.io/65gjf/).

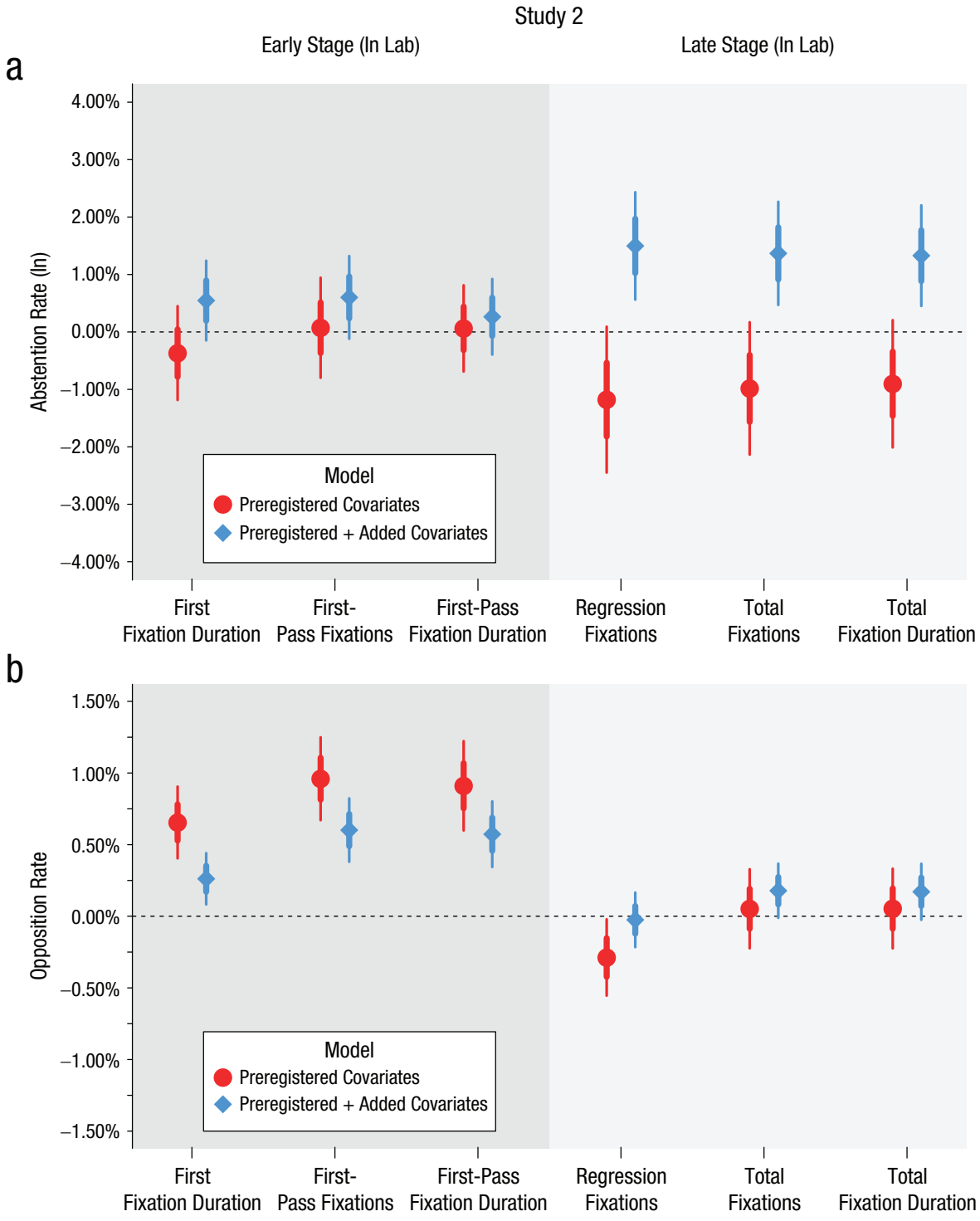


Fig. 2. Effect of a 1-*SD* increase on aggregate voting decisions in actual elections (Study 2; lab data were collected before actual voting occurred). The graphs show (a) the average abstention rate and (b) the average opposition rate for both the preregistered and alternative models, separately for each of the six-eye movement measures across the early and late stages. Point estimates are shown for both the preregistered and alternative models. Thicker lines represent ±1 cluster-robust standard errors, and thinner lines are 95% confidence intervals.

Finally, in terms of effect sizes, the effect of language-comprehension difficulties (as measured by eye movements) on aggregate voting decisions was small (see Fig. 2). For example, in the alternative models, a 1-*SD* (160.87 ms) increase in average total fixation duration was associated with a 0.11% increase in the abstention rate (95% CI = [0.043%, 0.18%]; see Note 1). The average total fixation duration was 326.73 ms. For the opposition analyses, a 1-*SD* (36.03 ms) increase in average first fixation duration was associated with a 0.26% increase in the opposition rate (95% CI = [0.083%, 0.44%]). The average first fixation duration was 152.45 ms.

In summary, for the preregistered analyses, eye movements predicted aggregate decisions to vote against the ballot measure but not the rate of abstentions. The alternative-model specifications that accounted for unexpected issues that we encountered for Study 2 show that the eye-movement measures predicted the rate of both abstention and opposition.

Comparison with common measures of language difficulty

We conducted exploratory analyses to examine whether eye movements also predicted aggregate voting decisions beyond what was accounted for by common measures of language difficulty. First, we assessed the predictive power of eye movements when accounting for processes captured by widely used linguistic metrics. Specifically, we estimated similar models as mentioned previously, but we added the Flesch-Kincaid Grade Level and the SUBTLEXUS median score for each ballot measure as covariates. The Flesch-Kincaid Grade Level assesses the readability of text and has been extensively used by researchers in the field of education, political scientists who study ballot language, and federal agencies in the U.S. government (Flesch, 1948; Reilly, 2010, 2015; Reilly & Richey, 2011).

We also used the SUBTLEXUS median score because word frequency is a common metric for assessing text difficulty in psycholinguistics (Hyönä & Olson, 1995; Rayner, 1998). Eye movements predicted aggregate voting decisions across the two studies even after we accounted for traditional text-based measures of language difficulty (see Tables S9 to S12 at https://osf.io/65gjf/). These results suggest that eye movements capture comprehension-related processes that are not accounted for by two commonly used metrics of language difficulty.

Second, we examined whether an alternative but common measure of text processing—total reading time—could robustly predict aggregate voting decisions. We measured total reading time as the amount of time from when a ballot measure appeared on the screen to when lab participants pressed a button allowing them to advance to the next screen. Longer reading time reflects greater difficulties in text comprehension, and this measure has been extensively used in the fields of education and psycholinguistics (Aaronson & Scarborough, 1977; Jegerski, 2014). The results were mixed (see Table S13 at https://osf.io/65gjf/). Longer reading times were associated with greater rates of opposition in Study 1 ($b = 0.00000035$, $SE = 0.00000015$, $p = .02$) and abstention in Study 2 ($b = 0.0000016$, $SE = 0.00000034$, $p < .001$). However, reading times were not associated with rates of abstention in Study 1 ($b = -0.00000034$, $SE = 0.00000035$, $p = .33$) and opposition in Study 2 ($b = 0.000000040$, $SE = 0.000000071$, $p = .58$). These and the eye-movement results suggest that eye movements robustly predict aggregate voting decisions when compared with a measure of total reading time.

Finally, we also examined the extent to which eye movements predicted aggregate voting decisions beyond what is accounted for by participants' in-lab voting decisions (i.e., decision to abstain/not abstain or oppose/support). Interestingly, participants' in-lab decisions to support or oppose a given ballot measure predicted aggregate rates of opposition for Study 1 (see Table S18 at https://osf.io/65gjf/). But, in-lab decisions to support/oppose and abstain/not abstain did not predict, respectively, aggregate rates of opposition for Study 2 (see Table S19 at https://osf.io/65gjf/) and abstention for Study 1 and Study 2 (see Tables S20 and S21 at https://osf.io/65gjf/). Furthermore, some of the eye-movement measures still predicted aggregate voting decisions for both Study 1 and Study 2 (see Tables S18 to S21).

Discussion

Across two studies, we found evidence that as real ballot measures became more difficult to understand, as indicated by eye-movement responses in the lab, the rate of aggregate decisions in actual elections to abstain from voting and vote against the ballot measure also increased. Furthermore, eye movements predicted aggregate voting decisions beyond what was captured by widely used measures of language difficulty and in-lab vote choices. Our study has several theoretical, methodological, and societal implications.

First, the findings expose the real-world importance and concerns of direct democracy. In particular, the results support the growing concern that how a ballot measure is written, rather than the substance of the policy itself, can influence voting decisions. This is an

important problem because politicians and special-interest groups may unintentionally or deliberately increase abstentions or votes against ballot measures by writing difficult-to-understand ballot language.

Second, our study also lays the groundwork for how these concerns may be addressed using eye-movement monitoring. Specifically, eye-movement measures are useful for explaining and predicting the consequences of ballot language on voting decisions. Eye movements have several advantages that make them ideally suited for examining the effects of ballot language on voting decisions. Eye movements can be collected without requiring participants to perform any task beyond silent reading, similar to what they would do inside the voting booth.

Additionally, eye-movement responses to linguistic features can be similar across languages and can be used to study voting decisions of non-English-speaking populations. For example, low-frequency words elicit greater gaze than high-frequency words in Spanish, German, and Chinese (Li et al., 2014; Tiffin-Richards & Schroeder, 2015; Whitford & Titone, 2017). In the context of the United States, this is important because the language-minority provisions of the Voting Rights Act have allowed millions of voters access to ballot measures translated in their non-English native language (Reilly, 2015). This suggests that eye movements can also be used to study the influence of non-English ballots on non-English-speaking voters.

Third, the results also support the notion that the psychological processes underlying voting decisions studied in a small group of individuals in a laboratory can generalize to a different group of voters in naturalistic settings. Indeed, it is striking that we observed the relationship between difficulties in text comprehension and voting decisions despite the variety of differences between the context of the lab and natural voting environments. For example, the lab participants knew that their vote choices were being observed, whereas choices in the voting booth are private. Choices in the voting booth occur in the informational and emotional environment of Election Day, whereas our lab studies occurred outside the context of Election Day. Furthermore, some lab participants evaluated several ballot measures that appeared in elections 6 years prior. Yet, despite these differences, we observed an association between difficulties in ballot comprehension and actual voting decisions.

Although the results are promising, the findings should also be interpreted in light of the study's limitations. Given that we used real ballot measures, we had less control over their characteristics. We used a careful ballot-selection procedure and covariate adjustment in the analyses to address possible confounding factors. But it is possible that factors other than difficulties in language comprehension accounted for the relationship between eye movements and aggregate voting decisions. Studies in which language difficulty is experimentally manipulated are therefore important for future work in this area.

We observed evidence for the predicted effects in our preregistered abstention and opposition analyses for Study 1 and the preregistered opposition analyses for Study 2. We did not, however, observe similar effects for the preregistered abstention analyses for Study 2. It was only after we accounted for several unexpected issues that could have affected voters' knowledge of the ballot measures in actual elections (e.g., campaign advertisement, local media coverage, explainers associated with ballots) that we observed the predicted relationship between eye movements and the abstention rate. This makes it less clear whether the abstention results in Study 2 can be viewed as a replication of the abstention results in Study 1.

We also took great care to ensure that both the voters in the lab and in actual elections were likely unfamiliar with the ballot measures. As a consequence, we intentionally did not select high-salience issues (e.g., gun control, affirmative action). Difficulties in comprehending ballot text may exert a weaker influence on voting decisions for well-known issues or ones in which voters possess strong prior attitudes. Therefore, future work should examine the extent to which the results we observed here generalize to ballot measures about which voters possess a high level of knowledge and emotional associations (i.e., ballot measures pertaining to highly partisan issues).

Furthermore, we selected ballot measures that varied in the number of familiar and unfamiliar words they contained. This may, in part, explain why difficulties in language comprehension (as measured by eye movements) had small effects on aggregate voting decisions. There are other features of ballot measures that can make them difficult to understand. One important source of difficulty is the manner in which words are arranged into phrases, clauses, and sentences (i.e., syntax). For example, individuals may be less likely to understand information conveyed via a long complex sentence (containing multiple clauses) than when the same information is conveyed through separate sentences (see Supplemental Discussion at https://osf.io/65gjf/).

These different sources of language difficulty—unfamiliar words and complex syntax—can simultaneously be present in ballot measures and, in combination, may produce larger effects on voting decisions. This is important given that even small effects can affect electoral outcomes in competitive elections. For example,

during the 2012, 2013, and 2014 U.S. elections, the margin of victory for approximately 10% of ballot measures was between 1% and 5%. Beyond electoral outcomes, incremental increases in the size of the margin of victory can also affect voter perceptions of the law's legitimacy (Arnesen et al., 2019). Our study, then, compels further investigation into other sources of language difficulty and their individual and joint effects on voting decisions.

Despite these limitations, our study highlights the usefulness of eye-movement measures for studying decision-making processes of voters in direct democracy elections. Here, we defined our areas of interest at the level of words. However, researchers can also define interest areas at other levels—phrases, sentences, paragraphs—and eye movements can provide an online record of reading performance at these levels (Hyönä & Lorch, 2004; Traxler et al., 2002). This property of eye movements is useful for future work examining the effects of complex syntax on voting decisions.

Beyond their capacity to predict voting decisions, eye movements may be used in future work to address other long-standing questions in political-science research, such as whether voting decisions are the product of careful versus cursory thinking (Lau & Redlawsk, 2006). Of relevance, eye movements have been used in studies on reading to distinguish skimming and mind wandering from careful reading of text (Reichle et al., 2010; Strukelj & Niehorster, 2018). Eye movements, then, can potentially be used in future work to examine the conditions that lead voters to carefully read, and deeply think about, the substantive content of ballot measures.

Finally, this study also contributes to the literature on eye movements. Our study is the first of its kind to show that the predictive power of eye movements extends to real-world voting decisions. In addition, our work demonstrates the utility of eye movements as an approach for understanding aggregate-level decisions. Individuals often extract information from reading text to inform their decisions such as whether to share a news article, comment on a social media post, sell a stock, and so on. Importantly, these individual-level decisions can scale up to aggregate-level social phenomena (e.g., virality of a news article, panic selling of stocks; Knutson & Genevsky, 2018; Scholz et al., 2017). Future work can explore whether eye movements can explain and predict aggregate-level choices in other domains.

In summary, as more countries adopt direct democracy elections, the question of how ballot language influences voting decisions will increasingly be an important issue for politicians, special-interest groups, and voters. Our work sets the foundation for the use of eye movements as an important tool to aid researchers and policymakers in creating ballot measures that promote comprehension and civic involvement among voters.

Transparency

Action Editor: Kate Ratliff
Editor: Patricia J. Bauer

Author Contributions

J. C. Coronel and H. C. Shulman conceived of the initial study concept. J. C. Coronel, O. M. Bullock, H. C. Shulman, M. D. Sweitzer, and R. M. Bond developed the research design. O. M. Bullock, S. Poulsen, and J. C. Coronel collected the lab-based data. O. M. Bullock obtained real-world ballot information and collected the norming data. M. D. Sweitzer obtained linguistic information about the ballot measures. J. C. Coronel, M. D. Sweitzer, R. M. Bond, and O. M. Bullock analyzed the data. J. C. Coronel wrote the manuscript, and all authors gave critical feedback. All the authors approved the final manuscript for submission.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

This work was supported by The Ohio State University's School of Communication Miller Award.

Open Practices

All data and code have been made publicly available via OSF and can be accessed at https://osf.io/jfxaz. Word-frequency norms from the SUBTLEXUS corpus are publicly available at http://brm.psychonomic-journals.org/content/supplemental. The design and analysis plans for Study 1 were preregistered at https://osf.io/hdc7x. Study 2 was not officially preregistered, but the preregistration plan was posted prior to data collection (see https://osf.io/srxyu). Changes to the preregistration of Study 1 were posted before data collection, and additional changes to the preregistrations are discussed in the text. This article has received the badges for Open Data, Open Materials, and Preregistration. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

ORCID iDs

Jason C. Coronel https://orcid.org/0000-0001-6397-3860
Matthew D. Sweitzer https://orcid.org/0000-0002-2075-6177

Acknowledgments

We thank Katie Ferriby, Morgan Hellmich, Alec Hinton, and Natalie Petit for their help in data collection or coding of the ballot measures. We also thank Christopher Dawes, Dave DeAndrea, Kara Federmeier, James Fowler, Kelly Garrett, David Hendry, Ian Krajbich, Molly Ritchie, Christin Scholz, Ralf Schmälzle, and graduate students enrolled in Communication Research Methods at The Ohio State University for their

helpful feedback and comments on early drafts of the manuscript.

Note

1. For ease of interpretation, this estimate is based on a model that uses the nontransformed version of the abstention rate.

References

Aaronson, D., & Scarborough, H. S. (1977). Performance theories for sentence coding: Some quantitative evidence. *Journal of Verbal Learning and Verbal Behavior*, *16*(3), 277–303.

Arnesen, S., Broderstad, T. S., Johannesson, M. P., & Linde, J. (2019). Conditional legitimacy: How turnout, majority size, and outcome affect perceptions of legitimacy in European Union membership referendums. *European Union Politics*, *20*(2), 176–197.

Barth, J., Burnett, C. M., & Parry, J. (2020). Direct democracy, educative effects, and the (mis)measurement of ballot measure awareness. *Political Behavior*, *42*, 1015–1034. https://doi.org/10.1007/s11109-019-09529-w

Bowler, S., & Donovan, T. (1998). *Demanding choices: Opinion, voting, and direct democracy*. University of Michigan Press.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990.

Burnett, C. M. (2019). Information and direct democracy: What voters learn about ballot measures and how it affects their votes. *Electoral Studies*, *57*(1), 223–244.

Cartwright, N., & Hardie, J. (2012). *Evidence-based policy: A practical guide to doing it better*. Oxford University Press.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, *32*(3), 221–233.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83.

Hyönä, J., & Lorch, R. F. (2004). Effects of topic headings on text processing: Evidence from adult readers' eye fixation patterns. *Learning and Instruction*, *14*(2), 131–152.

Hyönä, J., & Olson, R. K. (1995). Eye fixation patterns among dyslexic and normal readers: Effects of word length and word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(6), 1430–1440.

Jegerski, J. (2014). *Research methods in second language psycholinguistics*. Routledge.

Kaufmann, B., & Mathews, J. (2018, May 8). Democracy doomsday prophets are missing this critical shift. *The Washington Post*. https://www.washingtonpost.com/news/theworldpost/wp/2018/05/08/direct-democracy-is-thriving/

Knutson, B., & Genevsky, A. (2018). Neuroforecasting aggregate choice. *Current Directions in Psychological Science*, *27*(2), 110–115.

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, *13*(10), 1292–1298.

Lau, R. R., & Redlawsk, D. P. (2006). *How voters decide: Information processing during election campaigns*. Cambridge University Press.

Li, X., Bicknell, K., Liu, P., Wei, W., & Rayner, K. (2014). Reading is fundamentally similar across disparate writing systems: A systematic characterization of how words and characters influence eye movements in Chinese reading. *Journal of Experimental Psychology: General*, *143*(2), 895–913.

Lodge, M., Steenbergen, M. R., & Brau, S. (1995). The responsive voter: Campaign information and the dynamics of candidate evaluation. *American Political Science Review*, *89*(2), 309–326.

Lupia, A. (1992). Busy voters, agenda control, and the power of information. *American Political Science Review*, *86*(2), 390–403.

Nicholson, S. P. (2003). The political environment and ballot proposition awareness. *American Journal of Political Science*, *47*(3), 403–410.

Pärnamets, P., Johansson, P., Hall, L., Balkenius, C., Spivey, M. J., & Richardson, D. C. (2015). Biasing moral decisions by exploiting the dynamics of eye gaze. *Proceedings of the National Academy of Sciences, USA*, *112*(13), 4170–4175.

Quesenbery, W., & Chisnell, D. (2016, November 3). Ballot measures need to be written in plain language. *The New York Times*. https://www.nytimes.com/roomfordebate/2016/11/03/why-are-ballot-measures-so-confusing/ballot-measures-need-to-be-written-in-plain-language

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506.

Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading*, *10*(3), 241–255.

Rayner, K., Sereno, S. C., Morris, R. K., Schmauder, R. A., & Clifton, C. (1989). Eye movements and on-line language comprehension processes. *Language and Cognitive Processes*, *4*(3–4), SI21–SI49.

Reichle, E. D., Reineberg, A. E., & Schooler, J. W. (2010). Eye movements during mindless reading. *Psychological Science*, *21*(9), 1300–1310.

Reilly, S. (2010). *Design, meaning and choice in direct democracy: The influences of petitioners and voters*. Ashgate.

Reilly, S. (2015). *Language assistance under the Voting Rights Act: Are voters lost in translation?* Lexington Books.

Reilly, S., & Richey, S. (2011). Ballot question readability and roll-off: The impact of language complexity. *Political Research Quarterly*, *64*(1), 59–67.

Scholz, C., Baek, E. C., O'Donnell, M. B., Kim, H. S., Cappella, J. N., & Falk, E. B. (2017). A neural model of valuation and

information virality. *Proceedings of the National Academy of Sciences, USA*, *114*(11), 2881–2886.

Shockley, E., & Fairdosi, A. S. (2015). Power to the people? Psychological mechanisms of disengagement from direct democracy. *Social Psychological and Personality Science*, *6*(5), 579–586.

Strukelj, A., & Niehorster, D. C. (2018). One page of text: Eye movements during regular and thorough reading, skimming, and spell checking. *Journal of Eye Movement Research*, *11*(1), 1–22.

Tiffin-Richards, S. P., & Schroeder, S. (2015). Word length and frequency effects on children's eye movements during silent reading. *Vision Research*, *113*(Pt. A), 33–43.

Traxler, M. J., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language*, *47*(1), 69–90.

Whitford, V., & Titone, D. (2017). The effects of word frequency and word predictability during first- and second-language paragraph reading in bilingual older and younger adults. *Psychology and Aging*, *32*(2), 158–177.

Political Psychology, Vol. 29, No. 2, 2008

Framing Persuasive Appeals: Episodic and Thematic Framing, Emotional Response, and Policy Opinion

Kimberly Gross
George Washington University

Those seeking to frame political issues to their advantage recognize the power of emotional appeals. Yet the study of framing has focused mainly on the cognitive effects of framing rather than on its emotional effects. This study presents the results of two experiments designed to explore the effect of episodic and thematic framing on emotional response and policy opinion. Participants were randomly assigned to read a column arguing against mandatory minimum sentencing that employed either a thematic or one of two episodic frames featuring a woman who received a harsh sentence under the policy. Episodic framing was more emotionally engaging. Furthermore, the specific emotions elicited by the episodic frame—sympathy and pity for the woman featured in the column—were associated with increased opposition to mandatory minimum sentencing. Yet the thematic frame was actually more persuasive once this indirect effect of frame on emotional response was taken into account. The results are consistent with the conclusion that framing effects on policy opinion operate through both affective and cognitive channels. The theoretical and practical implications of the study are discussed.

KEY WORDS: Episodic Framing, Emotion, Mandatory Minimum Sentencing

Conventional wisdom tells us that emotional appeals matter. Those seeking to influence opinion and frame political issues to their advantage certainly seem to believe that appeals to emotion aid them in their attempts to gain public support. Moreover, a growing body of research demonstrates that emotion can play a crucial role in how citizens process political information and arrive at political judgments. Yet we know little about the possible effects of framing attempts on emotional response because the framing literature, with a few exceptions, has focused on cognitive reactions. This study extends Iyengar's (1991) work to examine how the use of episodic and thematic framing in a persuasive message affects emotional response and how these emotional reactions might

Published by Blackwell Publishing. Inc., 350 Main Street, Malden, MA 02148, USA, 9600 Garsington Road, Oxford, OX4 2DQ, and PO Box 378 Carlton South, 3053 Victoria Australia

help us to understand the link between these frames and policy views. It also focuses on emotions—sympathy and pity—that have received less attention in the political science literature. In doing this, the study not only increases our understanding of the dynamics of framing, but also contributes to a growing body of work that seeks to understand the role of emotion in political communication.

Framing and Emotion

According to Gamson and Modigliani (1987) a frame is "a central organizing idea or story line that provides meaning to an unfolding strip of events weaving a connection among them. The frame suggests what the controversy is about, the essence of the issue" (p. 143). The key premise of the framing literature is that frames will, by highlighting certain aspects of an event or policy, guide audience members' thoughts about that event or issue in predictable ways, to predictable conclusions. Numerous studies have shown that the particular frame imposed on an issue or event can shape opinion on related policies (e.g., Druckman 2001; Jacoby, 2000; Kinder & Sanders, 1990; Nelson, Clawson, & Oxley, 1997; Sniderman & Theriault, 2004).

Clearly, partisans try to tap the power of emotion when developing their frames. This makes the relative lack of attention to emotion in the framing literature surprising. There are, of course, exceptions. Kinder and Sanders (1990) showed that affirmative action framed as unfair advantage invoked indignation in a way that affirmative action framed as reverse discrimination did not. Negative emotions—anger, disgust, and fury—were associated more sharply with opinion on affirmative action under the unfair advantage frame. Masters and Sullivan (1993) reported the results of an experiment suggesting that people's emotional reactions toward leaders were shaped by news commentators' judgments and interpretations. Gross and D'Ambrosio (2004) found that frames alter the explanations citizens gave for their emotional response and that frames alter the relationship between emotional response and predispositions. Other work suggests that the persuasive effect of a frame depends on emotional reactions to that frame (Brewer 2001; Nabi 1998). These studies, consistent with the argument presented in this study, show that emotional response may be contingent upon how an issue is framed and that these emotional responses may mediate the effect of frames on opinion.

These studies focus on frames that are specific to an issue or context. Other types of frames might be thought of as more general rhetorical devices—ways of telling a story to make it more understandable, accessible, and compelling to the audience. Such frames can be employed across a range of issues. In this study, I focused on one of these more general rhetorical devices—episodic and thematic framing.

Episodic and Thematic Framing

Episodic frames present an issue by offering a specific example, case study, or event oriented report (e.g., covering unemployment by presenting a story on the plight of a particular unemployed person). Thematic frames, on the other hand, place issues into a broader context (e.g., covering unemployment by reporting on the latest unemployment figures and offering commentary by economists or public officials on the impact of the economy on unemployment).

For journalists, these frames may be employed in coverage of any number of issues. They are ways of telling the story that make it understandable and accessible to readers. Journalists often use episodic frames because they believe them to be more compelling and more likely to draw the reader or viewer into the story. Put another way, episodic frames are thought to be more emotionally engaging. Partisans also use what might be considered episodic frames for much the same reason. Advocates frequently highlight a particular individual's story as illustrative of a broader issue or problem as a compelling way to make their case. They select a particular example in order to maximize the persuasive appeal of their claim. Consider, for example, President Ronald Reagan's use of the "welfare queen" anecdote in the 1976 presidential campaign. This story of a woman who defrauded the system was seemingly selected (and embellished) because it provoked anger that reinforced opposition to the program and support for the candidate who would change it.

Prior research on episodic and thematic frames has mostly focused on their use in news coverage and has failed to explore the role of emotion. In his book, *Is Anyone Responsible?*, Iyengar (1991) examined the effects of episodic and thematic framing in television news. He argued that the prevalence of episodic framing in political news coverage diverts attention from societal responsibility and leads people to hold individuals responsible for their own predicaments, thereby dampening support for government programs designed to address problems and shielding leaders from responsibility. Episodic framing also encourages a "morselized" understanding of political problems by presenting recurring problems as discrete instances (Iyengar, 1991, 136). Citizens exposed to a steady stream of episodic frames fail to see the connections between problems such as poverty, racial discrimination, and crime when they are presented as discrete and unconnected.

His expectations that episodic coverage leads to individualistic attributions while thematic coverage engenders societal attributions were borne out in various experiments. In the face of episodic frames, individuals were more likely to offer individualistic attributions; in the face of thematic frames, they offered more societal attributions. Iyengar also demonstrated that causal attributions were associated with views on policy in systematic ways. For example, people who attribute causal responsibility for poverty to societal factors were particularly likely to support social welfare spending increases. Iyengar did not,

however, directly test the effects of episodic and thematic frames on emotional response. Nor did he investigate the role that emotion plays in determining the effect of these frames on attributions or policy views. In addition, he examined news coverage, not persuasive political communication. In this research, I examine the role episodic frames and thematic frames play in enhancing the role of a persuasive appeal and the role emotion plays in explaining these effects.

The Effect of Episodic Frames on Emotion

The general claim that framing should affect emotional response follows from appraisal theories of emotion. This approach posits that people experience emotion as a product of their cognitive appraisals of an event or phenomenon (e.g., Ellsworth, 1991; Lazarus, 1991; Ortony, Clore, & Collins 1988; Roseman, 1991). Appraisal theory focuses mainly on the patterns of appraisal that elicit specific emotions and the ways in which individual and situational factors interact to produce different emotional responses. If frames alter the information and considerations subjects have at hand, cognitive appraisal models would predict that emotional outputs should differ.

More specifically, in the case of episodic and thematic framing, I expect episodic frames to generate stronger emotional responses than thematic frames. The "human interest" details of an episodic frame should be more emotionally engaging than the pallid statistics of a thematic frame. After all, the emotional power of episodic framing to engage the audience is probably one of its strongest appeals. While episodic frames should elicit more emotions in general, the nature of emotional response (i.e., the specific emotions elicited) would depend on the content of a given frame. Thus, one must consider the details of the particular story used in an episodic frame to generate predictions about specific emotions.

In this study, I presented individuals with an opinion column arguing against mandatory minimum sentencing for drug offenses. In constructing my episodic frames, I featured a woman with no prior record who was sentenced to nearly 25 years in prison for conspiring to help the drug-dealing boyfriend who abused her. The column recognized she broke the law but suggested that the sentence was excessive given the extenuating circumstances. Though the name of the defendant and some details were changed, I based her on a real case; a case that received considerable attention in the press and was frequently cited in news reports and by advocates as illustrating the problems with mandatory minimum sentences. Given the particular details of this column, I expected the episodic frames to generate expressions of sympathy and pity in particular. Had I presented a column in favor of mandatory minimum sentencing featuring a repeat offender, I would expect different emotions (anger and perhaps anxiety) to dominate.

Using Episodic and Thematic Frames in Persuasive Appeals: Will Episodic Frames Enhance Persuasive Appeals?

Whether the use of episodic framing would enhance or diminish the effect of a persuasive appeal relative to a thematic frame is somewhat less clear from prior literature. Prior research, from a cognitive perspective, suggests two contradictory expectations for whether episodic or thematic framing of the column should produce greater opposition to mandatory minimums. Some psychologists have hypothesized that vivid information should be more persuasive than information presented in a dispassionate manner because it should be more memorable, receive greater attention, and be more easily brought to mind. This is likely the conventional wisdom among those using a vivid case (e.g., an episodic frame) to illustrate a claim. According to this line of reasoning, one might expect the episodic frame to produce greater opposition to mandatory minimum sentencing than the thematic frame because the arguments will be better remembered. Though this common wisdom is appealing, Fiske and Taylor (1991) argue there is little empirical evidence supporting the vividness hypothesis with one exception of relevance here: Individual case histories are more persuasive than group statistics (but see Allen & Preiss, 1997).

In contrast, there is strong evidence to suggest that episodic frames may actually be less persuasive in a context such as the one at issue in this study. As noted above, Iyengar (1991) found that episodic frames produced individual attributions for political problems and thematic frames produced societal attributions for political problems. Societal attributions of responsibility led people to offer greater support for government programs to solve political problems and to hold political leaders responsible. Episodic framing of political problems diverts attention from societal responsibility and leads people to hold individuals responsible for their own situation, thereby diminishing support for government programs designed to address the problem. According to this line of reasoning, one might expect that those in the episodic framing conditions would see the woman featured in the episodic column as responsible for her predicament, rendering the persuasive effect of the episodic frames weaker than that of the thematic frame.

None of these claims, however, take account of the potential role of the emotions elicited by these framing devices in facilitating persuasion. Iyengar's work suggests that the effect of these frames operates at a cognitive level. I argue that if my expectation of distinct emotional reactions between episodic and thematic frames holds, the effect of frames on policy views may *also* operate at least partly via these affective responses. A considerable body of research in political science has found that emotional reactions shape political judgments and that their effects vary in sensible ways (e.g., Abelson, Kinder, Peters, & Fiske, 1982; Brader, 2005, 2006; Conover & Feldman, 1986; Huddy, Feldman, Taber, & Lahev, 2005; Kinder, 1994; Marcus & MacKuen, 1993; Marcus, Neuman, & MacKuen, 2000). Although these authors employ different theoretical perspectives on emotion, taken together their work shows that emotional response can influence opinion. Assuming that

frames do alter emotional response, prior work on emotion and opinion would suggest that differences in opinion following exposure to episodic or thematic frames might be the result of the different emotional reactions they generate.

Schwarz (2000) suggests that emotional reactions themselves serve as a basis for judgment. This affect-as-information approach suggests that people rely on their feelings to make judgments as long as the feelings are perceived as relevant to the judgment at hand (Schwarz, 1990; Schwarz & Clore, 1983). Put another way, the emotions themselves are considerations used in coming to an opinion. Schwartz's work explores this in the context of mood; however, I follow Lerner and Keltner's (2000) argument that these influences should be emotion specific (see also Kinder, 1994). This suggests that the persuasive effect of any frame could depend on the specific emotion elicited by the frame. For example, sympathy for those who are sentenced to what may be seen as unduly harsh punishment under mandatory minimums would be expected to foster opposition to the policy whereas anger at those who commit drug offences might foster support for the policy. Because I expected the episodic frames employed in this study to generate sympathy and pity for someone who received an unduly harsh sentence, I expected the episodic frames to be associated with increased opposition to mandatory minimum sentencing.

Finally, I was interested in whether the race of the individual used in the episodic frame would alter the impact of that frame given that previous research has sometimes found evidence for differences in the effects of episodic (vivid) presentations based on the race of the individual featured (e.g., Iyengar & Kinder 1987). These differences could be the result of differences in emotional or cognitive responses to the individual featured. Iyengar's own experiments suggest that qualities of the particular individual highlighted in an episodic story may play a role in determining the effect of episodic framing. In the case of crime, which is of special relevance here, he found significant differences across episodic frames featuring white violent crime and episodic frames featuring black violent crime. Audience members cited societal causal attributions more frequently when coverage focused on white violent crime and less frequently when coverage focused on black violent crime; news about black crime also enhanced attributions directed at individuals (see also Gilliam & Iyengar, 2000). I included a race manipulation in the episodically framed column in this study to determine if respondents would express less sympathy and pity and to determine if the direct effects of the episodic frame would differ when the column featured a black individual.

Method

Design

To explore how episodic and thematic frames affect emotional response and policy views, I conducted two laboratory experiments with students at a private

mid-Atlantic university. The experiments varied the presentation of an opinion column arguing against mandatory minimum sentences for drug offenses. Specifically, I created three versions of the opinion column using information and quotations that appeared in newspaper coverage and editorials dealing with this issue: a thematic column, an episodic column featuring an African-American woman, and an episodic column featuring a white woman. These treatment articles were used in both experiments, though the procedure and the variables measured in each experiment differed somewhat (e.g., the first experiment included a control condition; the second experiment included additional emotions measures as detailed below).

The thematically framed column presented the case against mandatory minimum sentences by providing details on the sentencing guidelines, the exploding prison population, and the high costs of incarceration. This column also included quotations from an interest group representative and a Supreme Court justice opposing the policy. Both episodic frames used the story of a particular individual—"Janet Smith," a woman sentenced under the federal drug-sentencing laws for conspiring to help her drug-dealing boyfriend—to illustrate the case against mandatory minimum sentences. The column observed that she broke a number of laws and was "no angel" but argued that her sentence was "disproportionately harsh for a woman with no prior criminal history, who had arguable extenuating circumstances mitigating the severity of her conduct and who had never actually handled drugs." After featuring her story, the episodically framed columns included the same quotations from the interest group representative and Supreme Court justice. To assess whether the race of the individual featured in the episodic frame would matter there were two versions of this treatment. In one "Janet Smith" was white, in the other she was black. See Appendix A for the text of the treatment articles.

Procedure

I conducted the first experiment in the fall of 2001. Participants completed a pretest that measured opinion on mandatory minimum sentencing and a host of other policy and demographic variables. After a minimum of two weeks—an interval intended to minimize contamination by the pretest questions—each participant was randomly assigned to read one of the three mandatory minimum columns or a story on DC voting rights (the control condition). To prevent undue attention to the crime story, all participants also read two unrelated articles. They then completed a posttest questionnaire designed to measure views on mandatory minimum sentencing, opinions on a variety of other public policies, and emotional reactions to the articles. Specifically, I asked if they felt anger, disgust, sympathy, or pity while reading the article. These emotions—emotions that suggest empathy and aversion—were selected because I anticipated they were the emotions most likely to be evoked by the details of the columns.

The second experiment, administered during the fall of 2002, was designed to further explore the role of emotion by including additional emotions and by including questions designed to assess respondents' explanations for their emotional reactions. Participants read one of the same three columns dealing with mandatory minimum sentencing along with a filler story and then completed a posttest questionnaire designed to measure views on mandatory minimum sentencing, emotional reactions, and demographics. This experiment did not include a control condition or feature a pretest as in Experiment 1. It did feature two emotions assessing anxiety (fear and worry) that were not included in the first experiment. The measure of emotional response also differed between the two experiments. In the second experiment I asked participants whether they felt the emotion (binary variable) and, for those who said they felt an emotion, I used a follow-up question asking about strength of that emotional response. This experiment also included open-ended questions that asked about the object or impetus for a particular emotional response as well as questions developed by Wolski and Nabi (2000) to assess respondents' perceptions of their attention to, interest in, and objectivity toward the article. See Appendix B for more detail on these measures.

Sample

A total of 163 participants completed the first experiment. Fifty-two percent were women; 84% were white. Forty-seven percent of the sample identified as Democrats, 14% were Republicans, and 39% were Independents (58% of the Independents lean Democratic, and 16% lean Republican). A total of 105 participants completed the second experiment. Sixty-three percent were women, and 84% percent were white. Forty-seven percent of participants identified as Democrats, 13% identified as Republicans, and 40% identified as Independents (62% of Independents lean Democratic; 26% of Independents lean Republican). The use of student samples necessitates caution in generalizing from the results (Sears, 1986), a point I return to in the discussion.

In keeping with the democratic orientation of the samples, participants in both experiments tended to be fairly supportive of equality (mean .70, standard deviation .19 on the 6-item American National Election Study (ANES) equality scale in experiment 1; mean .70, standard deviation .21 in experiment 2) and less supportive of limited government (mean .39, standard deviation .22 on the three item ANES limited government scale in experiment 1; mean .31, standard deviation .37 in experiment 2). Participants in the second experiment were also less racially resentful (mean .36, standard deviation .25 on a 2-item racial resentment scale—there was no measure of resentment in the first experiment). I found no evidence of systematic differences across conditions on a variety of background measures for either of the two experiments.

Table 1. Emotional Response by Frame

Emotion	Episodic Frame White Individual	Episodic Frame Black Individual	Thematic Frame
	Experiment 1		
Sympathy	.61 (.28)***	.66 (.23)***	.43 (.29)
Pity	.53 (.30)***	.56 (.30)***	.33 (.28)
Anger	.53 (.36)*	.59 (.30)***	.39 (.34)
Disgust	.51 (.37)	.54 (.34)*	.40 (.34)
Aversion Scale	.52 (.35)*	.57 (.28)**	.39 (.31)
Empathy Scale	.57 (.28)***	.61 (.24)***	.38 (.26)
	N = 38	N = 40	N = 41
	Experiment 2		
Sympathy	.41 (.33)***	.52 (.28)***	.18 (.27)
Pity	.31 (.29)**	.37 (.31)***	.13 (.26)
Anger	.25 (.25)	.24 (.29)	.29 (.29)
Disgust	.31 (.31)	.33 (.31)	.28 (.31)
Worry	.16 (.28)	.10 (.24)	.20 (.29)
Fear	.09 (.25)	.04 (.14)	.07 (.19)
Aversion Scale	.28 (.25)	.29 (.26)	.28 (.27)
Empathy Scale	.36 (.26)***	.44 (.26)***	.16 (.19)
Anxiety Scale	.13 (.25)	.07 (.16)	.14 (.22)
	N = 31	N = 36	N = 38

Note. Table entry is mean emotional response by frame with standard deviation in parentheses. Emotional reactions are coded to range between 0 and 1. Asterisks indicate that the episodic framing condition differs significantly from the thematic framing condition (t-test on means): ***p < .01. **p < .05. *p < .10.

Results

Framing Effects on Emotional Response

I found support for the claim that episodic framing alters emotional response relative to a thematic article on the same topic. Table 1 presents mean emotional reaction by frame condition with standard deviation in parentheses. The emotional reaction variables were recoded to range between 0 and 1, where 1 signals stronger emotional response.[1] The table reports emotional reactions for individual emotions as well as scales for empathy (sympathy and pity combined), aversion (anger and disgust combined), and anxiety (fear and worry combined).

Those who read a column that featured an episodic frame were significantly more likely to express sympathy and pity in both experiments. Moreover, this was

[1] Emotional response was measured slightly differently in the two experiments, which results in differences in the reported level of emotional reaction between the two experiments. I am interested in the pattern of means across experimental conditions, not the differences between the two experiments. See Appendix B.

true whether Janet Smith was described as a black mother or a white mother. In the first experiment, those who read one of the episodically framed columns were also more likely than those in the thematic condition to report feeling angry. In addition, those in the episodic condition featuring the black mother were significantly more likely to report feeling disgust. However, these findings for anger and disgust were not replicated in the second experiment.[2]

In short, as I expected, the episodic frames seemed to have a greater emotional impact on the reader than the thematic frame. Given the nature of this particular treatment, it was not surprising that the effects found were mainly centered on empathetic emotions—pity and sympathy. The goal of the columnist in this case would be to use the story of Janet Smith to generate sympathy and pity that would translate into reduced support for mandatory minimum sentences. The results in Table 1 suggest that the first step was successful—the episodic frames generated increased expressions of sympathy and pity (empathetic emotions). Moreover, this increased empathy was not specific to the race of the individual portrayed. There were no differences across the two episodic frames on any of the emotions examined. The differences in emotional reactions were driven by the difference between the use of thematic and episodic storytelling.

Framing Effects on Opinion about Mandatory Minimum Sentencing

Next, I explored the effect of these different frames on the persuasive appeal of the column by examining views about mandatory minimum sentencing. All participants were asked "Do you favor or oppose mandatory minimum sentences for first or second time nonviolent drug offenses?" Responses to this question were coded to range between 0 and 1 where 1 represents those who strongly oppose mandatory minimum sentencing.[3] Table 2 presents mean opinion on mandatory minimum sentencing by framing condition for each experiment and opinion change by framing condition for the first experiment.

The argument developed above, in combination with the results of the analysis of framing effects on emotion, leaves me with no clear expectation about which frames will be more persuasive. If emotional reactions influence judgments, the sympathy and pity aroused by the episodic frames should lead to increased opposition to mandatory minimum sentencing relative to the thematic frame. Yet Iyengar's work suggests reasons that the episodic frames may not, in fact, be as persuasive as the thematic frame.

[2] All framing results hold when I estimate the effect of frame, controlling for gender, ideology, support for equality, support for limited government, and party identification (and in the second experiment racial resentment).

[3] In keeping with the liberal orientation of the student samples, participants on the whole opposed mandatory minimum sentences as evidenced by the high mean in the control condition (first experiment).

Table 2. Opposition to Mandatory Minimum Sentencing by Framing Condition

Episodic Frame White Individual	Episodic Frame Black Individual	Thematic Frame	Control Condition
	Experiment 1: Post test Opinion		
.72 (.24)	.64 (.29)	.75 (.27)**	.64 (.31)
N = 37	N = 39	N = 40	N = 42
	Experiment 1: Change in Opinion Pretest to Posttest		
.10 (.24)**	.03 (.29)	.14 (.35)***	.00 (.24)
N = 37	N = 39	N = 40	N = 41
	Experiment 2: Post test Opinion		
.67 (.25)	.61 (.28)	.70 (.26)	
N = 29	N = 36	N = 37	

Note. Table entries are mean opinion on mandatory minimum sentencing with standard deviation in parentheses. Opinion on mandatory minimum sentencing is coded to range between 0 to 1 where 1 represents those who strongly oppose mandatory minimum sentences for first time drug offenders. Change in opinion on mandatory minimum sentencing is coded so that positive numbers represented increased opposition between the pretest and posttest.
Asterisks indicate that a t-test on means shows framed condition differs significantly from control condition, **$p < .05$, ***$p < .01$ (one tail test).

The results of the first experiment, which included a control condition and pretest measure of opinion on mandatory minimum sentencing, suggest that the thematically framed column was most persuasive while the column featuring the story of a black Janet Smith did little to move opinion. I found significant differences in opinion change across experimental conditions. Those in the thematic frame condition and in the episodic frame condition featuring a white defendant exhibited significantly greater mean opinion change (e.g., they became significantly more opposed) than those in the control condition. By contrast, those who read the column featuring a black defendant were indistinguishable from the control group. This contrast is striking given that the race manipulation in the study was subtle: the conditions vary a single word within a 555-word column.

The pattern of means across framing conditions suggests a similar interpretation. In the first experiment, the difference between the control and thematic conditions was significant ($p < .05$, one-tail test). The difference between the episodic black and thematic condition was also significant at $p < .10$ (two-tail test).[4] The pattern looks similar in the second experiment—those who read about Janet Smith, the black mother, were least opposed to mandatory minimum sentencing, and those who read the thematic version were most opposed. However,

[4] Given the persuasive intent of the column, one-tail tests were appropriate when examining the framed conditions relative to the control condition while two-tail tests of significance were appropriate when examining differences between the differently framed columns (I did not have clear expectations regarding which frames would be more effective).

differences among the frames were not significant. Here, in contrast to the pattern for emotional response, the pattern for opinion on mandatory minimum sentencing suggests that the race of the individual used in the episodic frame did matter.

How Episodic Framing Effects Work: Affective and Cognitive Influences on Policy Views

The results in Table 1 show that episodic framing conditions invoked sympathy and pity to a greater degree than the thematic framing condition in both experiments. Given the persuasive intent of the column and the nature of the individual's story, this would lead one to expect the episodic framing conditions to be associated with greater opposition to mandatory minimum sentencing if emotional response (sympathy and pity in this case) influences policy views. Yet I found a pattern that suggests the opposite when I examined the direct effect of frames on policy attitudes. Those who read the thematic frame expressed greater opposition (Table 2). One possible explanation for these seemingly contradictory findings is that participants' sympathetic responses were not targeted in a way that would encourage opposition to mandatory minimum sentencing. For example, sympathy and pity for those hurt by Smith's actions in helping her boyfriend might not lead to greater opposition to mandatory minimums, though sympathy and pity for Smith herself might be expected to do this. Put another way, perhaps I had been assuming something about the nature of the empathy response that was untrue. The second experiment allowed me to explore this possibility by examining the questions which asked participants to explain their emotional reactions.

When I examined these open-ended responses, I found nothing to suggest that the episodic frames induced emotional reactions targeted in ways that might have encouraged support for, rather than opposition to, mandatory minimum sentencing. Sympathy and pity centered either on sympathy and pity for Janet Smith herself (e.g., the participant mentioned Smith in the open-ended response) or on sympathy and pity for "people" who were given what the participants saw as unfair and long sentences under the mandatory minimum guidelines.[5] Furthermore, no differences were found in the nature of the responses between the two episodic frame conditions that might account for differences in views on mandatory minimum sentencing. The two episodic conditions were equally likely to elicit expressions of sympathy and pity targeted at Smith individually. Thus, any differ-

[5] Not surprisingly, participants in the thematic frame condition had more expressions of sympathy directed at people generally, whereas those in the episodic frame conditions directed their sympathy to Smith. This seemed to be an informational effect, not a framing effect. I found no difference in the nature of sympathy responses across experimental groups after merging these categories (thinking about them as expression of sympathy for those receiving unfair sentences). The content of anger and disgust responses did not suggest that the experience of these emotions should be associated with decreased opposition to mandatory minimum sentences and thus would explain the pattern of support shown in Table 2. More detail on the content of the open-ended responses is available from the author.

ences in policy attitudes between the two episodic conditions cannot be explained by differences in the nature of the emotional response when faced with the story of a white mother versus a black mother.

There are two other possible explanations for the pattern of opinion across experimental conditions. Emotional response may be epiphenomenal; the emotional reactions generated by the episodic frames may not translate into effects on policy views as hypothesized. Alternatively, the effect of the frames may operate along both cognitive and affective tracks but in distinct ways. In addition to direct effects, frames may have had indirect effects on policy views through their effects on emotional response. In other words, emotional response may mediate the relationship between frames and policy views but it may not completely mediate the relationship.

In order to assess if and how frames and emotion together affect opinion on mandatory minimum sentencing, I estimated two ordered probit equations for each experiment. The first included only dummy variables for each of the episodic frame conditions (the thematic frame condition was the excluded category) to replicate the results in Table 2. The second included dummy variables for the episodic frame conditions and my combined measures of strength of emotional response. These results are presented in Table 3.[6]

Those who expressed empathy were significantly more opposed to mandatory minimum sentencing than those who did not, as one would expect given that sympathy and pity were targeted at the plight of Smith and others who might receive such sentences. Moreover, once I accounted for the effects of emotional response in determining opinion, I found that those in the episodic frame conditions were actually significantly less opposed to mandatory minimum sentencing than those in the thematic frame condition. This was particularly so in the case of those who read the episodic frame in which Janet Smith was described as a black mother.

The results suggest emotion was not epiphenomenal but neither did it carry the full effect of the frames. In these experiments, the partial effect (controlling for emotional response) of the episodic frame conditions was to lessen the persuasive effect of the column relative to the thematic frame. In other words, the episodic frames were less likely to produce opposition to mandatory minimum sentencing. However, episodic frames also produced greater empathy (Table 1) and that empathy was associated with greater opposition to mandatory minimum sentencing (Table 3). These results suggest that the effect of these frames may operate along two separate tracks—cognitive and affective.

[6] Because the patterns of emotional response across framing conditions were similar for related emotions (e.g., sympathy and pity, see Table 1) and the correlations between related emotions were high, I used the combined emotion variables (empathy, aversion, and anxiety scales) in these models. When I estimated the models using the individual emotion questions, I found significant effects for sympathy, pity, and anger in the first experiment. In the second experiment, I found significant effects for pity and worry (the effect for sympathy was significant at the $p = .10$ level).

Table 3. How Emotional Response and Frame Condition Effect Views on Mandatory Minimum Sentencing

	Experiment 1		Experiment 2	
Frame				
Episodic Frame with Black Individual	−.45 (.24)*	−1.10 (.28)***	−.35 (.25)	−.94 (.31)***
Episodic Frame with White Individual	−.17 (.25)	−.60 (.27)**	−.13 (.26)	−.45 (.29)
Emotion				
Empathy Scale (sympathy & pity)		1.28 (.48)***		1.77 (.51)***
Aversion Scale (anger & disgust)		1.43 (.41)***		.75 (.46)
Anxiety Scale (worry and fear)				−.92 (.56)
Cut1	−2.18	−1.48	−1.95	−1.73
Cut2	−1.27	−.52	−1.18	−.92
Cut3	−.66	.21	−.54	−.24
Cut4	.30	1.42	.70	1.15
Pseudo R2	.01	.13	.01	.07
N	116	116	102	101

Note. Table entries are ordered probit coefficients with standard errors in parentheses. Opinion on mandatory minimum sentencing is coded to range from 0 to 1 where 1 represents those who strongly oppose mandatory minimum sentences for first time drug offenders. The thematic frame is the excluded category.
*p < .10, **p < .05, ***p < .01.

The indirect effect of the frame via emotional response suppressed the direct effect of the frame. Sobel tests for mediation suggest that these indirect effects of the episodic frames on opinion were significant. I separately compared the effect of the episodic frame featuring a white individual to the thematic frame and the effect of the episodic frame featuring a black individual to the thematic frame. In both experiments, the indirect effect of the episodic frames through empathy was significantly different from zero. The effect on opposition to mandatory minimum sentencing of the episodic frame with a white defendant was mediated by empathy ($z = 2.68$, $p < .01$ for experiment 1; $z = 2.56$, $p < .05$ for experiment 2). Empathy also mediated the effect of the episodic black frame ($z = 2.91$, $p < .01$ for experiment 1; 2.19, $p < .05$ for experiment 2).[7] Traditionally, tests for mediation following Baron and Kenny (1986) require that the independent variable significantly affects the mediator. Table 1 demonstrated that the use of the story of a particular individual harmed by the policy did influence emotional response, specifically empathy. It also requires, as a first step, estimating the relationship between the independent variable and the dependent variable to show that there is something to be mediated. Shrout and Bolger (2002) suggest this is not required when suppression is present as is the case here.

Why might the direct effect of the episodic frame conditions dampen the persuasive effect of the column? Iyengar's work suggests one reason why the

[7] In the first experiment, the expression of aversion also varied by frame and I found aversion did mediate the effect of the episodic black frame ($z = 2.08$, $p < .05$).

episodic frames may not produce as much opposition as thematic frames. By centering attention on Janet Smith, these frames generate sympathy but they may also encourage individualistic attributions and less willingness to see a need for policy change. Furthermore, this was a conditional effect once I took account of emotional response. Thus, my results suggest that episodic and thematic frames can exert influence on policy views through their effect on emotions in ways that may not necessarily reinforce the influence through attribution.

An alternative possibility is that episodic frames are less persuasive than the thematic frame because the arguments are processed with less depth. Although I cannot directly test this interpretation, the second experiment did include a series of questions asking participants to assess their attention to, interest in, and impartiality toward the article. I found those who read the column featuring the story of Janet Smith reported greater attention (depth scale $F = 3.89$, $p < .05$) and less distraction (ability scale $F = 5.97$, $p < .01$), though no differences were found in reported interest. This pattern across frames indicates that the more emotionally engaging episodic frames draw the reader in, generating greater attention, not less. This is consistent with work in psychology that argues people pay more attention to personalized case history information (Nisbett & Ross, 1980) and appears to be at odds with the possibility that the arguments are processed with less depth. Yet, I have no way of knowing if this attention and interest was centered on Smith's particular plight, making the more general arguments against mandatory minimum sentencing less of a focus (Smith & Shaffer, 2000). More importantly, these questions do not actually test depth of processing or learning but rather respondents' perceptions of how much attention they paid to the article (see question wording in Appendix B). Research has shown that interest and motivation do not necessarily result in increased learning under certain conditions. For example, in an examination of the effects of anxiety on political learning in the context of the Iraq War, Feldman and Huddy (2005) show that anxiety can enhance motivation (measured as talking and thinking about the war) and increase exposure to news without increasing learning.

Discussion

The results presented here indicate that rhetorical devices such as episodic framing influence emotional response. This finding is consistent with others who demonstrate affective effects of context or policy-specific frames (e.g., Brewer, 2001; Gross and D'Ambrosio, 2004). Taken together, this research shows emotion is worthy of inclusion in our accounts of framing.

More importantly, the results presented here are consistent with the claim that episodic framing operates via both affective and cognitive routes. In the case at hand, I found the conditional effect of the episodic frames was to dampen the persuasive effect of the column, generating less opposition to mandatory minimum sentencing than a thematic frame. Yet these episodic frames also generated sym-

pathy for Janet Smith, and that sympathy led to greater opposition to mandatory minimum sentencing. The persuasive effect of the episodic frames in this instance would seem to run through their ability to engender sympathy and pity. This would be consistent with the claim that emotional reactions serve as information that individuals use in coming to their opinions. While the affective and cognitive effects of a frame or rhetorical appeal could certainly be reinforcing rather than countervailing, the finding that the direct and indirect effects work in opposite directions in this case lends support to the conclusion that these may be distinct routes for framing influence.

These findings seem to both reinforce and extend Iyengar's work in *Is Anyone Responsible?* Episodic frames appear to minimize attitude change by focusing on individual rather than societal forces (Iyengar, 1991). However, episodic frames can actually increase persuasion if the individual's story is compelling enough to generate intense emotional reactions from a significant portion of the audience. In other words, the value of episodic framing may be conditional on the ability of the subject featured in the frame to generate particular emotions. A story that generates incredible anger at how an individual has been treated by the government or that generates incredible sympathy and pity for an individual's plight can facilitate persuasive claims on behalf of policy change. By contrast, when the central character in a story does not arouse persuasive emotions as intended by the communicator, the persuasive appeal is more likely to fail.

The distinction that emerges between the two episodic frame conditions cannot be explained by a lack of sympathy among those reading about a black Janet Smith (if anything, they report greater sympathy and pity) or by differences in the target of that sympathy (in open-ended questions participants in the two different episodic frame conditions are equally likely to say they felt sympathy toward Smith). It would seem, then, that the difference between these frames operates on a cognitive level, though work needs to be done to disentangle this.

It is also important to raise the possibility these may be short-term effects. While thematic frames may prove more persuasive in the short run, as demonstrated in the experiments presented here, the emotional engagement of the episodic frames might produce persuasive effects over a longer period. An individual case like Janet Smith's may be remembered long after the arguments of the column fade precisely because the story is more emotionally engaging. I do not have a way to test this possibility in this study.

Given my use of student samples, I must be careful in generalizing from this study. In particular, both student samples are more democratic and liberal than the general population. As such, the participants in this study may be particularly disposed to agree with the arguments in the column, as liberals tend to be less supportive of mandatory minimum sentencing.[8] Such a sample may also be pre-

[8] Unfortunately, I do not have enough cases to analyze and test the models separately for democrats and republicans.

disposed to sympathy and pity for the woman used to illustrate the problem in the episodic conditions. The experience of emotion or the content of the emotional response may have been different with a different population. For example, an overwhelmingly conservative sample, whose predispositions inclined them against mandatory minimums, might have reported less sympathy and more anger and disgust in general or might have reported anger and disgust directed in different ways (e.g., more anger and disgust directed at Smith's actions or at the notion that mandatory minimums might be changed). By the same token, the effects may be particular to the issue of mandatory minimum sentencing or the specific exemplar employed in the episodic frames. The content of the stimulus matters, but predispositions likely matter as well, in determining emotional reactions (Gross & D'Ambrosio, 2004). Thus, further work with more diverse samples is needed to show how emotional response is contingent on both the nature of the frame and an individual's predispositions toward the issue at hand.

Although this study only explored the effect of a particular set of episodic and thematic frames on emotional response, the results have some practical implications. Partisans assume that emotional appeals are an important element of any persuasive message. The results here support the notion that episodic frames are more emotionally engaging, but this alone does not make them a more compelling rhetorical device for someone seeking to change opinion. Advocates must carefully select the individual stories they choose to feature if they want to facilitate persuasive appeals through effects on emotion. In short, the choice of how to frame a story may have important consequences for how citizens feel and think about the policies implicated by that story.

Appendix A

In each case, the treatments were formatted to appear as if they were real newspaper articles. The text of the episodic frame stories is exactly the same except the fourth sentence which either describes her as "a single, white mother" or "a single, black mother." The italicized text is present in both conditions.

Episodic Frame (White)

Frederick Jackson
The Case Against Mandatory Minimum Sentencing

Janet Smith stands for the controversy surrounding federal drug-sentencing laws passed in the 1980s.

In 1995, a federal judge sentenced Smith to 24.5 years in prison without parole. Smith did not murder anyone. She wasn't a major crime figure. She was a single, white mother in her 20s with the bad judgment to have fallen for a drug dealer who abused her.

The middle-class only daughter of a schoolteacher and an accountant, Smith was a college sophomore when she became involved in a relationship with Peter Hall. Hall turned out to be a drug dealer in a murderous East Coast crack ring. Smith helped Hall in a number of ways, bailing him out of jail, carrying money for him, lying to authorities. By the time she was ready to cooperate, Hall was dead and she no longer had information to trade for a reduced sentence.

She pleaded guilty to three charges including conspiring to help Hall with crack and powder cocaine trafficking.

A fringe player in a crack cocaine ring and a first time nonviolent offender—Smith's penalties were greater than the average state sentence for murder or voluntary manslaughter. It will cost close to half a million dollars to keep Smith behind bars.

Clearly, Janet Smith was no angel. She broke a number of laws and deserved punishment. But 24.5 years is a disproportionately harsh sentence for a woman with no prior criminal history, who had arguable extenuating circumstances mitigating the severity of her conduct and who had never actually handled drugs.

Even a judge who rejected one of Smith's appeals wrote that her prison sentence was "truly heavy" and represented "the unintended consequences of Congressional legislation." But his hands were tied.

Under current law, judges have little discretion over whether a drug offender will be imprisoned and for how long. Instead, they must operate within a range of minimum and maximum sentences that fail to take account of whether the crime involved violence or whether there are mitigating circumstances. Only the government can seek a reduction in the minimum sentence. Since Smith failed to cooperate until it was too late, prosecutors never asked for a reduction.

Monica Pratt, spokeswoman for Families Against Mandatory Minimums, explains, "The more information you have to trade, the more information you can give to prosecutors to reduce your sentence. The less information you have to trade, the less of a chance you will have to reduce your sentence." So federal and state prisons are full of low-level nonviolent drug offenders instead of drug kingpins.

According to 1993 statistics from the Department of Justice, low-level players—first-time, nonviolent offenders whose criminal activity was not sophisticated—numbered over one-third of those in federal prison on drug charges. Under mandatory minimums the prison population is exploding and prison costs are skyrocketing.

It is time to end federal mandatory minimums for drug offenses. As Supreme Court Chief Justice William H. Rehnquist argues, "mandatory minimums impose unduly harsh punishment for first-time offenders and have led to an inordinate increase in the prison population." Former Attorneys General Janet Reno and Edwin Meese, the American Bar Association, and judges across the nation have all called for a reassessment of mandatory minimums. It is time for Congress to listen. The alternative is more wasted lives and wasted dollars.

Thematic Frame

Frederick Jackson

The Case against Mandatory Minimum Sentencing

Congress responded to the crack cocaine epidemic of the mid-1980s with get-tough measures like mandatory minimum sentencing for drug offenses. It is now clear that mandatory minimums and their ripple effects are not punishing the major drug player they were intended for. Instead we have a system where first time nonviolent offenders can receive penalties greater than the average state sentence for murder or voluntary manslaughter.

Under current law, judges have little discretion over whether a drug offender will be imprisoned and for how long. Instead, they must operate within a range of minimum and maximum sentences.

The federal mandatory minimums are determined by the amount of drugs—for example, a 10-year sentence is imposed for possession of 1,000 marijuana plants, while 5 grams of crack cocaine will send a defendant to jail for five years. They *fail to take account of whether the crime involved violence or whether there are mitigating circumstances.*

Under this system, *only the government can seek a reduction in the minimum sentence*. These restrictions give prosecutors greater control of cases. When prosecutors are in charge, the importance of cooperating with investigators is magnified.

Monica Pratt, spokeswoman for Families Against Mandatory Minimums, explains, "The more information you have to trade, the more information you can give to prosecutors to reduce your sentence. The less information you have to trade, the less of a chance you will have to reduce your sentence."

Because those who are more involved have more information to trade, it is the drug users and those who are caught up with the actions of loved ones who are put into jail. *In fact, federal and state prisons are full of low-level nonviolent drug offenders instead of drug kingpins.*

According to 1993 statistics from the Department of Justice, low-level players—first-time, nonviolent offenders whose criminal activity was not sophisticated—numbered over one-third of those in federal prison on drug charges.

Under mandatory minimums the prison population has exploded and prison costs are skyrocketing. The national crime rate has been dropping for seven years, yet more Americans are going to jail than ever before. The number of prisoners nationwide has more than tripled over the past 20 years, according to Justice Department statistics. More than half of these prisoners were locked up for non-violent crimes, most of them drug driven.

States and the federal government are spending $31 billion a year on corrections. From 1987 to 1995 money spent by states on prisons rose by 30% while expenditures for universities dropped by 19%. This is not what we intended with the War on Drugs.

It is time to end federal mandatory minimums for drug offenses. As Supreme Court Chief Justice William H. Rehnquist, argues "mandatory minimums impose unduly harsh punishment for first-time offenders and have led to an inordinate increase in the prison population." Former Attorneys General Janet Reno and Edwin Meese, the American Bar Association and judges across the nation have all called for a reassessment of mandatory minimums. It is time for Congress to listen. The alternative is more wasted lives and wasted dollars.

Appendix B

Measuring Policy Views

In both experiments the following question was used to measure opinion on mandatory minimum sentencing: "Do you favor or oppose mandatory minimum sentences for first or second time non-violent drug offenses?" Responses were coded to range between 0 and 1 as follows: strongly favor (0), favor (.25), neutral (.50), oppose (.75), strongly oppose (1).

Measuring Emotional Response

The measures of emotional response varied slightly between the two experiments. In the first experiment, respondents were asked for their emotional reactions to the articles on a five point scale ranging from "none" to "a lot" for four different emotions (anger, sympathy, disgust, and pity) using the following question: "How much anger did you feel while reading the article?" This was recoded to range between 0 and 1, where 0 represents those who said "none" and 1 represents those who said "a lot."

Subjects in the second experiment were asked whether or not they felt a series of emotions—anger, sympathy, worry, disgust, pity, and fear—while reading the article. These questions took the following form: "While reading the article did you feel angry?" Those who said yes were asked to answer two follow-up questions: They were asked about the strength of their emotional response measured on a ten point scale ranging from "only a little angry" to "extremely angry"; and were asked about the object or impetus for their emotional response ("If yes, why did you feel this? What was it that made you feel angry?"). The two closed-ended questions were combined to create a measure of strength of emotional response that ranged between 0 and 1. This was done as follows: Those who reported they did not feel the emotion when asked the first binary (yes-no) emotional response question were coded as 0 on the new measure of emotional response. For those who said they did feel the emotion on the initial binary emotional response measure, I recoded the strength of emotional response to range from .10 (for those who selected a 1, "only a little angry") to 1 (for those who selected a 10, "extremely angry"). I also created

three scales for use in the analysis: an empathy scale which combined sympathy and pity (Pearson correlation .77 in experiment 1 and .50 in experiment 2); an aversion scale which combined anger and disgust (Pearson correlation .71 in experiment 1 and .56 in experiment 2); and an anxiety scale (combining fear and worry, Pearson correlation .65 experiment 2 only).

How one measures the experience of emotion does affect the number of people claiming to have experienced the emotion. When offered the five point scale anchored 1 "none" and 5 "a lot" in experiment 1 fewer individuals report feeling no emotion (e.g., select "none") than report not feeling the emotion when presented with the yes or no question in experiment 2. This change in question wording does result in different mean levels of emotional response between the two experiments. However, I am interested in variations across treatments not differences in the level of emotion between the two experiments.

Detail on the coding of the open-ended items measuring emotion is available from the author. Open-ended items measuring emotion are rarer than closed-ended items measuring emotion at least partly because of the demands they place on researchers (e.g., the time-consuming reading and coding that must be done to make them useful as data). However, such items allowed me to more fully understand emotional reactions by helping me to get a better handle on their meaning. Individuals might report feeling the same emotion but for very different reasons. For example in the case at hand, one might express disgust with mandatory minimum sentencing because it forced someone like Smith to spend such a long time in jail or express disgust with the article for trying to portray Smith as deserving of a lesser sentence. Wilson and his colleagues raise concern about the use of open-ended responses as legitimate measures (Erber, Hodges, & Wilson 1995; Nisbett & Wilson, 1977; Wilson, Dunn, Bybee, Hyman, & Rotondo, 1984). They suggest that people have little ability to understand their mental processes and thus the true underlying reasons for their attitudes (or in this case emotion). Their concerns suggest caution in interpreting the meaning of the open-ended measures.

Measures of Message Processing

A series of questions designed by Wolski and Nabi (2000) to measure different aspects of message processing were included in the second experiment. Respondents were asked "Still thinking about the article on mandatory minimum sentencing, please tell us how much you agree or disagree with the following statements." These 16 statements were recoded into four scales: motivation, ability, depth, and bias. Motivation Questions: This issue is interesting to me; I was interested in what the author had to say; I don't find this issue very interesting; I was motivated to read this article (Scale alpha = .79). Depth Questions: I focused on the arguments the author made; While reading the article, I paid close attention to each point that was made; I didn't pay close attention to the author's arguments; I concentrated on the article arguments (Scale alpha = .83). Ability Questions: My

mind kept wandering as I read the message; While reading, I didn't let myself get distracted from focusing on the article content; While reading the message, thoughts about other things kept popping up in my head; My mind did not wander as I read the article (Scale alpha = .86). Bias Questions: I remained objective about the article content; My prior beliefs about the issue prevented me from being objective; I tried not to let how I feel about the issue influence how I read the article; I tried to remain impartial as I read the article. (Scale alpha = .79). Motivation, depth, and ability scales are highly correlated (Pearson correlation ranges from .48 to .59), but bias scale is not particularly correlated with the others (−.19 with motivation, −.01 with depth, and .07 with ability).

ACKNOWLEDGMENTS

The author would like to thank Leonie Huddy, Paul Brewer, James Druckman, Jarol Manheim, and three anonymous reviewers for their helpful comments on earlier versions of this paper. Correspondence concerning this article should be addressed to Kimberly Gross, School of Media and Public Affairs, George Washington University, 805 21st Street NW, Suite 400, Washington, DC 20051. Email: kimgross@gwu.edu

REFERENCES

Abelson, R. P., Kinder, D. R., Peters, M. D., & Fiske, S. T. (1982). Affective and semantic components in political person perception. *Journal of Personality and Social Psychology, 42*, 619–610.

Allen, M., & Preiss, R. W. (1997). Comparing the persuasiveness of narrative and statistical evidence using meta-analysis. *Communication Research Reports, 14*, 125–131.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173–1182.

Brader, T. (2005). Striking a responsive chord: How political ads motivate and persuade voters by appealing to emotions. *American Journal of Political Science, 49*, 388–405.

Brader, T. (2006). *Campaigning for hearts and minds: How emotional appeals in political ads work*. Chicago: University of Chicago Press.

Brewer, P. (2001). Value words and lizard brains: Do citizens deliberate about appeals to their core values? *Political Psychology, 22*, 45–64.

Conover, P. J., & Feldman S. (1986). Emotional reactions to the economy: I'm mad as hell and I'm not going to take it anymore. *American Journal of Political Science, 30*, 50–78.

Druckman, J. N. (2001). The implications of framing effects for citizen competence. *Political Behavior, 23*, 225–256.

Ellsworth, P. (1991). Some implications of cognitive appraisal theories of emotion. In K. T. Strongman (Ed.), *International review of studies on emotion* (Vol. 1, pp. 143–161). New York: Wiley.

Erber, M. W., Hodges, S. D., & Wilson, T. D. (1995). Attitude strength, attitude stability, and the effects of analyzing reasons. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 431–454). Mahwah, NJ: Lawrence Erlbaum Associates.

Feldman, S., & Huddy, L. (2005). "The paradoxical effects of anxiety on political learning." Presented at the 2005 Annual Meeting of the Midwest Political Science Association.

Fiske, S., & Taylor, S. (1991). *Social cognition* (Second edition). New York: McGraw-Hill Inc.

Gamson, W. A., & Modigliani, A. (1987). The changing culture of affirmative action. In R. Braungart (Ed.), *Research in political sociology* (Vol. 3, pp. 137–177). Greenwich, CT: JAI Press.

Gilliam, F. D., & Iyengar, S. (2000). Prime suspects: The influence of local television news on the viewing public. *American Journal of Political Science, 44*, 560–573.

Gross, K. A., & D'Ambrosio L. A. (2004). Framing emotional response. *Political Psychology, 25*, 1–29.

Huddy, L., Feldman, S., Taber, C., & Lahev, G. (2005). Threat, anxiety, and support for antiterrorism policies. *American Journal of Political Science, 4*, 593–608.

Iyengar, S. (1991). *Is anyone responsible? How television frames political issues*. Chicago: University of Chicago Press.

Iyengar, S., & Kinder, D. R. (1987). *News that matters: Television and American opinion*. Chicago: University of Chicago Press.

Jacoby, W. G. (2000). Issue framing and public opinion on government spending. *American Journal of Political Science, 44*, 750–767.

Kinder, D. R. (1994). Reason and emotion in American political life. In R. C. Schank and E. J. Langer (Eds.), *Beliefs, reasoning and decision-making: Psycho-logic in honor of Bob Abelson* (pp. 277–314). Hillsdale, NJ: Lawrence Erlbaum.

Kinder, D. R., & Sanders, L. M. (1990). Mimicking political debate with survey questions: The case of white opinion on affirmative action for blacks. *Social Cognition, 8*, 73–103.

Lazarus, R. S. (1991). *Emotion and adaptation*. New York: Oxford University Press.

Lerner, J. S., & Keltner, D. (2000). Beyond valence: Toward a model of emotion-specific influences on judgment and choice. *Cognition and Emotion, 14*, 473–493.

Marcus, G. E., & MacKuen, M. B. (1993). Anxiety, enthusiasm, and the vote: The emotional underpinnings of learning and involvement during presidential campaigns. *American Political Science Review, 87*, 672–685.

Marcus, G. E., Neuman, W. R., & MacKuen, M. (2000). *Affective intelligence and political judgment*. Chicago: University of Chicago Press.

Masters, R. D., & Sullivan, D. G. (1993). Nonverbal behavior and leadership: Emotion and cognition in political information processing. In S. Iyengar & W. J. McGuire (Eds.), *Explorations in political psychology* (pp. 150–182). Durham, NC: Duke University Press.

Nabi, R. L. (1998). The effect of disgust-eliciting visuals on attitudes toward animal experimentation. *Communication Quarterly, 46*, 472–484.

Nelson, T. E., Clawson, R. A., & Oxley, Z. M. (1997). Media framing of a civil liberties conflict and its effect on tolerance. *American Political Science Review, 91*, 567–583.

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231–259.

Ortony, A., Clore, G. L., & Collins, A. (1988). *The cognitive structure of emotions*. Cambridge: Cambridge University Press.

Roseman, I. J. (1991). Appraisal determinants of discrete emotions. *Cognition and Emotion, 5*, 161–200.

Sears, D. (1986). College sophomores in the laboratory: Influences of a narrow database on social psychology's view of human nature. *Journal of Personality and Social Psychology, 51*, 515–530.

Schwarz, N. (1990). Feelings as information: Information and motivational functions of affective states. In E. T. Higgins & R. M. Sorrentino (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (Vol. 2, pp. 527–561). New York: Guilford Press.

Schwarz, N. (2000). Emotion, cognition and decision making. *Cognition and Emotion, 14*, 433–440.

Schwarz, N., & Clore, G. L. (1983). Mood, misattribution and judgments of well being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology, 45*, 513–523.

Smith, S. M., & Shaffer, D. R. (2000). Vividness can undermine or enhance message processing: The moderating role of vividness congruency. *Personality and Social Psychology Bulletin, 26* (7), 769–779.

Sniderman P. M., & Theriault, S. M. (2004). The structure of political argument and the logic of issue framing. In W. E. Saris & P. M. Sniderman (Eds.), *Studies in public opinion* (pp. 133–165). Princeton: Princeton University Press.

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods, 7*, 422–445.

Wilson, T. D., Dunn, D. S., Bybee, J. A., Hyman, D. B., & Rotondo, J. A. (1984) Effects of analyzing reasons on attitude-behavior consistency. *Journal of Personality and Social Psychology, 47*, 5–16.

Wolski, S., & Nabi, R. (2000). Message processing quality: Confirmatory analysis of an elaboration depth measure. Paper presented at the annual meeting of the International Communication Association, Acapulco, Mexico.

Playing the Race Card in the Post-Willie Horton Era: The Impact of Racialized Code Words on Support for Punitive Crime Policy

Author(s): Jon Hurwitz and Mark Peffley

Source: *The Public Opinion Quarterly* , Spring, 2005, Vol. 69, No. 1 (Spring, 2005), pp. 99-112

Published by: Oxford University Press on behalf of the American Association for Public Opinion Research

Stable URL: https://www.jstor.org/stable/3521605

Public Opinion Quarterly, Vol. 69, No. 1, Spring 2005, pp. 99–112

PLAYING THE RACE CARD IN THE POST-WILLIE HORTON ERA

THE IMPACT OF RACIALIZED CODE WORDS ON SUPPORT FOR PUNITIVE CRIME POLICY

JON HURWITZ
University of Pittsburgh
MARK PEFFLEY
University of Kentucky

Abstract To date, little is known about the precise impact of racially coded words and phrases. Instead, most of what we know about racialized messages comes from studies that focus on pictorial racial cues (for example, the infamous "Willie Horton" ad) or on messages with an extensive textual narrative that is laced with implicit racial cues. Because in a "post-Horton" era strategic use of racially coded words will often be far more subtle than those explored in past studies, we investigate the power of a single phrase believed by many to carry strong racial connotations: "inner city." We do so by embedding an experiment in a national survey of whites, where a random half of respondents was asked whether they support spending money for prisons (versus antipoverty programs) to lock up "violent criminals," while the other half was asked about "violent *inner city* criminals." Consistent with the literature on issue framing, we find that whites' racial attitudes (for example, racial stereotypes) were much more important in shaping preferences for punitive policies when they receive the racially coded, "inner city" question. Our results demonstrate how easy it is to continue "playing the race card" in the post–Willie Horton era, as well as some of the limits of such framing effects among whites with more positive racial attitudes.

In 1994, during the House debate on the Violent Crime Control and Law Enforcement Bill supported by President Bill Clinton, House Republicans

The order of the authors' names is alphabetical. This material is based upon work supported by the National Science Foundation under Grant No. 9906346. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The authors would like to thank Nicholas Valentino and David Barker for helpful comments on an earlier draft. Address correspondence to Mark Peffley; e-mail: mpeffl@uky.edu.

doi:10.1093/poq/nfi004

seized on a minor provision of the legislation that designated block grants for midnight basketball programs, which were designed to provide recreational activities for inner-city youth. Even though the basketball line item represented less than two-hundredths of 1 percent of the bill's expenditures (the majority of which were earmarked for punitive anticrime measures), 29 Republican legislators spoke derisively about midnight basketball on the House floor between August 3 and August 21 before the bill's passage, characterizing the program as "hugs for thugs." A few Democrats charged Republicans with "playing the race card," or using coded language to racialize the bill. Knowing that the American public sees professional basketball as a sport played overwhelmingly by African-Americans, Republicans, according to Democrats, attempted to portray the legislation as something to coddle black, inner-city youth.

This example neatly raises the central issue of the present inquiry: to what degree does such language affect the audience's expression of support for the policy? While numerous studies have investigated the impact of racially coded language on candidate preferences and vote intentions (Mendelberg 2001; Valentino, Hutchings, and White 2002) and on a slew of race-related policy attitudes, little is known about the power of such language to affect policy attitudes in the one domain in which, intuitively, it should have its most profound impact—crime.[1] To the degree that citizens conflate race and crime, and to the extent that individuals consider crime to be highly salient, politicians can be expected to manipulate public attitudes by injecting race into this very emotional arena.

Public debate pursuant to "playing the race card" began with the airing of the infamous "Willie Horton" ad, run by the National Security Political Action Committee (NSPAC) against Democrat Michael Dukakis during the 1988 presidential campaign. As meticulously detailed by Jamieson (1992, pp. 15–42), the narrator of the spot states that Willie Horton, a convicted murderer, received multiple weekend furlough passes from prison, during the last of which, the narrator informs us, he "fled, kidnapping a young couple, stabbing the man and repeatedly raping his girlfriend." While the ad could have conveyed exactly the same information without graphics, NSPAC elected to superimpose the most menacing possible picture of Horton, an African-American, over the narrative.

Tali Mendelberg (2001) convincingly argues that the Horton ad is effective because of its implicitness. White Americans, despite their resentment toward blacks, are committed to a "norm of equality," which causes them to reject blatantly racial appeals, but not those that are implicit—that is, those not recognized as racial. The NSPAC spot fulfills the implicitness requirement in the sense that it never explicitly mentions that the subject is African-American,

1. Mendelberg (2001) found that exposure to news about the Willie Horton ad, used during the 1988 presidential campaign, strengthened the connection between racial attitudes and opposition to economic policies (e.g., welfare), but not crime policy.

nor does it make an explicit linkage between candidate Dukakis, his policies, and the black recipient (Horton) of his policy.

We believe that the Horton ad has made it more difficult to play the race card in campaigns. Put differently, it may have helped transform what Mendelberg terms an "implicit" appeal into a more explicit appeal that would, doubtless, be recognized for its blatant racial component today.[2] While, technically, the NSPAC commercial may have been implicit, racialized content can be, and in the future is likely to be, far more subtle and implicit. Valentino, Hutchings, and White (2002) demonstrate convincingly that race-related pictures that are *far* more subtle than that in the Horton spot can strongly influence both individuals' racial policy positions and their candidate preferences.

One must wonder if the norm of political correctness and the fear of being accused of racialization have driven racial appeals even further underground, possibly to the point where virtually *any* presentation of race—in audible or printed narrative or in pictures—has become practically verboten. If so, the racialization of politics will increasingly take place mainly at the level of code words, or words that are fundamentally nonracial in nature that have, through the process of association, assumed a strong racial component. Martin Gilens (1996), for example, has argued that "welfare" is one such word.

To date, little is known about the impact of racially coded words and phrases. Instead, most of what we know about racialized messages comes from studies whose major focus is on pictorial racial cues (for example, Mendelberg 2001; Valentino, Hutchings, and White 2002) or an extensive textual narrative (for example, a political ad or news segment) that is laced with implicit racial cues (Valentino, Hutchings, and White 2002). However, we presuppose that modern battles will be far less obvious in nature and will rely instead on the subtle introduction of strategic words or phrases with racial connotations.

What follows is a systematic analysis of the impact of a phrase believed by many to carry strong racial connotations: "inner city." After discussing the cognitive process by which racial coding may be effective, we use a survey experiment embedded within the 2001 National Race and Crime Survey to examine not only how this phrase influences individuals' beliefs about anticrime policies but also, and more important, how they reach their decisions regarding these policies. We undertake this analysis precisely because the implications of coding are, to the degree that the technique succeeds, both frightening and potentially pervasive.

2. For example, even though crime and prison furloughs were featured in political ads in statewide and national election campaigns in 1994, the ads avoided any reference (pictorial or otherwise) to race because "[candidates knew that if they did] they'd spend weeks responding to the counterattack" (Kurtz 1994).

Group-Centric Framing

Studies of issue framing (Druckman 2001; Nelson, Clawson, and Oxley 1997; Nelson and Kinder 1996) demonstrate that frames affect opinions by making certain considerations (for example, values, predispositions, groups) seem more important than others, thus affecting the way people judge the issue. If, for example, a narrative about a Ku Klux Klan rally is framed as a story about First Amendment rights (rather than as a matter of public order), then individuals are more likely to base their decision to tolerate such a rally on their support for civil liberties, while they are more likely to base their decision on concerns for public order if the same story is framed in terms of safety (Nelson, Clawson, and Oxley 1997). In essence, framing not only creates a tighter linkage between the consideration and the policy attitude, but it also elevates the importance of the consideration as a decisional criterion.

According to Nelson and Kinder (1996, pp. 1055–56), "Public opinion on matters of government policy is group-centric: shaped in powerful ways by the attitudes citizens possess toward the social groups they see as the principal beneficiaries (or victims) of the policy." They further argue that framing is particularly effective at heightening group-centrism—or the tendency to base policy decisions on the group given prominence in the frame. Not surprisingly, then, recent research has linked support for ostensibly nonracial policies, such as welfare (Gilens 1999; Peffley, Hurwitz, and Sniderman 1997) and crime (Gilliam and Iyengar 2000; Hurwitz and Peffley 1997; Peffley and Hurwitz 2002) to attitudes toward African-Americans.

As noted, at least in the racial domain, framing may be less effective as a determinant of group-centric policy judgments when it is explicit (Mendelberg 2001). Unfortunately, virtually all of the growing literature from both the framing and priming paradigms has examined the power of either extensive textual narratives, such as 30-second "spot" ads, or visuals to frame a message in a particular way.[3] We also maintain that modern politics in the United States has seen an increase in more subtle means to frame racial messages. It is necessary, therefore, to examine the power of single words or phrases that have no explicit racial content.

Our method, more specifically, involves asking respondents about the preferred way to prevent crime (spending money for prisons or for antipoverty programs), while randomly asking one-half about "violent criminals" and the other half about "violent *inner-city* criminals." In our analysis, the following question serves as the independent variable:

> Some people want to increase spending for new prisons to lock up violent [inner-city/xxx] criminals. Other people would rather spend this money for antipoverty programs to prevent crime. What about you? If you had to choose, would you

3. On framing, see Nelson, Clawson, and Oxley 1997; Nelson and Kinder 1996; Nelson and Oxley 1999. On priming, see Iyengar and Kinder 1987; Valentino, Hutchings, and White 2002.

rather see this money spent on building new prisons, or on antipoverty programs? Do you feel strongly or not very strongly about this?

Notably, the frame manipulation is restricted to a single phrase: "inner-city." Because everything else in the question is identical, any differences between the control and treatment groups can only be attributed to the inclusion of the inner-city frame.

Based on the framing literature, and more specifically on that which delves into the question of group-centric framing (Nelson and Kinder 1996), we anticipate finding two differences between experimental groups. First, we expect the frame to alter the evaluative criteria by which respondents select their preferred policy. Given the modern association between inner cities and African-Americans in the minds of many individuals,[4] those in the racially coded, inner-city frame should be more likely to base their policy decisions on their evaluations of African-Americans (H_1). Second, we expect more punitive responses (use the money for prisons) from white respondents in the racially coded, inner-city condition, at least to the degree that they view African-Americans, who are the presumptive inhabitants of inner cities, negatively (H_2). There is ample evidence that many whites who hold racially prejudicial views favor harsher responses to crime when the criminals are identified as black (Hurwitz and Peffley 1997). When the policy is framed with racial connotations, therefore, we can expect these respondents' attitudes to become more punitive.

We underscore the subtlety of the frame in this experiment, which is precisely why we consider it to be the appropriate methodology in the post-Horton era. We cannot, of course, claim that the inner-city frame is the most subtle of all possible racial frames, as evidenced by our finding (see footnote 4) that many clearly equate the inner city with blacks. Doubtless, media consultants will introduce increasingly clever and subtle racial language to future campaigns. However, the frame is, by definition, implicit inasmuch as there is no explicit mention or graphic representation of race. Mendelberg (2001, p. 11) defines an implicitly racial appeal as one that "contains a recognizable—if subtle—racial reference, most easily through visual references." Surely, our frame is far more implicit than a visual reference and far more subtle than some of the other so-called implicit, racially coded phrases, such as "welfare queens." We also believe the frame is substantially more implicit than the frames embedded in either pictures or in extensive narratives that have been studied by others, if only because the race in these other studies is obvious to the audience. Not only is our approach more realistic, given the climate in which more obvious forms of racialization are likely to backfire, but it is also

4. Respondents in the National Race and Crime Survey were asked to guess the percentage of all Americans living in the inner city who are African-American; the median guess was 60 percent, a much higher figure than the actual percentage of residents of central cities who are African-American, which was 20 percent in 1990, according to the U.S. Census.

more potentially frightening to the degree that our expectations are confirmed. For if this simple phrase is found to effectively frame responses to the crime problem, we will know how easily public opinion can be shaped.

Analysis

DATA AND MEASURES

The data for the analysis are drawn from the sample of (non-Hispanic) whites in the National Race and Crime Survey, a nationwide random digit dial (RDD) telephone survey administered by the Survey Research Center (SRC) at the University of Pittsburgh between October 19, 2000, and March 1, 2001. Interviews were completed with a total of 602 white respondents, for an overall response rate (response rate 3) of 48.64 percent (American Association for Public Opinion Research [AAPOR] 2004, p. 31).[5] To conserve space on the survey, the experiment described below was randomly administered to half of the white sample ($N = 290$). Further details on the sample are available from the authors on request.[6]

Dependent Measure. Responses to the crime policy question are assessed on a 4-point scale ranging from strongly preferring new prisons (point 1) to strongly preferring antipoverty programs (point 4) as the best way to deal with crime.

Racial Attitudes. We included two sets of racial attitudes in the survey: stereotypes of African-Americans (Black Stereotypes) and beliefs about racial discrimination in the criminal justice system (Racial Fairness). Whites who accept negative stereotypes of African-Americans–viewing them as lazy, violent, and dishonest—should be more likely to prefer the punitive policy option in the inner-city condition. Accordingly, our measure of Black Stereotypes is an additive index (ranging from 5 to 35) of the extent to which whites rate "most blacks" negatively on five, 7-point trait scales (see appendix, items 1–5).

Our index of Racial Fairness—the second racial attitudes variable—was created by adding responses to four items that assessed whether the police or the courts "in your community" treat blacks less fairly than they treat whites (see appendix items, 6 and 7a–c). The resulting scale ranges from 4 (very unfair) to 23 (very fair). Whites who continue to deny that blacks are treated unfairly by the police and the courts—in spite of an abundance of evidence to the contrary (Lauritsen and Sampson 1998; Walker, Spohn, and DeLone

5. The survey data are scheduled to be deposited at Inter-University Consortium for Political and Social Research (ICPSR).

6. The study also included a sample ($N = 579$) of African-American respondents who were not used in this analysis, in part because there was virtually no variation in responses to the dependent measure among African-Americans, who overwhelmingly rejected the prison option regardless of experimental condition.

2000)—are likely to favor punitive measures to control violent inner-city crime because, in their view, the higher crime rate of African-Americans is due more to the failings of blacks than to the justice system, which they view as imminently fair and color-blind.

Control Variables. We include a range of controls in the analysis to guard against the possibility that racial attitudes are tied to crime policy preferences due to their association with other, spurious factors.[7] Most important in this regard are two controls designed to serve as baselines against which the racial attitudes variables can be compared: White Stereotypes and General Fairness. White Stereotypes is assessed over the same five traits as Black Stereotypes, but in reference to "most whites." In contrast to the Racial Fairness scale, the two items assessing the General Fairness of the criminal justice system make no reference to race (appendix, items 8 and 9). In addition, because racial attitudes are associated with punitiveness, or a desire to punish those who break the rules (Hurwitz and Peffley 1992), we included such a measure (Punitiveness) as a control (see appendix, items 10 and 11), as well as Fear of Crime (appendix, items 12 and 13), which is often associated with a desire to punish criminals (Ferraro 1995).

Other political orientations (Ideology, Party Identification, and Equality) and social demographic factors (education, gender, age, income, and residing in the South) may also shape crime policy attitudes and thus are also included as controls (see the appendix for details about the measurement of the Equality and the demographic control variables).[8]

To test our hypotheses, we regressed Anticrime Policy Preferences on the predictor variables, a dummy variable representing the question frame (coded 1 for inner-city reference and 0 otherwise), and interactions for each of the predictors and question frame, using ordered probit analysis.[9] The regression results are displayed in table 1. The first two columns of coefficients provide the estimates for the model without the reference to the inner cities (when question frame = 0), and the second two columns give the computed coefficients for the model with the inner-city reference. As the coefficients in the left-hand side of the table make clear, racial attitudes (black stereotypes and racial fairness) have no discernible impact on crime policy preferences when no reference is made to inner-city criminals. The effects of black stereotypes and racial fairness are tiny and are far from being statistically significant. Rather, whites' preferences for fighting crime in the baseline (no inner city)

7. It is ordinarily considered unessential with an experimental design to incorporate control variables. In this case, however, it is important to be able to demonstrate that racial attitudes are more responsive to racial coding than other variables and that the effects of racial attitudes are not due to their association with other variables that might be affected by coded language.

8. Ideology and Partisanship are both measured in the standard way, ranging from "strong conservative" (Republican) at point 1 to "strong liberal" (Democrat) at point 7.

9. Rather than assume that the effects of the nonracial variables must be constant across question frame, we allowed the effects of all predictors—racial and otherwise—to vary across the experimental conditions.

Table 1. Predicting Whites' Crime Policy Preferences across Question Frames

	No Inner-City Reference		Inner-City Reference (Computed)	
	Coefficient	Standard Error	Coefficient	Standard Error
Racial Attitudes				
Black Stereotype	.009	.028	−.081**[a]	.030
Racial Fairness	−.020	.031	−.144**[a]	.032
Control Variables				
General Fairness	.045	.077	.129	.093
White Stereotype	−.027	.028	.041	.035
Equality	.081	.086	−.136	.097
Punitiveness	−.252**	.103	−.318**	.109
Fear of Crime	.186	.124	.007	.132
Ideology	.070	.082	−.134	.091
Party Identification	−.203**	.076	.005[a]	.078
Education	.035	.089	.159	.114
Gender	.780**	.247	.250	.289
Age	−.017	.008	−.007	.009
Income	−.096	.092	−.269**	.100
South	−.066	.254	−.381	.304
Question Frame (Inner city = 1)	3.437	1.984		
Interactions with Frame				
Black Stereotype × Frame	−.090*	.041		
Racial Fairness × Frame	−.124**	.045		
Intercept 1	−.451	1.444		
Intercept 2	−.057	1.444		
Intercept 3	.609	1.443		
N	250			
Pseudo R^2	.196			

NOTE.—Entries are ordered probit coefficients, with standard errors. Higher values on the above variables indicate the following: favor spending on antipoverty programs versus prisons, more negative stereotypes, justice system is racially fair, justice system is generally fair, egalitarian, fear of crime, punitive, conservative, Republican, female, older, higher income, residing in the South, and the question frame contained the inner-city reference.

[a] Coefficients are statistically different across experimental conditions at the .05 level.

* $p < .05$.

** $p < .01$.

condition are a function of more traditional determinants of crime policy attitudes—partisanship, gender, and punitiveness. As one might expect, Republicans, males, and those with more punitive orientations are significantly more likely to favor fighting crime by building more prisons than by spending money on antipoverty programs.

When the frame of the question is altered by inserting the phrase "inner city," however, policy preferences are driven by very different evaluative criteria. Consistent with our expectation (H_1), when whites are asked about locking up violent *inner-city* criminals, their policy preferences are based much more strongly on their racial attitudes. Not only are the coefficients for black stereotypes and racial fairness significantly larger in the racially coded (than the baseline) condition, but their impact in the coded condition is significant beyond the .01 level (coefficients in the right-hand side of the table). And consistent with the second hypothesis, when crime is framed as a racially coded issue, whites who endorse negative racial stereotypes and who view the justice system as racially fair are much more likely to favor punitive (building prisons) versus preventive (antipoverty) policies as the solution to fight crime.

To gain a better understanding of the impact of the two racial attitudes, we display in figure 1 the predicted probabilities of whites who favor new prisons for the two experimental conditions across both the black stereotype variable (the top portion of the figure) and the racial fairness variable (the bottom portion of the figure).[10] As the figures make plain, the impact of the racial variables is essentially nonexistent when no reference is made to inner cities; the plots are basically flat in the control condition. In the racially coded condition, however, the probability of favoring punitive measures to fight crime increases dramatically among whites who are more accepting of racial stereotypes or who believe the justice system is racially fair. Moreover, while the percentage of whites favoring prisons is uniformly low in the control condition (ranging from 14 percent to 25 percent), in the inner-city condition the punitive option is preferred by a fairly large percentage of whites with extremely negatives views of African-Americans (50 percent and 80 percent, respectively, of whites at the extremes of the racial fairness and black stereotype scales).

The figure also suggests that racial liberals as well as racial conservatives react differently to the two policy frames.[11] While, as expected, racial conservatives (those on the right-hand side of the two racial attitude scales) are *more* likely to endorse punitive crime policies in the treatment versus control conditions,

10. Predicted probabilities were generated for favoring prisons (either strongly or not so strongly) based on the ordered probit results in the first two columns of coefficients in table 1 using the Spost program developed by Long and Freese for Stata 7.0 (2001). Probabilities are computed by varying the racial attitude variable, holding other predictors in table 1 constant at their sample means and setting gender to male.

11. We use these terms only as a shorthand suggesting that racial conservatives hold more negative racial stereotypes and are more likely to deny that the justice system discriminates against African-Americans.

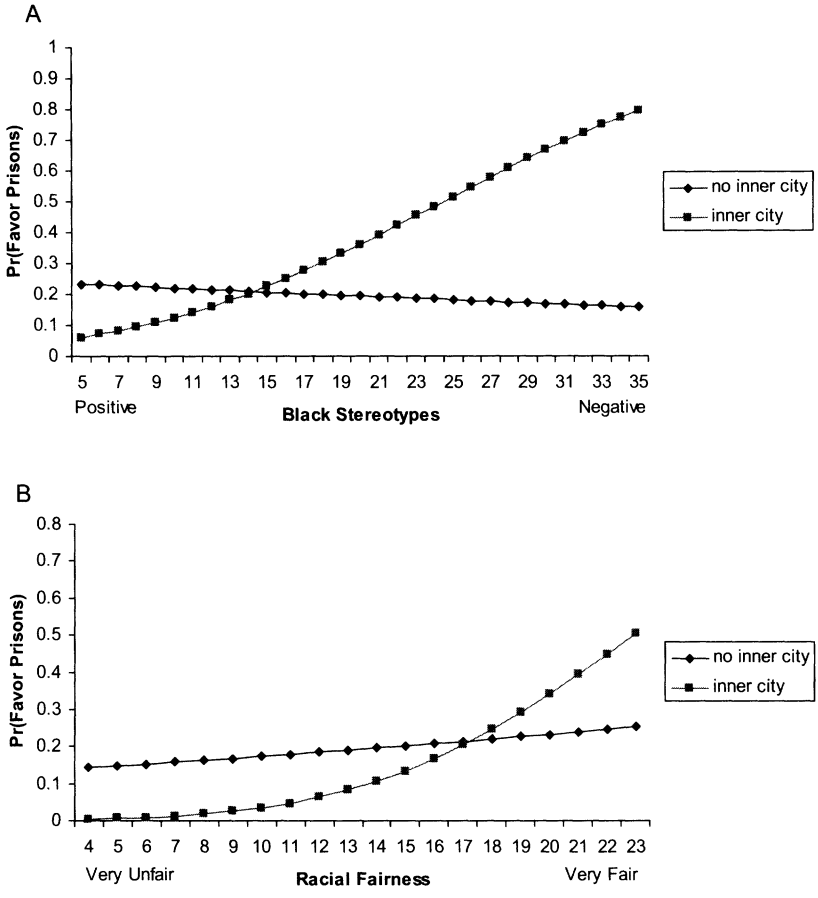


Figure 1. Predicted probabilities across racial attitudes (based on table 1). Probability in favor of prisons across black stereotypes by inner-city reference, whites (A). Probability in favor of prisons across racial fairness by inner-city reference, whites (B).

we also find that racial liberals (those at the left end of the racial attitudes scales) are *less* likely to endorse punitive crime policies in the inner-city condition. Although the differences across conditions appear larger, overall, for racial conservatives (especially for the black stereotype scale), racial liberals are also affected by the frame. Their support for the punitive option in the inner-city frame is close to 0 percent. Thus, the use of the racially coded phrase, "inner city," appears to be as much of a cue to racial liberals to reject punitive solutions as it is to conservatives to endorse them.

Conclusions

This analysis has been designed to address, in the most specific fashion possible, racialized code words in modern political discourse. The findings, we believe, are highly informative in a number of different ways. Most generally, they are fully consistent with, and add to, the literature on issue framing—particularly the literature focusing on the group-centric nature of a frame (Nelson and Kinder 1996). When messages are framed in such a way to reinforce the relationship between a particular policy and a particular group, it becomes far more likely that individuals will evaluate the policy on the basis of their evaluations of the group.

From our results, it apparently does not take much to reinforce this relationship—at least in the racial domain. Because of the very simple experimental design employed, in which differences between the inner-city and the baseline group can only be attributed to the introduction of the code word, it seems clear that respondents who think about criminals from the inner city have been encouraged to evaluate anticrime policies on the basis of the group that they associate with inner cities—African-Americans. As such, these respondents are far more likely (relative to those in the baseline group) to link their policy preferences to their feelings toward blacks.

This result is chilling in its implications, for it demonstrates how easily opinion can be manipulated and, more specifically, how easy it can be to "play the race card." As we argued at the outset, one of the legacies of the Willie Horton ad, and its ensuing discussion, may have been to make it more difficult to manipulate opinion with blatantly racialized messages. Most likely, such a message would, today, be either unproductive or counterproductive because, as a number of studies suggest, implicit appeals are much more effective. We have carried implicitness to its extreme: we have used no visuals, no mention of race, and no mention of characteristics often associated with race (like "welfare queens"). Still, we have found a way to encourage respondents to evaluate government policies on the basis of racial beliefs instead of, say, partisanship. While our research has focused on policy preferences, we see no reason why we would not obtain the same results if we focused on candidate preferences—a topic for future analysis.

Clearly, racially coded language can affect citizens' political judgments in insidious ways. At the same time, however, our results also suggest that people are not likely to be helpless victims of elites' attempts to manipulate public opinion through the use of racial code words. Although we found that a reference to inner-city criminals pushed racial conservatives to more punitive policy preferences, racial liberals moved in the opposite direction, favoring more preventive (in this case, antipoverty) policies, when asked about criminals in the inner city. Racial liberals thus appear to resist even the most subtle racial appeals. Our findings on this score are consistent with studies that show that framing does not affect political judgments in a mindless way but works through a psychological process in which individuals consciously and deliberately think about the relative importance of different considerations suggested by

a frame (Nelson and Oxley 1999). Thus, while our findings provide ample evidence of the power of racial coding to affect citizens' judgments, they are also consistent with recent studies that demonstrate that there are limits to the ability of elites to use frames to manipulate mass opinion (Druckman 2001).

One reason for confidence in our findings is that the wording of the crime policy question in the survey experiment was designed to provide a conservative test of the power of the inner-city phrase to lift white support for the building of new prisons. Had we simply offered respondents a one-sided question with no alternative to the punitive option, the racial coding frame would undoubtedly have increased support for new prisons beyond what we observed in our experiment. However, by employing a two-sided question format where respondents were offered a choice between punitive and preventive (antipoverty) policy options, resulting responses are not only more valid but are presumably less influenced by framing effects than if respondents had been asked to agree (or disagree) with, say, a single Likert statement. Still, even under these circumstances—circumstances designed to discourage susceptibility to framing—we find it relatively easy to push people to base their policy preferences on racial beliefs. And we find it relatively easy to push racial conservatives to become proponents of more punitive strategies.

Our findings also have important implications for studies of media framing of news stories on violent crime that take place in the inner city. As numerous content analysis studies have shown, journalists are not only more likely to disproportionately portray violent crime as perpetrated by black males, but such stories are also much more likely to be set in the inner city, in part because of the closer proximity of crime scenes to urban news organizations (Entman and Rojecki 2001; Gilens 1999). Our results suggest that even if blacks are not portrayed in news stories, the inner-city setting of the story is likely to provide an implicit racial frame for the story, thus reinforcing the connection between race and crime in the minds of many whites (Gilliam and Iyengar 2000; Peffley, Shields, and Williams 1996).

Appendix

SURVEY ITEMS

Racial Stereotypes

On a scale from 1 to 7, where 1 means that it is a very poor description and 7 means that it is a very accurate description, how well do you think [. . .] describes most whites/most blacks?

1. lazy
2. prone to violence
3. prefer to live on welfare
4. hostile
5. dishonest

Racial Fairness (all items are reverse coded)

6. Do you feel that African-Americans in your community are treated less fairly than whites in dealing with the police, such as traffic incidents? (no = 1, yes = 2)
7. Now I'm going to read you several statements that some people make about problems with the justice system in their community. As I read each one, please rate how serious it is in your community on a 7-point scale, where 1 means it is not a problem and 7 means it is a serious problem.

    a. Courts that give harsher sentences to African-Americans than to whites.
    b. Police who stop and question blacks far more often than they stop whites.
    c. Police who care more about crimes against white people than crimes against minorities.

General Fairness

8. The justice system in this country treats people fairly and equally.
9. The courts in your area can usually be trusted to give everyone a fair trial.

Punitiveness

10. One good way to teach certain people right from wrong is to give them a good stiff punishment when they get out of line.
11. Parents need to stop using physical punishment as a way of getting their children to behave properly.

Fear of Crime

12. First, over the last five years or so, would you say that violent crime in our nation has increased, decreased, or stayed about the same?
13. Of all the problems facing the country today, such as education, taxes, and the environment, how would you rate the importance of the crime problem? Would you say it's the most important problem, no more important than other problems, or less important than other problems facing the nation today?

Equality

14. One of the big problems in this country is that we don't give everyone an equal chance. (reverse coded)
15. We shouldn't worry so much about how equal people are in this country.

Demographic Variables

Education: Respondents were assigned into the following categories based on their highest grade or level of education completed: (1) 8th grade or less; (2) 9–11th grade; (3) high school graduate/GED; (4) some college or post secondary school; (5) bachelor's degree; (6) some graduate studies; (7) master's degree; (8) doctoral degree.

South: A dummy variable coded 1 if respondents reside in one of the states of the former Confederacy, and 0 otherwise.

General Fairness, Punitiveness, and Equality: These items were measured using Likert scales ("strongly agree," "somewhat agree," "somewhat disagree," and "strongly disagree").

References

American Association for Public Opinion Research (AAPOR). 2005. *Standard Definitions: Final Disposition of Case Codes and Outcome Rates for Surveys.* 3d ed. Lenexa, KS: AAPOR. Available online at http://www.aapor.org (accessed January 30, 2005).

Druckman, James N. 2001. "On the Limits of Framing Effects: Who Can Frame?" *Journal of Politics* 63 (November): 1041–66.

Entman, Robert M., and Andrew Rojecki. 2001. *The Black Image in the White Mind: Media and Race in America.* Chicago: University of Chicago Press.

Ferraro, Kenneth. 1995. *Fear of Crime: Interpreting Victim Risk.* Albany, NY: State University of New York Press.

Gilens, Martin. 1996. " 'Race Coding' and White Opposition to Welfare." *American Political Science Review* 90 (September): 593–604.

———. 1999. *Why Americans Hate Welfare.* Chicago: University of Chicago Press.

Gilliam, Franklin D., Jr., and Shanto Iyengar. 2000. "Prime Suspects: The Influence of Local Television News on the Viewing Public." *American Journal of Political Science* 44 (July): 560–73.

Hurwitz, Jon, and Mark Peffley. 1992. "Traditional versus Social Values as Antecedents of Racial Stereotyping and Policy Conservatism." *Political Behavior* 14(4): 395–421.

———. 1997. "Public Perceptions of Race and Crime: The Role of Racial Stereotypes." *American Journal of Political Science* 41 (April): 375–401.

Iyengar, Shanto, and Donald R. Kinder. 1987. *News That Matters: Television and American Public Opinion.* Chicago: University of Chicago Press.

Jamieson, Kathleen Hall. 1992. *Dirty Politics: Deception, Distraction, and Democracy.* New York: Oxford University Press.

Kurtz, Howard. 1994. "Ads Use Crimes' Pain for Candidates' Gain: Focus on Victims Can Ricochet." *Washington Post,* November 2, p. A1.

Lauritsen, J. L., and R. J. Sampson. 1998. "Minorities, Crime, and Criminal Justice." In *The Handbook of Crime and Punishment,* ed. M. H. Tonry et al., pp. 58–84. New York: Oxford University Press.

Long, Scott, and Jeremy Freese. 2001. *Regression Models for Categorical Dependent Variables Using STATA.* College Station, TX: STATA Press.

Mendelberg, Tali. 2001. *The Race Card: Campaign Strategy, Implicit Messages, and the Norm of Equality.* Princeton, NJ: Princeton University Press.

Nelson, Thomas, and Donald R. Kinder. 1996. "Issue Frames and Group-Centrism in American Public Opinion." *Journal of Politics* 58 (November): 1055–78.

Nelson, Thomas E., Rosalee A. Clawson, and Zoe M. Oxley. 1997. "Media Framing of a Civil Liberties Conflict and Its Effect on Tolerance." *American Political Science Review* 91 (September): 567–83.

Nelson, Thomas E., and Zoe M. Oxley. 1999. "Issue Framing Effects on Belief Importance and Opinion." *Journal of Politics* 61 (November): 1040–67.

Peffley, Mark, and Jon Hurwitz. 2002. "The Racial Component of 'Race-Neutral' Crime Policy Attitudes." *Political Psychology* 23 (March): 59–74.

Peffley, Mark, Jon Hurwitz, and Paul Sniderman. 1997. "Racial Stereotypes and Whites' Political Views of Blacks in the Context of Welfare and Crime." *American Journal of Political Science* 41:30–60.

Peffley, Mark, Todd Shields, and Bruce Williams. 1996. "The Intersection of Race and Crime in Television News Stories: An Experimental Study." *Political Communication* 13:309–27.

Valentino, Nicholas A., Vincent L. Hutchings, and Ismail K. White. 2002. "Cues That Matter: How Political Ads Prime Racial Attitudes during Campaigns." *American Political Science Review* 96 (March): 75–90.

Walker, Samuel, Cassia Spohn, and Miriam DeLone. 2000. *The Color of Justice: Race, Ethnicity, and Crime in America.* 2d ed. Belmont, CA: Wadsworth.

ANNALS, *AAPSS*, **546**, July 1996

Framing Responsibility for Political Issues

By SHANTO IYENGAR

ABSTRACT: This article examines the influence of television news on viewers' attributions of responsibility for political issues. Television's systematic reliance on episodic as opposed to thematic depictions of political life elicits individualistic attributions of responsibility for national problems such as poverty and terrorism. These attributions emphasize the actions of private rather than governmental actors. By obscuring the connections between political problems and the actions or inactions of political leaders, television news trivializes political discourse and weakens the accountability of elected officials.

Shanto Iyengar is professor of political science and communication studies at the University of California, Los Angeles. His principal publications include News That Matters *(coauthored with D. R. Kinder);* Is Anyone Responsible? How Television Frames Political Issues; Explorations in Political Psychology *(coedited with William J. McGuire), and the forthcoming* Going Negative: How Political Advertisements Shrink and Polarize the Electorate *(coauthored with Stephen Ansolabehere).*

THE concept of responsibility is an essential building block of all social knowledge. From the demeanor of one's next-door neighbors to the behavior of elected officials in the nation's capital, people spontaneously attribute responsibility for the behaviors they observe. Attributions of responsibility are known to exert powerful influence over a broad spectrum of interpersonal and social attitudes.[1]

The two principal types of attributions correspond to causal and treatment responsibility.[2] Causal responsibility concerns the origin of a problem, while treatment responsibility focuses on who or what has the ability to alleviate the problem. Both types of attributions are especially relevant for understanding political life. Why political problems occur and recur and how they might be appropriately remedied are perennial themes in political campaigns. Do the poor prefer to remain wards of the state rather than to work for a living, or are they victims of circumstances and forces beyond their control? What course of action is likely to reduce poverty, and who are the persons or institutions with the ability to carry it out?

The importance of people's causal and treatment attributions for political issues has not been lost on those who seek public office. Since voters tend to punish or reward politicians depending upon the state of national—especially economic—conditions, incumbent officials from the president on down are quick to claim responsibility for outcomes deemed favorable and disclaim responsibility for events or decisions with negative implications.[3] The increasingly partisan and vitriolic debates over "who really did it" have, by some accounts, contributed to considerable public disillusionment with political leaders.[4]

How do people decide questions of responsibility? The dominant paradigm treats attributions as residues of political socialization and accul-

1. The psychological evidence is reviewed in David J. Schneider, Albert H. Hastorf, and Phoebe C. Ellsworth, *Person Perception* (Reading, MA: Addison-Wesley, 1979); Phillip Brickman et al., "Models of Helping and Coping," *American Psychologist*, vol. 37 (1982); James R. Bettman and Barton A. Weitz, "Attributions in the Board Room: Causal Reasoning in Corporate Annual Reports," *Administrative Science Quarterly*, vol. 28 (1983); James Lemkau, F. B. Bryant, and Phillip Brickman, "Client Commitment in the Helping Relationship," in *Basic Processes in Helping Relationships*, ed. T. A. Mills (New York: Aldine, 1982); Valerie S. Folkes, "Consumer Reactions to Product Failure: An Attributional Analysis," *Journal of Consumer Research*, vol. 10 (1984).

2. Alan I. Abramowitz, David Lanoue, and Subha Ramesh, "Economic Conditions, Causal Attributions, and Political Evaluations in the 1984 Presidential Election," *Journal of Politics*, vol. 50 (1988); Shanto Iyengar, *Is Anyone Responsible? How Television Frames Political Issues* (Chicago: University of Chicago Press, 1991).

3. Morris P. Fiorina, *Retrospective Voting in American National Elections* (New Haven, CT: Yale University Press, 1981); Douglas A. Hibbs, *The American Political Economy* (Cambridge, MA: Harvard University Press, 1987); Stephen Ansolabehere, Shanto Iyengar, and Adam Simon, "Good News, Bad News, and Economic Voting" (Paper delivered at the annual meeting of the American Political Science Association, San Francisco, 1990); Donald R. Kinder and Roderick Kiewiet, "Economic Discontent and Political Behavior," *American Journal of Political Science*, vol. 23 (1979).

4. Stephen Ansolabehere and Shanto Iyengar, *Going Negative: How Political Advertisements Shrink and Polarize the Electorate* (New York: Free Press, 1995).

turation. In this view, an individual's political ideology or worldview provides the dominant influence over attributions of responsibility. Newt Gingrich holds the poor responsible for poverty; Bill Clinton attributes responsibility to societal forces and institutions. In short, how individuals assign responsibility is considered part and parcel of long-standing political predispositions.

While the influence of culture or ideology on attribution of responsibility cannot be denied, there is considerable evidence that short-term factors are just as important. Beliefs about who or what is responsible are likely to shift depending upon the information environment in which political issues and events are presented. Today, the most important of these contextual influences is television news.

FRAMING EFFECTS OF NEWS COVERAGE

The concept of framing refers to the effects of presentation on judgment and choice. In the psychological literature, it is well known that individuals' choices vary dramatically depending upon whether the options are presented as potential gains or losses. When faced with prospects that are presented as relative gains (such as winning $1), experimental participants exhibit risk aversion—they prefer a sure gain to a gamble. When faced with a prospective loss, however, they become risk seeking and prefer to gamble than to accept a certain loss.[5] Analogous framing effects have been obtained by public opinion researchers who elicit diverging responses by varying the form and wording of survey questions. For example, the stimulus "people on welfare" typically elicits more disapproving and less charitable responses than the stimulus "poor people."[6]

Given the widespread presence of framing effects associated with wording shifts in the presentation of choice problems or opinion questions, similar effects might be expected with media news presentations. Most people encounter the world of public affairs through the language of television, and television news coverage of political issues embodies two distinct frames or modes of presentation: the episodic news frame and the thematic news frame. The research that is summarized here was designed to investigate the effects of these alternative frames on viewers'

5. Daniel Kahneman and Amos Tversky, "Choices, Values, and Frames," *American Psychologist*, vol. 39 (1984); Amos Tversky and Daniel Kahneman, "Rational Choice and the Framing of Decisions," in *Rational Choice: The Contrast Between Economics and Psychology*, ed. Hillel Einhorn and Robin Hogarth (Chicago: University of Chicago Press, 1987); George A. Quattrone and Amos Tversky, "Contrasting Rational and Psychological Analyses of Political Choice," *American Political Science Review*, vol. 82 (1988); Richard Thaler, "The Psychology and Economics Conference Handbook," in *Rational Choice: The Contrast Between Economics and Psychology*, ed. Einhorn and Hogarth.

6. Tom Smith, "That Which We Call Welfare by Any Other Name Would Smell Sweeter: An Analysis of the Impact of Question Wording on Response Patterns," *Public Opinion Quarterly*, vol. 51 (1987); Howard Schuman and Stanley Presser, *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context* (New York: Academic Press, 1982).

attributions of responsibility for political issues.

The episodic news frame depicts issues in terms of specific instances—for example, a terrorist bombing, a homeless person, or a case of illegal drug usage. Episodic reports are essentially illustrations of issues. The thematic frame, by contrast, depicts political issues more broadly and abstractly by placing them in some appropriate context—historical, geographical, or otherwise. A thematic report on poverty might present information about recent trends in the rate of poverty and the areas with the greatest concentration of poor people. In appearance, the thematic frame takes the form of a backgrounder report featuring a series of people talking.

In practice, of course, few news reports are purely episodic or thematic. A close-up portrait of an unemployed worker will invariably make reference to the national rate of unemployment. Conversely, a news story about congressional efforts to reform welfare programs might include an interview with a welfare recipient. Content analyses of television news coverage, however, suggests that in most cases one frame or the other predominates.[7]

The nature of television news and the increasingly competitive nature of the news business have combined to create a premium for episodic coverage of political issues. Episodic reports tend to provide good pictures; they do not require reporters with subject-matter expertise; and, being devoid of interpretive analysis, they are less likely to be labeled as biased by media critics. A number of content-analytic studies have documented the pervasiveness of episodic framing in broadcast news.[8]

With the support of the National Science Foundation, I was able to carry out a series of experiments designed to investigate how television's reliance on episodic framing affects the viewing public's understanding of political issues. The results showed that episodic framing breeds individualistic as opposed to societal attributions of responsibility; national issues are traced to private actions and motives rather than deep-seated socioeconomic or political conditions. Given the public's sustained exposure to episodic framing, these results suggest that the ultimate effect of television news is to protect elected officials from policy failures or controversies and thus strengthen their legitimacy.

RESEARCH METHODS

The experiments were aimed at two sets of national issues: issues concerning public order and issues concerning social or economic welfare. The former category consisted of crime and terrorism; the latter included unemployment, poverty, and racial inequality. These five issues have been at the forefront of political life in recent years and show no signs of abatement.

In each experiment, media framing of the target issue was manipu-

7. Iyengar, *Is Anyone Responsible?* chap. 3.

8. Ibid.; David L. Altheide, "Format and Symbol in Television Coverage of Terrorism in the United States and Great Britain," *International Studies Quarterly*, vol. 31 (1987); William A. Gamson, "News as Framing," *American Behavioral Scientist*, vol. 33 (1989).

lated so that some participants were exposed to episodic news while others watched thematic coverage. To use the example of poverty, one set of viewers watched a news report that described the financial woes of an unemployed autoworker in Ohio; others watched a report that juxtaposed the national unemployment rate with the size of the federal budget deficit. Since participants were exposed to experimental conditions at random and since the conditions differed only with respect to episodic or thematic coverage of the target issue, differences in viewers' responses between conditions can be treated as evidence of framing effects.[9]

Participants in the experiments were residents of the Three Village area of Suffolk County, New York, who were recruited through newspaper and other forms of advertising that offered payment of $10. When potential subjects responded to the advertisement, they were screened to exclude noncitizens, college students, and people under the age of 18. Pooled across all experiments, the participants were generally representative of the local area in terms of their social background and political orientation.[10]

All studies were conducted at the Media Research Laboratory, an office suite located on the campus of the State University of New York at Stony Brook. When participants arrived, they were informed that the study concerned selective perception of television news and that they were to watch a videotape of randomly selected news stories broadcast by the three major networks over the past six months.[11] After receiving their instructions, participants completed a short pretest questionnaire concerning their personal background, degree of attention to politics, and media habits and tastes.

In order to reduce the aura of the research laboratory, participants were encouraged to come with a friend, spouse, or coworker. Typically, participants watched the videotape with someone they knew. In addition, the viewing environment was furnished casually to resemble a typical living room or family room, and participants could sip coffee or browse through newspapers and magazines.

Following completion of the pretest, participants watched a twenty-

9. Of course, this is a probabilistic argument. It is possible that despite the use of randomization, experimental conditions will still differ in composition. It is standard practice in experimental research to verify that random assignment of participants achieves the desired effect. In none of the framing experiments described here did the conditions differ significantly on any relevant background characteristic.

10. The nine separate experiments were administered between June 1985 and September 1987. The number of individuals who participated in each experiment ranged from 40 to 244. While the participants constituted a reasonable approximation of Suffolk County, New York, they were, in comparison with the American electorate, more affluent, more Catholic and more Jewish, more educated, and more likely to have voted in the last election. For further details on the sample participants, see Iyengar, *Is Anyone Responsible?* chap. 3.

11. The instructions stated, "Today television news is the major source of information for Americans. This study is about how people evaluate, understand, and interpret television news stories. We are particularly interested in 'selective perception.' Do people's opinions about politics and government influence how they react to news? Do Republicans and Democrats really see the same news?"

minute videotape that included seven news stories. The fourth story on the tape was the experimental manipulation. This story framed the issue under investigation in either thematic or episodic terms. The treatment story was between two and three minutes in length.

On completion of the videotape, participants received the posttest questionnaire, which included a battery of open-ended questions probing attributions of causal and treatment responsibility for the target and other political issues. Specifically, individuals were asked, "In your opinion, what are the most important causes of _____?" They were then asked, "If you were asked to suggest ways to reduce ____, what would you suggest?" Each individual was allowed to answer freely, without prompting. Up to four separate responses were coded for each question.[12] In addition, participants completed numerous other questions concerning their attitudes toward groups, political leaders, and public policies. Once participants had completed the posttest, they were debriefed in full and paid.[13]

The experimental design described permits estimation of the effects of the treatment story—episodic or thematic—on viewers' attributions. However, because the experimental manipulations were based on actual news reports, they necessarily differ in many ways. These differences were minimized by editing (using studio-quality equipment) the episodic and thematic versions of the report so as to maximize the similarity of their content. For instance, they were edited to be of equal length, and when taken from the same network, the anchor's lead-in remarks were used in both cases.[14]

In short, the design used here is subject to the usual trade-off between realism and precision. The use of actual news stories created extraneous differences between the different news frames. By constraining the treatment stories to a high degree of visual and semantic similarity, however, the role of these differences was minimized.

RESULTS

Responses to the open-ended attribution questions were plentiful. On average, participants nominated two causes and two treatments for each of the five issues.[15] Causal responsibility for both crime and terrorism was assigned to the individual perpe-

12. Although these responses were unwieldy and coding intensive, they have the advantage of nonreactivity; unlike closed-ended questions, open-ended questions do not cue respondents to think of particular causes or treatments. Two coders read each questionnaire and classified each response. Despite the large number of raw categories, intercoder agreement indicated acceptable levels of reliability.

13. The experimental procedures adhered fully to the American Psychological Association's guidelines on the conduct of experimental research.

14. To guard against the possibility of confounded variables, the experimental tests of framing were always replicated with an entirely different set of news stories. To the degree that the observed framing effects appear across different sets of manipulations, the possibility that idiosyncratic differences between episodic and thematic reports are at work is minimized.

15. Crime elicited the largest number of causal and treatment attributions—2.7 and 2.1, respectively—presumably because of its relatively obtrusive nature.

trator (individualistic responsibility), to a variety of societal conditions including economic deprivation and political oppression (societal responsibility), or to the failure of the criminal justice process to mete out adequate retribution (punitive responsibility). Some 50 percent of the causal responses referred to societal responsibility; 36 percent suggested individualistic responsibility; and only 12 percent cited judicial leniency.

Treatment responsibility for crime and terrorism was reserved exclusively for society and government and took two forms—addressing the underlying political or economic grievances (societal treatment responsibility) or imposing more severe retaliation and punishment against terrorists and criminals (punitive treatment responsibility). Sixty percent of the responses called for more punitive policies, and 40 percent referred to economic and political change.

Turning to the issues of social welfare, causal responsibility was assigned either to individuals or to society at large. Attribution of individual responsibility referred to insufficient achievement motivation or inadequate training and education. Attributions of societal responsibility encompassed economic conditions, such as the changing nature of work; cultural and ideological values, such as racial prejudice; and unresponsive public policies, such as cuts in federal social welfare programs. For unemployment and racial inequality, societal causes outnumbered individual causes by a considerable margin, but in the case of poverty, the balance was relatively even.

Unlike crime and terrorism, the suggested cures for issues of social welfare were directed at actions by individuals and the collectivity (societal responsibility). Once again, poverty elicited an equal number of individualistic and societal prescriptions, whereas calls for societal action far exceeded references to individual action in the cases of racial inequality and unemployment. Relative to the unemployed and racial minorities, poor people are thought to have more control over their fates.

How did the relative prominence of individualistic and societal attributions change as a result of media framing? The most consistent framing effects were detected with news coverage of poverty and terrorism. Less clear-cut results, in which the effects of framing varied with the subject matter of the news, were observed with crime and racial inequality. Finally, attributions of responsibility for the issue of unemployment were unaffected by the experimental manipulations of framing.

Poverty and terrorism

Attributions of responsibility for poverty—both causal and treatment—became significantly more individualistic when news coverage was episodic. Conversely, thematic coverage elicited a greater preponderance of societal attributions. In effect, news that dwells on particular instances of poverty encourages viewers to blame the victim.

In addition to these significant framing effects, viewers' attributions of responsibility for poverty were also sensitive to the specific categories of

poor people depicted in the news. In one of the experiments, participants were exposed to one of four different episodic reports corresponding to different groups of welfare recipients: elderly widows, unemployed male workers, children, and single mothers. Within each of these categories, the treatment report varied the race of the poor person. Thus some participants watched a news report on hungry black children, while others watched the identical report, except that the children shown were white.

The results of this more elaborate study suggested that attributions of responsibility for poverty were not only contingent upon the news frame—in the sense that thematic coverage made viewers more societal in their reasoning, while episodic coverage had the opposite effect—but also sensitive to the particular class of poor people encountered in the news. The class that elicited the harshest responses—that is, the most numerous individualistic attributions of cause and treatment—was single mothers. In effect, compared with elderly widows, children, and unemployed males, single mothers are considered less deserving.

The race of the poor person depicted in the news also proved to be a meaningful cue. When viewers encountered black poor people, they often argued in favor of individual as opposed to societal treatment responsibility. Differences in causal attribution associated with the race of the poor person were in the same direction, but failed to achieve statistical significance. These racial differences were especially pronounced in the single-mother condition, where the black mother attracted more than double the volume of individualistic treatment responses of her white counterpart.

Turning to the issue of terrorism, two separate experiments yielded evidence of significant framing effects. When the news depicted terrorism in thematic terms—for instance, by noting recent changes in U.S. diplomatic policy toward countries suspected of fomenting international terrorism—viewers' causal and treatment attributions gravitated toward societal factors. When the news depicted a particular act of terrorism, however, attributions became significantly more individualistic and punitive in orientation. The effects were especially strong in the area of treatment responsibility; under the episodic framing of terrorism, the ratio of punitive to societal treatment attributions was 3:1; under conditions of thematic framing, the ratio was 1:1.

Crime and racial inequality

Unlike the experiments on poverty and terrorism, where the framing manipulation independently influenced viewers' attributions, the studies bearing on crime and racial inequality yielded more ambiguous and complex results. For both issues, the importance of the episodic versus thematic framing manipulation depended upon the particular subject-matter focus of the news. In the study on crime, participants were exposed to news coverage of violent crime (by far the most common theme in television news about crime), drug abuse, or the workings of the criminal jus-

tice process. The violent-crime category was subdivided into "black" and "white" categories depending upon the race of the individuals depicted as the perpetrators of criminal activity.[16]

The effects of framing in this experiment were weak and varied with the subject matter of crime news. The episodic conditions were characterized by significantly higher levels of individualistic causal attributions than the thematic conditions only when the report on crime concerned white violent crime or the criminal justice process.

While this experiment failed to generate a consistent pattern of framing effects, it did shed further light on the importance of racial cues conveyed in news coverage. Societal causes of crime were cited least frequently—29 percent of all causal attributions were societal—in the black crime condition. In this respect, the black crime condition deviated significantly from the three remaining subject-matter conditions. Not only did news about black crime divert attention from societal responsibility, but it also highlighted individual responsibility. More than 60 percent of all causal attributions were directed at individuals when the news reported on black violent crime, which was double the comparable percentage in the white-violent-crime condition.

The experiment featuring racial inequality as the target issue incorporated three subject-matter categories, which corresponded to the prevailing themes in network news coverage of this issue. Participants watched a news report on racial discrimination against blacks, on economic inequality between whites and blacks, or on affirmative action. The framing manipulation was confined to the affirmative action and economic inequality conditions.

The effects of framing were limited to news coverage of economic inequality. In the case of affirmative action, attributions of responsibility were unaffected by framing; societal attributions of responsibility were the dominant responses in both the episodic and the thematic conditions. When the focus of the news was black poverty, however, episodic framing significantly increased the prominence of individualistic attributions and reduced the frequency of societal attributions. By directing attention to a black poor person, the episodic report prompted a relative outpouring of individualistic accounts of racial inequality in American society.

Unemployment

Two separate studies were devoted to unemployment. In Study 1, two of the three conditions represented thematic framing. The news reports described trends in the recent unemployment rate and the difficulties facing the U.S. steel industry. The episodic framing condition described

16. The "black crime–episodic" condition described a violent gang-related confrontation in Los Angeles. The "black crime–thematic" condition described the extent of crime in inner-city neighborhoods in several major cities. The "white crime–episodic" condition focused on the well-known shooting incident in the New York City subway in which Bernhard Goetz, a white passenger, shot at four black teenagers who allegedly approached him in a menacing manner. The "white crime–thematic" condition detailed the growing importance of organized crime.

the economic difficulties facing an unemployed white male.

In all three conditions, attributions of societal responsibility overwhelmed all other responses. The level of societal responsibility was lowest in the episodic condition, but the magnitude of the differences was trivial.

The second effort to detect framing effects included one thematic condition—on regional differences in unemployment—and two episodic conditions on unemployed autoworkers, one of whom was black and the other, white. Once again, no significant framing effects emerged. Attributions of responsibility for unemployment were predominantly societal regardless of the news frame or the race of the unemployed worker.

Overall, with the exception of unemployment, the results suggest that media framing does shape attributions of responsibility for political issues.[17] Considering the relatively minute scope of the experimental manipulations (a single news report) and the fact that they addressed highly visible issues with which the participants had extensive familiarity, it is notable that any of the experimental effects proved significant. In general, the dominant episodic frame used by television news increased viewers' reliance on individualistic or nonsocietal constructions of political issues, in which the characteristics or motives of private citizens are the most relevant causes or cures. Prior efforts to explain the prominence of individualistic attributions in the public's reasoning about political issues have cited mainstream cultural values as the source. However, the well-documented tendency of Americans to consider poor people, minorities, or those who break the law as personally responsible for their actions may be due not only to cultural norms but also to a pattern of television news coverage in which the spotlight is directed at poor people, criminals, or terrorists.

17. The distinctiveness of unemployment may stem from the unusually high level of prominence accorded the issue during the period of this research. Thus people may have regarded unemployment as a particular instance of national economic problems.

Spillover effects

To this point, the tests of framing have been confined to particular issue areas. It is possible that framing effects may affect related issues. Episodic framing of poverty, for instance, may carry over to influence attributions of responsibility for unemployment or racial inequality. There were two tests of spillover effects. The first concerned the degree to which episodic or thematic framing of the target issue affects attributions of responsibility for a closely related issue. The second test addressed the degree to which episodic or thematic framing affects the consistency of viewers' attributions for related issues.

The first study investigated spillover between poverty and unemployment. If the two issues are psychologically connected, thematic framing of poverty should strengthen attribution of societal responsibility for unemployment, while episodic framing of poverty should strengthen attribution of individual responsibility for unemployment. The results demonstrated that, with one excep-

tion (thematic framing of poverty significantly increased the prominence of societal causal attribution for unemployment), the manner in which the news framed poverty had no relevance to individuals' understanding of unemployment.

As a second test of spillover, between-issue correlations between summary measures of causal and treatment responsibility were computed.[18]

These correlations were small (.22 for causal responsibility and .26 for treatment responsibility), suggesting that individuals did not attribute responsibility consistently. The level of between-issue consistency was uniformly low in both episodic and thematic conditions.

The second study of spillover focused on crime and terrorism. Of the five different categories of causal and treatment attributions (individual, societal, and punitive causal responsibility; societal and punitive treatment responsibility), only one showed traces of spillover; episodic framing of terrorism induced significantly fewer societal treatment attributions for crime. In all remaining cases, the framing effects were issue specific.

Unlike the earlier results on poverty and unemployment, this study showed that thematic framing significantly increased the consistency of the attributions across issues. When exposed to episodic framing, viewers tended to offer distinct attributions for crime and terrorism; when shown thematic coverage, their attributions became more convergent, suggesting that they linked the two issues. All told, the analyses of spillover showed that individuals tend to exhibit little consistency across issues when attributing responsibility to individual, societal, or other factors. People tend to consider issues discretely rather than developing an overarching schema for political responsibility. Attribution of responsibility is domain specific rather than general. These data thus confirm the classic findings of Lane and Converse that political belief systems are narrow rather than broad and that particular issue publics exhibit distinct opinion profiles.[19]

CONCLUSION

Two sets of implications can be drawn from the media-framing experiments. First, the evidence on framing suggests a circumstantially bounded process of political reasoning in which attributions of responsibility are buffeted about by the prevailing winds of news coverage. While core values such as individualism and the work ethic encourage citizens to hold individuals rather than society responsible for issues such as poverty or racial inequality, exposure to thematic framing of issues can and does override these dispositions. Therefore, the dominant dispositional model of public opinion and political behavior that grants monopoly status to stable personal

18. Each respondent's percentage of individualistic attributions was subtracted from the percentage of societal attributions to arrive at a net responsibility score.

19. Robert E. Lane, *Political Ideology: Why the Common Man Believes What He Does* (New York: Free Press, 1962); Philip E. Converse, "The Nature of Belief Systems in Mass Publics," in *Ideology and Discontent*, ed. David Apter (New York: Free Press, 1964).

influences—most notably, party identification—must be revised to allow for circumstantial influence.

In addition to framing, the news shapes attributions of responsibility in other ways. Racial cues in the news proved significant in shaping viewers' responses to crime and poverty. It is possible that by depicting African Americans in the role of poor people or perpetrators of criminal activity, television may contribute to a reaffirmation of old-fashioned racial prejudice. That is, the frequent association of African Americans with economic failure and violent crime encourages and justifies the expression of racist attitudes.[20]

The second set of implications concerns the dominance of the episodic frame in television news and the resulting distortions in the political process. By reducing complex issues to the level of anecdotal cases, episodic framing leads viewers to attributions that shield society and government from responsibility. Confronted with a parade of news stories describing particular instances of national issues, viewers come to focus on the particular individuals or groups depicted in the news rather than historical, social, political, or other such structural factors.

Americans are not, however, intrinsically averse to structural accounts of responsibility for political issues. When the news presents a general frame of reference for national problems, viewers' reasoning about causal and treatment responsibility shifts accordingly. For example, following exposure to increases in malnutrition nationwide, poverty becomes a matter of inadequate governmental action; confronted with news about deteriorating economic conditions in inner-city areas, individuals cite increased economic opportunity as the appropriate remedy for crime.

Because the public's reasoning about responsibility is susceptible to framing effects and because the episodic frame is so predominant, television news, in the final analysis, is a significant resource for political elites. Instead of serving as a restraining or monitoring force, television news enables incumbent officials to distance themselves from any rising tide of disenchantment over the state of public affairs.

20. For experimental evidence that demonstrates this connection, see Franklin Gilliam, Shanto Iyengar, and Adam Simon, "The Intersection of Race, Crime and Television News" (Paper delivered at the annual meeting of the Midwestern Political Science Association, Chicago, 1995).

Article

Journal of Language and Social Psychology
2021, Vol. 40(2) 277–296

DOI: 10.1177/0261927X20949594
journals.sagepub.com/home/jls
SAGE

The Deception Spiral: Corporate Obfuscation Leads to Perceptions of Immorality and Cheating Behavior

David M. Markowitz[1], Maryam Kouchaki[2], Jeffrey T. Hancock[3], and Francesca Gino[4]

Abstract

In four studies, we evaluated how corporate misconduct relates to language patterns, perceptions of immorality, and unethical behavior. First, we analyzed nearly 190 codes of conduct from S&P 500 manufacturing companies and observed that corporations with ethics infractions had more linguistically obfuscated codes than corporations without ethics infractions. Next, we tested perceptions of a company based on values statements modified by obfuscation (Study 2). Participants perceived low-obfuscation companies as more moral, warmer, and more trustworthy than high-obfuscation companies. Finally, behavioral experiments (Studies 3a and 3b) revealed that group members cheat more after reading a high-obfuscation values statement than a low-obfuscation values statement. The results provide evidence of a potentially troublesome cycle: corporate unethicality has linguistic traces, can affect how people appraise a company, and can change ethical behavior.

Keywords

obfuscation, corporate unethicality, deception, deception spiral

Moral decisions are a recurrent part of everyday life. People regularly face opportunities to cheat (Ariely, 2012) and decisions about whether to report the misconduct of others (Bird, 1996). The ubiquity of moral decision making suggests people should be

[1]University of Oregon, Eugene, OR, USA
[2]Kellogg School of Management, Northwestern University, Evanston, IL, USA
[3]Stanford University, Stanford, CA, USA
[4]Harvard Business School, Harvard University, Boston, MA, USA

Corresponding Author:
David M. Markowitz, School of Journalism and Communication, University of Oregon, Allen Hall, Eugene, OR 97403, USA.
Email: dmark@uoregon.edu

adept at recognizing and reporting immoral behavior. Decades of social psychological research, however, suggest people are poor deception detectors (Bond & DePaulo, 2006; Markowitz, 2020) and hesitant to blow the whistle (Miceli et al., 2009). When people judge hypothetical and unambiguous moral situations, however, ethical assessments become clearer.

Our investigation tests how people make moral judgments about corporations using language cues, an idea rooted in a tradition of research that uses words to infer social and psychological processes (Maass et al., 1989; Pennebaker, 2011). Language analysis is effective for evaluating psychological dynamics about people (e.g., social status) and networks or organizations (Kacewicz et al., 2014; Margolin & Markowitz, 2018). By looking at language data, we can learn about social and psychological processes that are representative of a group.

In this article, we evaluate the bidirectional effect of language cues and moral behavior in large groups, particularly how corporations' writing style is linked to deception and ethics infractions. We predict that deception changes how corporations communicate about themselves and that these changes influence perceptions of the group and the moral behaviors of group members. Indeed, the language used by corporations reflects the organization's culture and shapes employees' perceptions and behaviors (Weick, 1979). We identify one style of language, linguistic obfuscation, that may reflect and subsequently encourage unethical behavior. Our findings suggest that obfuscated language is a subtle cue that can undermine the ethicality of decision makers. We examine the possibility of a troublesome "deception spiral" in which both the language and its recipients perpetrate the deceptive behavior.

Language and Deception: The Linguistic Obfuscation Hypothesis

Corporations often use obfuscated, or difficult to understand and obscured language, in documents that involve deception or indicate poor financial performance. For example, companies with negative annual earnings produce more obfuscated reports to the Securities and Exchange Commission than do companies with positive earnings (Humpherys et al., 2011; Li, 2008). These data are consistent with the obfuscation hypothesis (Bloomfield, 2002; Courtis, 1998), which predicts that corporations hide problematic behavior or performance in complex and less understandable reports. Therefore, obfuscated language is a consequence of a corporation's intent to mislead an audience after unfavorable performance or behavior. This form of impression management is deliberate, where language patterns reflect a company's interest in opacity and manipulation of information. Obfuscation as a form of impression management is also theoretically consistent with research on strategic ambiguity, which suggests that people in organizations often "respond with communicative strategies which do not always minimize ambiguity, but may nonetheless be effective" (Eisenberg, 1984, p. 228). Therefore, obfuscation does not need to be overtly deceptive, but instead, a goal-oriented strategic communication method that allows a corporation to self-present in a way that honesty could not.

Obfuscation is therefore defined as a strategic form of communication used to build "cohesion and unity across diverse audience segments," but is achieved verbally with unclear and difficult-to-understand language (Clementson, 2018, p. 481). This concept is different from other forms of deception, such as equivocation, because of its effects on an audience. As Clementson (2018) suggests, people should respond favorably to equivocal language, though people should respond unfavorably to obfuscation because obfuscating leaders "are often not as clear in their communication as we would like them to be" (Dewan & Myatt, 2008, p. 353). Therefore, obfuscation is a particular form of deception that uses unclear, ambiguous, and often incomprehensible language for an audience to interpret. The effects of obfuscation, as we demonstrate, are far-reaching: We predict and find evidence that obfuscation affects how people perceive the morality, warmth, and trustworthiness of a company, and a company's muddled values statement can cause people to cheat for monetary gain.

Evidence for the obfuscation hypothesis is robust (Bloomfield, 2002; Burgoon et al., 2016; Courtis, 1998; Garrett et al., 2016; Humpherys et al., 2011; Li, 2008; Mann et al., 2014; Riley & Luippold, 2015) and has been extended to domains where people write about other forms of unethical behavior. For example, deceptive scientists—who tried to present their studies as genuine but had their papers retracted for fraud—used a more abstract writing style, more jargon, and less readable writing as compared with honest scientists (Markowitz & Hancock, 2016). The linguistic obfuscation hypothesis suggests that verbal content (e.g., as indicated by rates of jargon, positive emotion words, and causal terms), style (e.g., as indicated by abstraction), and structure (e.g., as indicated by readability) are related to deceptive intent. Given the strong empirical evidence supporting the obfuscation hypothesis (Bloomfield, 2002; Burgoon et al., 2016; Courtis, 1998; Garrett et al., 2016; Humpherys et al., 2011; Li, 2008; Mann et al., 2014; Riley & Luippold, 2015), which suggests the lack of clarity in writing often signals false speech (see also McCornack, 1992) and is indicated by specific language patterns, we predict that companies with ethics infractions will use more linguistic obfuscation in their values statements than those without ethics infractions.

An important, yet understudied test of the linguistic obfuscation hypothesis is how obfuscation affects perceptions of the ethicality of a target. Related fluency research suggests that high text complexity often leads to negative judgments and perceptions. Oppenheimer (2006) observed that participants who read a high-complexity admissions essay by a prospective student appraised the student more negatively than did those who read a moderate- or low-complexity essay. Other evidence suggests that low-complexity text is preferable to high-complexity text across domains, including corporate communications (Chou et al., 2017). In the current research, we evaluate if manipulating linguistic obfuscation in ethics documents affects how people judge a company's morality and ethicality. We predict that when participants are presented with a values statement—a section from a code of conduct that outlines the company's honor code, ideals, and mission—a high-obfuscation writing style will lead to more negative moral perceptions of the corporation (e.g., less morality, less warmth, less

trust) as compared with a low-obfuscation writing style. We further suspect that high-obfuscation writing in a values statement would affect perceptions related to dimensions stated in the message—that is, moral perceptions of the firm—and not necessarily other perceptions, such as competence. We do not offer a formal prediction for competence but rather consider this variable exploratory.

So far, we argued that linguistic obfuscation can affect perceptions of morality. Does linguistic obfuscation in a values statement also affect ethical behavior? Prior work suggests that members of unethical groups can be affected by corporate dishonesty. For example, in one study, employees of a large international bank demonstrated more cheating behavior on an experimental task when they were reminded of their professional identity (Cohn et al., 2014). The researchers hypothesized that the banking industry's culture of dishonesty prescribes bad behavior to its members, an argument echoed in other work as well (Gino et al., 2009).

We therefore expect that exposure to an obfuscated values statement will lead to higher rates of cheating as people will perceive the company as less moral. An alternative explanation to this account, however, is that people tend to make fewer ethical decisions after expending the high levels of cognitive effort, which is typically required by obfuscated text (Gino et al., 2011; Kouchaki & Smith, 2014). Therefore, we test whether resource depletion predicts the possible cheating effects of obfuscated language. In the final studies, we investigate if participants cheat more after reading their group's high-obfuscation values statement versus a low-obfuscation values statement.

Social scientific research evaluating obfuscation as a form of deception has largely investigated its verbal correlates and testing their applicability in new settings (e.g., financial fraud, deceptive conference calls). Our studies provide evidence that obfuscation affects how people think and feel about a company and cheating behavior. Therefore, we show that obfuscation is not just a linguistic artifact, but a psychological phenomenon that can affect perceivers and ethical decision making.

Study 1: Field Study

Method

Almost all companies, including those on the S&P 500 stock market index, write a code of conduct for their employees. While there is no universal template for such codes of conduct, most include letters from executive officers (e.g., the chairman, the chief financial officer), a values statement (e.g., an honor code, including the company's ideals and mission statement), guidelines, and frequently asked questions (e.g., what an employee should do when ethical issues arise). These documents offer important information to employees, providing a vision for the company and its future, while modeling the company's current positions on ethical issues within the organization. In this study, we use several automated text-analysis tools to evaluate dimensions of the linguistic obfuscation hypothesis. Specifically, we investigate word pattern differences in codes of conduct from companies with or without ethics infractions.

Data Collection and Database Information

We used code of conduct text to understand the psychology of the corporation. The codes were catalogued by prior work (Kouchaki et al., 2019), excluding six files that could not be parsed for an automated text analysis. This resulted in a final database of 188 codes from U.S. companies on the S&P manufacturing list (from 1990 to 2012). Prior authors compiled the companies' ethics infractions (e.g., environmental violations, fraud, anticompetitive activity) by searching media and Internet sources (Mishina et al., 2010). For each year, each company was dichotomously coded as unethical (e.g., had ethics infractions; indicator variable = 1) or ethical (e.g., did not have ethics infractions; indicator variable = 0). The corpus of 188 codes of conduct contained a total of 1,548,237 words (M = 8235.30 words, SD = 5154.92 words).

Automated Text-Analysis Approach

We analyzed the field study data with two automated text-analysis programs, Linguistic Inquiry and Word Count (LIWC; Pennebaker et al., 2015) and Coh-Metrix (McNamara et al., 2014). LIWC and Coh-Metrix quantify word patterns across a variety of social (e.g., words related to family, friends), psychological (e.g., emotion terms), part of speech (e.g., pronouns, articles), and discourse categories (e.g., readability), and have been applied to evaluations of linguistic obfuscation (Li, 2008; Markowitz & Hancock, 2016).

Each LIWC dimension is calculated as a percentage of each code's word count based on words incremented by the tool's internal dictionary (Tausczik & Pennebaker, 2010). For example, the sentence "We value honesty and integrity" contains five words across several categories, including but not limited to: first-person plural pronouns (e.g., *we*; 20% of the total word count), positive affect (e.g., *value, honesty*; 40% of the total word count), and conjunctions (e.g., *and*; 20% of the total word count). All language categories described below were drawn from the standard LIWC2015 or Coh-Metrix dictionaries unless otherwise specified.

The Linguistic Obfuscation Index

We analyzed the language patterns of each corporate statement along the linguistic obfuscation index (Markowitz & Hancock, 2016), a composite variable characterized by high rates of jargon, abstraction, and causal terms, but low scores of positive emotion terms and Flesch Reading Ease readability (Flesch, 1948). See Table 1 for the correlation matrix for obfuscation variables.

Jargon. Jargon was operationalized as the percentage of words not incremented by the LIWC dictionary, a proxy for the number of common words used in everyday English (Pennebaker et al., 2015; Tausczik & Pennebaker, 2010). This measure reflects complexity through the presence of specialized content (e.g., what is said) rather than structure (e.g., the length of a word or sentence). For example, the phrase "Honesty is

Table 1. Correlation Matrix for Obfuscation Variables ($N = 188$).

Variable	Jargon	Abstraction	Causal terms	Positive emotion	Readability
Jargon	—				
Abstraction	.439***	—			
Causal terms	−.136	.071	—		
Positive emotion	−.100	.325***	.192**	—	
Readability	−.109	.271***	.046	.299***	—

Note. Readability was assessed using Flesch Reading Ease scores.
**$p < .01$. ***$p < .001$.

important" contains three words in the LIWC dictionary, as compared with "Honesty is crucial," which contains one word outside of the dictionary, which would be considered jargon (*crucial*). The jargon formula for the linguistic obfuscation index is (100-Dictionary); this value was standardized for the obfuscation index. A high score represents a higher rate of jargon and specialized terms than a low score.

Abstraction. Linguistic abstraction is a composite variable derived from three function word categories, including articles (e.g., *a, the*), prepositions (e.g., *to, from*), and quantifiers (e.g., *more, less*). Function words, also known as style words, describe *how* a person is communicating instead of *what* a person is communicating about (e.g., the language content, typically expressed through nouns or verbs; Chung & Pennebaker, 2007).

Style words are important indicators of a range of social and psychological processes, from social status (Kacewicz et al., 2014; Markowitz, 2018) to persuasion (Larrimore et al., 2011). Specific function words indicate concrete writing patterns: Articles refer to concrete nouns (Tausczik & Pennebaker, 2010), prepositions are markers of complexity and critical thinking (Pennebaker et al., 2014), and quantifiers express degrees of difference between objects (Markowitz & Hancock, 2016). People who use high rates of articles, prepositions, and quantifiers often communicate with a concrete language style (e.g., "*The* fabric of *a* strong company . . .") relative to people who use low rates of these word types (e.g., "Strong companies . . .") and an abstract language style (Larrimore et al., 2011). To have a single measurement of linguistic abstraction, or the opposite of concreteness, we added the inverse of the standardized rates of articles, prepositions, and quantifiers. A high score on this index suggests a more abstract writing style than a low score (Larrimore et al., 2011; Margolin & Markowitz, 2018; Markowitz & Hancock, 2016). Intercorrelations between articles, prepositions, and quantifiers were all positive and significant ($rs > .172$, $p < .018$), suggesting that indexing these language variables was empirically valid.

While there are many operationalizations of abstraction or its opposite, concreteness (Pollock, 2018), we chose function words because they are often less susceptible to contextual constraints than other word types (e.g., verbs, nouns; Chung & Pennebaker, 2007). An abstract writing style therefore reflects the company's reduced focus on specific objects and details that are problematic for an underperforming company (e.g., earnings, employee matters).

Causal Terms and Positive Emotion Terms. Causal terms explain relationships between objects, such as *led, reacted*, and *made*. These words are often overused in obfuscated writing to explain why corporate performance may have failed to meet expectations (Li, 2008) and to overexplain the value of fraudulent science (e.g., relationships between variables; Markowitz & Hancock, 2016). Positive emotion terms (e.g., *happy, pleasure, strong*) are words that describe positive affect. Obfuscated text reporting on deceptive behaviors often contains fewer positive emotion terms than nonobfuscated text because authors understand their corporate performance is problematic or that data in a scientific paper is unverifiable (Li, 2008; Markowitz & Hancock, 2016). Both language dimensions were standardized for the creation of the obfuscation index.

Readability. Each text received a readability score using the Flesch Reading Ease metric (Flesch, 1948). A low score on this measure suggests that the text is more difficult to read (e.g., more words per sentence and more syllables per word) than a high score. Readability was calculated using Coh-Metrix and standardized for the index.

Results

We fit a logistic regression model to predict corporations with ethics infractions, using the overall obfuscation index as a single language predictor. We also include three primary control variables that may affect a company's inclination to act unethically. First, we include year as an indicator variable to control for systematic differences in unethical behavior over time. Second, we control for corporation size, operationalized as the natural logarithm of the number of annual employees, to account for potential inconsistencies among firm composition. Finally, we control for three types of slack resources, defined as available resources that a company can use to achieve their goals (George, 2005). We control for slack resources because companies with more resources may find it unnecessary to pursue unethical or illegal activities. Our slack controls included absorbed slack (e.g., the ratio of administrative expenses to sales), unabsorbed slack (e.g., the ratio of cash and marketable securities to liabilities), and potential slack (e.g., the ratio of debt to equity). Including these controls in the models ensured the best chance of predicting unethicality from linguistic obfuscation and ruling out alternative explanations from prior work.

Consistent with the linguistic obfuscation hypothesis, companies with infractions wrote their corporate statements in a more obfuscated style than did companies without infractions ($\beta = 0.08$, $SE = 0.02$, $z = 4.75$, $p < .001$). We also evaluated the dimensions of the index in separate logistic regression models with controls to assess each feature individually (see Table 2).

Most of the language dimensions, except for causal terms, operated in the predicted direction, and three of the five obfuscation features reached significance. Companies with ethics infractions wrote more abstractly, ($\beta = 0.07$, $SE = 0.03$, $z = 2.50$, $p = .012$), with more jargon, ($\beta = 0.07$, $SE = 0.01$, $z = 5.11$, $p < .001$), fewer positive emotion terms, ($\beta = -0.15$, $SE = 0.07$, $z = -1.98$, $p = .048$], and marginally less readable text ($\beta = -0.01$, $SE = 0.005$, $z = -1.83$, $p = .068$).

Table 2. Field Study Results Across the Obfuscation Index.

Language variable	β	*SE*	*z*	*p*	OR
Obfuscation Index	0.08	0.02	4.75	<.001	1.085
Jargon	0.07	0.01	5.11	<.001	1.075
Abstraction	0.07	0.03	2.50	.012	1.068
Causal terms	−0.03	0.12	−0.27	.788	0.968
Positive emotion terms	−0.15	0.07	−1.98	.048	0.864
Readability	−0.01	0.005	−1.83	.068	0.991

Note. Each logistic regression model included the *year*, corporation size, and slack resources control variables. OR = odds ratio and represent exponentiated logit betas.

The results from this field study suggest that unethical companies' code of conduct are written in a language style that is consistent with the linguistic obfuscation hypothesis. The obfuscation effect in this setting is a novel contribution, given that tests of the obfuscation hypothesis have not been applied to corporate records with documented ethics infractions. Instead, obfuscation has been evaluated in cases of financial fraud (Courtis, 2004; Humpherys et al., 2011; Li, 2008) and the writing style of deceptive scientists (Markowitz & Hancock, 2016). Therefore, this study suggests that corporate documents and guidelines can offer a valuable lens to evaluate unethical behavior. Not all obfuscation dimensions are unique contributors to identify companies with ethics infractions, however. Corporate unethicality is primarily associated with more linguistic abstraction (e.g., fewer articles, prepositions, quantifiers) and more jargon.

Next, we investigated if reading a high- or low-obfuscation values statement changes people's perceptions of a corporation's morality, warmth, competence, and trust in the company.

Study 2: Perceptions of Obfuscation

Person-perception is often evaluated along two dimensions, warmth and competence (Fiske et al., 2007). Warm and competent people are perceived more favorably (e.g., reliable, intelligent) than cold or incompetent people (e.g., dishonest, unintelligent). Goodwin et al. (2014) advanced this work by extending warmth-competence models to include morality. For example, they had coders assess obituaries (Study 7) and participants then provided an impression rating of the deceased individual based on the obituary text. The data suggested that morality was a unique construct relative to warmth and competence, was more prevalent than warmth in the obituary writing, and was a stronger indicator of impression ratings than warmth.

Together, person-perception can be captured by the prior dimensions that are independent from each other. We use these data as evidence to examine if word patterns

that are typically associated with unethicality (e.g., obfuscation) modify how people perceive a company based on its corporate writing.

Method

We created values statements (e.g., the section most similar to an honor code) from our code of conduct database to test how people rate the morality, warmth, competence, and trustworthiness of a company based on high- or low-obfuscation text. Participants judged values statements instead of complete codes of conduct because obfuscation is typically found throughout an entire document of text (e.g., science papers; Markowitz & Hancock, 2016) and such statements are likely familiar even to people who are not affiliated with a corporation (e.g., people likely have seen a university honor code or are familiar with the Ten Commandments; Mazar et al., 2008). We also wanted to keep the participants' effort burden reasonable. Values statements provide a succinct section of a code of conduct to evaluate how obfuscated language patterns affect perceptions of a company.

Using the Study 1 findings as inspiration to create our stimuli, we first successfully validated that values statements with high and low levels of obfuscation were indeed different on typical obfuscation perceptions (e.g., the clarity and complexity of the writing, how well the writing could be understood; see below for pilot study details). The high-obfuscation values statement contains high levels of abstraction (e.g., low rates of articles, prepositions, and quantifiers; Larrimore et al., 2011; Margolin & Markowitz, 2018; Markowitz & Hancock, 2016) and high rates of jargon (e.g., words outside of the LIWC dictionary; following a stimulus creation similar procedure to Oppenheimer, 2006). The high-obfuscation text example is below, with an overall jargon rate of 28.30% and no articles, prepositions, or quantifiers.

Be Good:
Obey laws and this code.

Be Honest:
Act honestly and scrupulously.

Be Equitable and Impartial:
Follow ordinances when helping government, customers, or suppliers.

Be Staunch:
Protect this Company's interests, assets, and data.

Be Accurate:
Keep complete and veracious business documentation.

Be Deferential:
Respect your peers and our social and physical environment.

A low-obfuscation values statement contains low levels of abstraction (e.g., high rates of articles, prepositions, and quantifiers) and low levels of jargon (e.g., more words captured by the LIWC dictionary). The low-obfuscation text example is also displayed below, with a jargon rate of 13.51%, and higher rates of articles (13.51%; *a, the*), prepositions (8.11%; *of, with, by, around*), and quantifiers (4.05%; *most, another*) than the high-obfuscation values statement.

Be Good:
Obey the law and the Company's Code of Conduct.

Be Honest:
Act with the most honesty and a high sense of integrity.

Be Fair and Impartial:
Play by the rules, whether working with government, a customer, or a supplier.

Be Loyal:
Protect the Company's interests, assets, and information.

Be Accurate:
Keep the most complete and accurate business records.

Be Respectful:
Respect one another and our social and physical environment around the world.

Participants and Power Analysis

We targeted a sample size of approximately 100 participants per cell, subject to the availability of participants, which would provide nearly 95% power to detect a medium effect size using a two-tailed test ($f = 0.25$, $\alpha = .05$).

Participants in Study 2 were recruited from Amazon Mechanical Turk ($M_{age} = 35.37$ years, $SD_{age} = 11.06$ years). Our study had 113 males and 113 females, and one participant who designated "other" as a gender. Gender was evenly represented across our two experimental conditions, $\chi^2(2) = 1.05$, $p = .592$. Experiments in this article received institutional review board approval from the second and fourth author's institutions.

Procedure

Participants entered the Qualtrics survey environment and were randomly assigned to view a high- or low-obfuscation values statement from the pilot study; results for this pilot are located in an endnote.[1] Specifically, participants read a document from an unnamed corporation that explained the values and responsibilities of the company. They were instructed to "read the following values statement from the company's

Code of Conduct." After reading the values statement, participants were provided with a list of attributes and were asked, "To what extent do you believe this company is . . ." followed by traits used to judge the company's morality, warmth, competence, and trust (see Measures). Finally, participants provided demographic information (e.g., age, gender) and exited the survey.

Measures

Morality, Warmth, and Competence Perceptions. Participants reported their perceptions of the company along morality, warmth, and competence dimensions. Goodwin et al. (2014) made a clear theoretical distinction between warmth perceptions (conceptualized as inclusive of morally relevant traits, such as sincerity, as well as less morally relevant traits, such as friendliness) and moral character traits that are low on warmth (e.g., principled, just, trustworthy). Using a selection of traits adapted from prior work (Goodwin et al., 2014), participants rated the company across psychological categories.

The categories and their traits included *High Morality, High Warmth* (forgiving, helpful, kind, empathetic, cooperative; Cronbach's $\alpha = .91$), *High Morality, Low Warmth* (fair, principled, responsible, ethical; Cronbach's $\alpha = .89$), *High Warmth, Low Morality* (warm, agreeable; Cronbach's $\alpha = 0.81$; $r = .695$, $p < .001$), and *Competence* (innovative, organized, logical, competent, efficient, effective; Cronbach's $\alpha = .89$). Participants rated each trait on scale of (1) *Not at all* to (7) *Very much*, and each category was calculated as an average of the traits. In rare cases, when a participant did not provide a rating for an individual trait, a measure of morality, warmth, competence, and trust was not calculated.

Trust. Participants rated their agreement with three statements on a scale of (1) *Strongly disagree* to (7) *Strongly agree* (Cronbach's $\alpha = .95$): (a) "This company can generally be trusted," (b) "I trust this company," and (c) "I can depend on this company to do the right thing." These three dimensions were averaged to create a single measure of perceived trust.

Together, we evaluated if participants who read a high-obfuscation values statement ($n = 115$) perceived differences in morality, warmth, competence, and trust as compared to participants who read a low-obfuscation values statement ($n = 112$).

Results and Discussion

Morality, Warmth, Competence, and Trust Perceptions. As Table 3 displays, participants rated the high-obfuscation values statements as less moral, less warm, and less trustworthy than the low-obfuscation values statements (ts > 2.44, ps $< .015$). Perceptions of competence were not statistically different across values statements.

Crucially, these effects were largely replicated in several other experiments reported in the Supplemental Material (available online). Our other experiments address important alternative explanations for the Study 2 effects, namely that the

Table 3. Morality, Warmth, Competence, and Trust Perceptions for Study 2.

	High obfuscation		Low obfuscation				
Category	*M*	*SD*	*M*	*SD*	*t*	*p*	*d*
High Morality, High Warmth	4.59	1.39	5.19	1.13	3.59	<.001	0.47
High Morality, Low Warmth	5.71	1.10	6.04	0.93	2.44	.015	0.32
High Warmth, Low Morality	4.63	1.50	5.30	1.22	3.72	<.001	0.49
Competence	5.47	1.06	5.61	0.97	1.02	.308	0.14
Trust	5.21	1.33	5.64	1.19	2.57	.011	0.34

text length, specific manipulations of abstraction, jargon, or their combination, and employee perspective-taking might affect the perceptions results. Such manipulations did not lead to systematic perceptions differences after participants read a high- or low-obfuscation values statement. Furthermore, using ecologically valid stimuli, we failed to obtain results. We provided these supplementary studies online in the spirit of transparent science. It is important to not store away these data since they helped to inform our understanding of the relationship between language, obfuscation, and perceptions of corporations. They also serve as positive replications for our Study 2 perceptions experiment.

Next, to explore the potential consequences of obfuscation and achieve our other empirical aim, two studies examined if reading a high-obfuscation values statement of one's in-group leads to a change in one's unethical behavior as compared with reading a low-obfuscation values statement. If we observe that reading a high-obfuscation values statement leads to more unethicality (e.g., cheating) from in-group members than reading a low-obfuscation values statement, there is sufficient evidence of a problematic "deception spiral."

Study 3a: Obfuscation Leads to Cheating With Anagrams

Method

Participants. Subjects were recruited from Amazon Mechanical Turk (M_{age} = 35.29 years, SD_{age} = 10.90 years) and paid for their time in a study with an opportunity to earn additional money based on their performance on a task (up to $2).

Procedure. This study began by telling participants to assume the role of a member of a research lab called the Behavioral Insights Group. Participants were told that they would read the group's values statement from its code of conduct (using the high- and low-obfuscation stimuli from the experiment using long obfuscation texts; see Supplemental Study 1 available online) and then complete a performance measure that was self-scored. The performance measure was modeled after a standard anagram or word unscrambling task with eight total trials. To ensure that participants understood the procedure, an example of the unscrambling task was provided (e.g., the letters T A C can be rearranged to spell CAT or ACT).

To incentivize participants to take the task seriously, they also were told that they had been randomly assigned to a group of 100 people and that their performance on the anagram task would be compared with the group. Participants were instructed that if their performance was in the top 20% of the group, they would earn a bonus of $0.25 for each word jumble they unscrambled. They were provided with a comprehension check to ensure that they understood the rules of self-scoring and additional payment, which was fundamental to our cheating measure. Any participant who failed these checks was excluded ($n = 65$). We decided on this exclusion criterion a priori and, given our past experience running this cheating task, we recruited more participants to have sufficient power after excluding participants. We were left with 189 participants (87 males and 102 females). Gender was evenly represented across the experimental conditions after exclusions, $\chi^2(1) = 0.04$, $p = .85$.

After reading a high- or low-obfuscation values statement about their group, participants were given 2 minutes to complete the self-scored anagram task in which they selected a radio button to indicate "solved" or "not solved" anagrams. Subjects were also told that any blank response would be marked as "not solved." We made three of the eight anagrams unsolvable (OPOER, ALVNO, ANHDU). Solvable anagrams and possible solutions included the following: (1) "TTISRA": "ARTIST"; (2) "SREETD": "RESTED"; (3) "LONSEM": "LEMONS"; (4) "EESPRMU": "SUPREME"; and (5) "TTEDES": "TESTED." We counted responses on the unsolvable anagrams as evidence of cheating.

Results

Participants who read a low-obfuscation values statement cheated less and claimed to solve fewer unsolvable anagrams ($M = 0.75$, $SD = 0.99$), as compared with participants who read a high-obfuscation values statement ($M = 1.05$, $SD = 1.05$); $t(187) = 2.03$, $p = .044$, $d = 0.29$. High levels of linguistic obfuscation can increase dishonest behavior.

To investigate whether our cheating results are specific to the materials used in this study, Study 3b reports a replication using different stimuli and a different cheating measure. We also perform a mechanism test by having participants complete a measure of resource depletion after reading their values statement. After expending high levels of cognitive effort, which is typical with obfuscated text, people may make less effortful decisions. A measure of depletion evaluated if possible cheating effects are due to energy expenditure or language.

Study 3b: Obfuscation Leads to Cheating With Number Matrices

Method

Participants. This study took place at a large Midwestern university in the United States. A total of 200 students participated in this online study for $5 pay ($M_{age} = 20.97$ years, $SD_{age} = 3.68$ years) with an opportunity to earn additional money based on their

performance on a task (up to $5). Our experiment had 121 females and 79 males. Gender was evenly represented across the experimental conditions, $\chi^2(1) = 1.86$, $p = .17$.

Procedure. This study began by telling participants to assume the role of a member of a research lab called the Behavioral Insights Group. Participants were told that they would read the values statement of the group and then complete several self-report measures.

After participants were randomly assigned to read a high- or low-obfuscation values statement (the statements used to disentangle obfuscation with both abstraction and jargon manipulated, located in Supplemental Study 2 available online), they completed a measure of resource depletion (Kouchaki & Smith, 2014). After reading a high- or low-obfuscation values statement, we asked participants, "Which of the following magazines would you most like to spend time reading, right now?" and had subjects select either the *New York Review of Books* (a "should" choice) or *People* magazine (a "want" choice). These choices required people to imagine expending different levels of effort while reading. That is, reading *People* magazine is simple and easy for most individuals, while reading the *New York Review of Books* is generally effortful and taxing. If more participants in the high-obfuscation condition selected the "want" option (e.g., *People*) than the "should" option (e.g., the *New York Review of Books*), this evidence would support the idea that reading a high-obfuscation values statement consumes more psychological resources and depletes more cognitive energy.

After the resource depletion question, participants solved number matrices as a performance measure, which involved finding two numbers in a 4 (row) × 3 (column) grid that sum to 10 (e.g., 5.81 + 4.19). If participants located two numbers that added to 10, they clicked a radio button, "Found it," and progressed to the next matrix and would earn $0.50. If participants did not find a matrix solution, they did not click a radio button. There were 10 trials during the matrix task, but three were unsolvable (Kouchaki & Smith, 2014). We counted responses on these three unsolvable matrices as cheating.

Results

Participants who read a low-obfuscation values statement cheated less and claimed to solve fewer unsolvable matrices, ($M = 1.44$, $SD = 1.22$), than participants who read a high-obfuscation values statement ($M = 1.81$, $SD = 1.25$); $t(198) = 2.12$, $p = .035$, $d = 0.30$. This evidence is consistent with our third prediction and Study 3a.

Participants in the high-obfuscation condition and low-obfuscation condition were equally likely to select the "want" versus the "should" reading option, $\chi^2(1) = 0.24$, $p = .62$. These data offer no evidence for a resource depletion mechanism.

General Discussion

This article provides a multidimensional view of linguistic obfuscation. The evidence from our studies using manipulated texts reached a clear consensus: People who read

a high-obfuscation values statement rated the company as less moral, less warm, and less trustworthy than did people who read a low-obfuscation values statement. Obfuscated text is not simply rated as less positive overall, however. Competence judgments were not modified by obfuscation, suggesting that the prior dimensions are indeed orthogonal (Goodwin et al., 2014).

We also demonstrated that people who read a high-obfuscation values statement of their group cheated more than did people who read a low-obfuscation values statement. Therefore, obfuscation can lead to a cycle, or "deception spiral," where obfuscated writing is both perceived as representing less ethical groups and also leads to less ethical behavior. The deception spiral describes a phenomenon where corporate communication reflects a company's unethical actions through linguistic obfuscation, which influences perceptions of the group as less moral, warm, and trustworthy, which in turn leads in-group members to act less ethically. While codes of conduct for employees are aimed at guiding ethical actions, they can have negative consequences for moral behavior because of how the messages are communicated.

This article makes several contributions. First, our package of studies, combining studies from the main text and online supplement, totaled nearly 5,000 participants and through a variety of tests, we extend a morality judgment paradigm to organizations instead of individuals or hypothetical situations (Haidt, 2001). People can accurately discriminate between high- and low-obfuscation text on morality dimensions in a corporate setting suggesting common, understandable language should replace complex speech.

Second, we provide evidence that people can accurately judge morality based on writing style. We extend fluency research by suggesting that function words can affect perceptions of a target and cheating behavior. Reading challenging content words often leads to negative perceptions, but our data suggest that the absence of function words such as articles (e.g., *a, the*) and prepositions is a source of discord as well. Why did such language modifications lead to perception and behavior changes? One potential mechanism offered by prior research suggests that the processing of certain function words and concrete content words may require less cognitive resources than abstract words because they are familiar and syntactically expected in language (Friederici et al., 2000). Indeed, function word processing often leads to a *prototypicality effect*, described as reduced brain activation when the familiarity of a word stimulus increases (Friederici et al., 2000; Raichle et al., 1994). Reading low-obfuscation text may lead to more favorable perceptions of a corporation because the text is familiar and fulfills the reader's expectations. More important, we show that the perceptions effect is related to dimensions in the message—that is, moral perceptions of the firm—and not other perceptions, such as competency.

Other takeaways from our research may have tangible consequences for corporations and related groups. Corporations should be incentivized to write in a nonobfuscated manner since people may act ethically after reading low-obfuscation text relative to high-obfuscation text. Otherwise, a values statement may backfire in its intended purpose of promoting a moral and ethical company. Our evidence suggests language matters, especially in high-stakes settings when deception might incur a cost for corporations.

Limitations and Future Directions

In our tests of the linguistic obfuscation hypothesis, we addressed several alternative explanations for the outcome that people perceive high-obfuscation texts as less moral, less warm, and less trustworthy than low-obfuscation texts. We explored if the length of the values statements or employee perspective-taking affected perceptions of immorality, and if fluency explained how people make judgments about a company (see Supplemental Material available online). We did not find evidence that length or perspective-taking change perceptions, and fluency only explains results related to warmth. Our tests were therefore limited to features that could be manipulated, but there may be others that should be explored as well.

Future versions of our perspective-taking manipulation should also include a manipulation check to ensure that participants identified with the organization if they were expected to adopt an in-group employee mentality (per the online supplement). Furthermore, it might be informative to collect additional demographic data (e.g., ethnicity, employment status) and investigate how they moderate the perceptions and cheating effects. The cheating tasks in Studies 3a and 3b were also low-stakes situations used to evaluate if obfuscation affects unethical behavior. Future research should investigate if obfuscation affects cheating behavior in high-stakes settings. It is important to consider how our findings can be scaled to understand how unethicality might have high-stakes financial stakes and social implications.

Conclusion

The writing style of corporations can have downstream perceptions-based and behavioral consequences. The evidence in this article suggests that high-obfuscation text leads to negative appraisals of an organization (e.g., people perceive the organization to be less moral, less warm, less trustworthy) than low-obfuscation text. High levels of obfuscation can also lead to cheating. Therefore, unethicality has a linguistic trace that affects how people appraise a company and their likelihood of engaging in unethical behavior, providing evidence for a worrisome deception spiral that can perpetuate unethical behavior. It is continually important to consider how corporations communicate, as their word patterns reveal social and psychological dynamics such as deception and further connect to how people think, feel, and behave.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

David M. Markowitz https://orcid.org/0000-0002-7159-7014

Supplemental Material

Supplemental material for this article is available online.

Note

1. Participants in this pilot study were recruited from Amazon Mechanical Turk, randomly assigned to read a high-obfuscation values statement ($n = 50$) or a low-obfuscation values statement ($n = 50$) and paid for their time. They rated how clear, complex, and understandable each text was, with scale ratings ranging from (1) *Not [clear/complex/did not understand] at all*, to (9) *Extremely [clear/complex/understood completely*]. The manipulation checks were successful, as participants who read a high-obfuscation values statement ($M = 7.40$, $SD = 1.62$) rated the writing as less clear than did those who read the low-obfuscation values statement, ($M = 8.08$, $SD = 1.50$); $t(98) = 2.18$, $p = .031$, Cohen's $d = 0.44$. Participants who read a high-obfuscation values statement ($M = 3.38$, $SD = 2.02$) perceived the text as more complex than did participants who read a low-obfuscation values statement, ($M = 1.98$, $SD = 1.36$); $t(85.96) = -4.06$, $p < .001$, $d = 0.81$. Finally, participants who read a high-obfuscation values statement rated the writing as less understandable ($M = 7.64$, $SD = 1.37$) than did participants who read a low-obfuscation value statement, ($M = 8.26$, $SD = 1.38$); $t(98) = 2.26$, $p = .026$, $d = 0.45$.

References

Ariely, D. (2012). *The (honest) truth about dishonesty: How we lie to everyone—especially ourselves*. HarperCollins.

Bird, F. B. (1996). *The muted conscience: Moral silence and the practice of ethics in business*. Quorum Books.

Bloomfield, R. J. (2002). The "incomplete revelation hypothesis" and financial reporting. *Accounting Horizons*, *16*(3), 233-243. https://doi.org/10.2139/ssrn.312671

Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, *10*(3), 214-234. https://doi.org/10.1207/s15327957pspr1003_2

Burgoon, J., Mayew, W. J., Giboney, J. S., Elkins, A. C., Moffitt, K., Dorn, B., Byrd, M., & Spitzley, L. (2016). Which spoken language markers identify deception in high-stakes settings? Evidence from earnings conference calls. *Journal of Language and Social Psychology*, *35*(2), 123-157. https://doi.org/10.1177/0261927X15586792

Chou, E. Y., Halevy, N., Galinsky, A. D., & Murnighan, J. K. (2017). The Goldilocks contract: The synergistic benefits of combining structure and autonomy for persistence, creativity, and cooperation. *Journal of Personality and Social Psychology*, *113*(3), 393-412. https://doi.org/10.1037/pspi0000096

Chung, C. K., & Pennebaker, J. W. (2007). The psychological functions of function words. In K. Fiedler (Ed.), *Social communication* (pp. 343-359). Psychology Press.

Clementson, D. E. (2018). Deceptively dodging questions: A theoretical note on issues of perception and detection. *Discourse & Communication*, *12*(5), 478-496. https://doi.org/10.1177/1750481318766923

Cohn, A., Fehr, E., & Maréchal, M. A. (2014). Business culture and dishonesty in the banking industry. *Nature*, *516*(7529), 86. https://doi.org/10.1038/nature13977

Courtis, J. K. (1998). Annual report readability variability: tests of the obfuscation hypothesis. *Accounting, Auditing & Accountability Journal*, *11*(4), 459-472. https://doi.org/10.1108/09513579810231457

Courtis, J. K. (2004). Corporate report obfuscation: artefact or phenomenon? *British Accounting Review*, *36*(3), 291-312. https://doi.org/10.1016/j.bar.2004.03.005

Dewan, T., & Myatt, D. P. (2008). The qualities of leadership: Direction, communication, and obfuscation. *American Political Science Review*, *102*(3), 351-368. https://doi.org/10.1017/S0003055408080234

Eisenberg, E. M. (1984). Ambiguity as strategy in organizational communication. *Communication Monographs*, *51*(3), 227-242. https://doi.org/10.1080/03637758409390197

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, *11*(2), 77-83. https://doi.org/10.1016/j.tics.2006.11.005

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, *32*(3), 221-233. https://doi.org/10.1037/h0057532

Friederici, A. D., Opitz, B., & von Cramon, D. Y. (2000). Segregating semantic and syntactic aspects of processing in the human brain: An fMRI investigation of different word types. *Cerebral Cortex*, *10*(7), 698-705. https://doi.org/10.1093/cercor/10.7.698

Garrett, N., Lazzaro, S. C., Ariely, D., & Sharot, T. (2016). The brain adapts to dishonesty. *Nature Neuroscience*, *19*(12), 1727-1732. https://doi.org/10.1038/nn.4426

George, G. (2005). Slack resources and the performance of privately held firms. *Academy of Management Journal*, *48*(4), 661-676. https://doi.org/10.5465/amj.2005.17843944

Gino, F., Ayal, S., & Ariely, D. (2009). Contagion and differentiation in unethical behavior. *Psychological Science*, *20*(3), 393-398. https://doi.org/10.1111/j.1467-9280.2009.02306.x

Gino, F., Schweitzer, M. E., Mead, N. L., & Ariely, D. (2011). Unable to resist temptation: How self-control depletion promotes unethical behavior. *Organizational Behavior and Human Decision Processes*, *115*(2), 191-203. https://doi.org/10.1016/j.obhdp.2011.03.001

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, *106*(1), 148-168. https://doi.org/10.1037/a0034726

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*(4), 814-834. https://doi.org/10.1037/0033-295X.108.4.814

Humpherys, S. L., Moffitt, K. C., Burns, M. B., Burgoon, J. K., & Felix, W. F. (2011). Identification of fraudulent financial statements using linguistic credibility analysis. *Decision Support Systems*, *50*(3), 585-594. https://doi.org/10.1016/j.dss.2010.08.009

Kacewicz, E., Pennebaker, J. W., Davis, M., Jeon, M., & Graesser, A. C. (2014). Pronoun use reflects standings in social hierarchies. *Journal of Language and Social Psychology*, *33*(2), 125-143. https://doi.org/10.1177/0261927x13502654

Kouchaki, M., Gino, F., & Feldman, Y. (2019). The ethical perils of personal, communal relations: A language perspective. *Psychological Science*, *30*(12), 1745-1766. https://doi.org/10.1177/0956797619882917

Kouchaki, M., & Smith, I. H. (2014). The morning morality effect:The influence of time of day on unethical behavior. *Psychological Science*, *25*(1), 95-102. https://doi.org/10.1177/0956797613498099

Larrimore, L., Jiang, L., Larrimore, J., Markowitz, D. M., & Gorski, S. (2011). Peer to peer lending: The relationship between language features, trustworthiness, and persuasion success. *Journal of Applied Communication Research*, *39*(1), 19-37. https://doi.org/10.1080/00909882.2010.536844

Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics*, *45*(2-3), 221-247. https://doi.org/10.1016/j.jacceco.2008.02.003

Maass, A., Salvi, D., Arcuri, L., & Semin, G. R. (1989). Language use in intergroup contexts: The linguistic intergroup bias. *Journal of Personality and Social Psychology*, *57*(6), 981-993. https://doi.org/10.1037/0022-3514.57.6.981

Mann, H., Garcia-Rada, X., Houser, D., & Ariely, D. (2014). Everybody else is doing it: Exploring social transmission of lying behavior. *PLOS ONE*, *9*(10), Article e109591. https://doi.org/10.1371/journal.pone.0109591

Margolin, D., & Markowitz, D. M. (2018). A multitheoretical approach to big text data: Comparing expressive and rhetorical logics in Yelp reviews. *Communication Research*, *45*(5), 688-718. https://doi.org/10.1177/0093650217719177

Markowitz, D. M. (2018). Academy Awards speeches reflect social status, cinematic roles, and winning expectations. *Journal of Language and Social Psychology*, *37*(3), 376-387. https://doi.org/10.1177/0261927x17751012

Markowitz, D. M. (2020). The deception faucet: A metaphor to conceptualize deception and its detection. *New Ideas in Psychology*, *59*, 100816. https://doi.org/10.1016/j.newideapsych.2020.100816

Markowitz, D. M., & Hancock, J. T. (2016). Linguistic obfuscation in fraudulent science. *Journal of Language and Social Psychology*, *35*(4), 435-445. https://doi.org/10.1177/0261927X15614605

Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, *45*(6), 633-644. https://doi.org/10.1509/jmkr.45.6.633

McCornack, S. A. (1992). Information manipulation theory. *Communication Monographs*, *59*(1), 1-16. https://doi.org/10.1080/03637759209376245

McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press.

Miceli, M. P., Near, J. P., & Dworkin, T. M. (2009). A word to the wise: How managers and policy-makers can encourage employees to report wrongdoing. *Journal of Business Ethics*, *86*(3), 379-396. https://doi.org/10.1007/s10551-008-9853-6

Mishina, Y., Dykes, B. J., Block, E. S., & Pollock, T. G. (2010). Why "good" firms do bad things: The effects of high aspirations, high expectations, and prominence on the incidence of corporate illegality. *Academy of Management Journal*, *53*(4), 701-722. https://doi.org/10.5465/amj.2010.52814578

Oppenheimer, D. M. (2006). Consequences of erudite vernacular utilized irrespective of necessity: Problems with using long words needlessly. *Applied Cognitive Psychology*, *20*(2), 139-156. https://doi.org/10.1002/acp.1178

Pennebaker, J. W. (2011). *The secret life of pronouns: What our words say about us*. Bloomsbury Press.

Pennebaker, J. W., Booth, R. J., Boyd, R. L., & Francis, M. E. (2015). *Linguistic inquiry and word count: LIWC2015*. Pennebaker Conglomerates.

Pennebaker, J. W., Chung, C. K., Frazee, J., Lavergne, G. M., & Beaver, D. I. (2014). When small words foretell academic success: The case of college admissions essays. *PLOS ONE*, *9*(12), Article e115844. https://doi.org/10.1371/journal.pone.0115844

Pollock, L. (2018). Statistical and methodological problems with concreteness and other semantic variables: A list memory experiment case study. *Behavior Research Methods*, *50*(3), 1198-1216. https://doi.org/10.3758/s13428-017-0938-y

Raichle, M. E., Fiez, J. A., Videen, T. O., MacLeod, A.-M. K., Pardo, J. V, Fox, P. T., & Petersen, S. E. (1994). Practice-related changes in human brain functional anatomy during nonmotor learning. *Cerebral Cortex*, *4*(1), 8-26. https://doi.org/10.1093/cercor/4.1.8

Riley, T. J., & Luippold, B. L. (2015). Managing investors' perception through strategic word choices in financial narratives. *Journal of Corporate Accounting & Finance*, *26*(5), 57-62. https://doi.org/10.1002/jcaf.22064

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*(1), 24-54. https://doi.org/10.1177/0261927X09351676

Weick, K. E. (1979). *The social psychology of organizing*. McGraw-Hill Humanities.

Author Biographies

David M. Markowitz is an assistant professor in the School of Journalism and Communication at the University of Oregon. He uses language data from natural repositories to make inferences about what people are thinking, feeling, and experiencing psychologically. He received his PhD from Stanford University and his undergraduate and master's degrees from Cornell University.

Maryam Kouchaki is an associate professor of Management and Organizations at Kellogg School of Management. Maryam is an organizational psychologist who seeks to understand everyday moral encounters, particularly at work. Her research is organized around two conceptual themes that involve 1) understanding the dynamic nature of moral decision-making and 2) understanding how individuals psychologically experience everyday moral encounters. Maryam examines these with a particular emphasis on the consequences of these encounters for individuals and groups.

Jeffrey T. Hancock is a professor in the Department of Communication at Stanford University. He works on understanding psychological and interpersonal processes in social media. He uses computational linguistics and experiments to understand how the words we use can reveal psychological and social dynamics, such as deception and trust, emotional dynamics, intimacy and relationships, and social support.

Francesca Gino studies how people can have more productive, creative and fulfilling lives. She is a professor at Harvard Business School and the author, most recently, of "*Rebel Talent: Why it Pays to Break the Rules in Work and Life*." Gino regularly gives keynote speeches, delivers corporate training programs and serves in advisory roles for firms and not-for-profit organizations across the globe.

PNAS RESEARCH ARTICLE | PSYCHOLOGICAL AND COGNITIVE SCIENCES

Even lawyers do not like legalese

Eric Martínez[a,1], Francis Mollica[b], and Edward Gibson[a]

Edited by Susan Goldin-Meadow, University of Chicago, Chicago, IL; received February 17, 2023; accepted April 8, 2023

Across modern civilization, societal norms and rules are established and communicated largely in the form of written laws. Despite their prevalence and importance, legal documents have long been widely acknowledged to be difficult to understand for those who are required to comply with them (i.e., everyone). Why? Across two preregistered experiments, we evaluated five hypotheses for why lawyers write in a complex manner. Experiment 1 revealed that lawyers, like laypeople, were less able to recall and comprehend legal content drafted in a complex "legalese" register than content of equivalent meaning drafted in a simplified register. Experiment 2 revealed that lawyers rated simplified contracts as equally enforceable as legalese contracts, and rated simplified contracts as preferable to legalese contracts on several dimensions–including overall quality, appropriateness of style, and likelihood of being signed by a client. These results suggest that lawyers who write in a convoluted manner do so as a matter of convenience and tradition as opposed to an outright preference and that simplifying legal documents would be both tractable and beneficial for lawyers and nonlawyers alike.

psycholinguistics | experimental jurisprudence | cognitive science | empirical legal studies | law and psychology

Significance

Why do lawyers write in such a convoluted manner? Across two preregistered experiments, we find that lawyers a) like laypeople, were less able to understand and recall "legalese" contracts than content of equivalent meaning drafted in a simplified register; and b) rated simplified contracts as equally enforceable as legalese contracts, and rated simplified contracts as preferable to legalese contracts on several important dimensions. Contrary to previous speculation, these results suggest that lawyers who write in a convoluted manner do so as a matter of convenience and tradition as opposed to an outright preference and that simplifying legal documents would be beneficial for lawyers and nonlawyers alike.

There is a burgeoning psycholinguistics literature documenting the various domains in which efficiency shapes human language, such that successful communication can be achieved with minimal effort on average by the sender and receiver (1–12). Two ways in which this efficiency manifests itself relate to word length and syntax. For example, words that are more frequent (such as "the") tend to be shorter than less frequent words (such as "accordion"), such that utterances tend not to be longer than necessary given one's communicative aims (13). With regard to syntax, it has been observed across languages that words that depend on each other tend to be close together in an utterance (14), so as to (by hypothesis) avoid overloading working memory capacity when interpreting an utterance.

However, one domain in which this efficiency has been attested to not apply is in the context of the legal system, as the language in contracts, statutes, and other legal documents is often observed to be notoriously inaccessible to a typical layperson, such that legal content seems to not be understood by a listener with minimal effort, e.g., refs. 15–22. Recent empirical work has supported the longstanding anecdotal observation/intuition that legal language is complex. For example, on a syntactic level, the language in contracts (23) and legislation (24) has been found to be laden with center-embedded clauses (leading to long-distance syntactic dependencies) at a rate several times higher than standard English texts, including academic articles and other texts aimed at an educated audience.

Meanwhile, on a word level, legal documents have also been found to be laden with words that are infrequently used in everyday speech. Previous research had long identified center-embedding (25, 26) and word frequency (27) to be reliable proxies for processing difficulty in normal texts. Recent work confirmed this to be true in legal documents, also, as contracts drafted with these features were recalled and comprehended at a lower rate than legal documents of equivalent meaning drafted without these features (and center-embedding, in particular, was found to inhibit recall to a greater degree than word frequency) (23).

While the above studies have shed insight into the question of how legal language is complicated to understand, it remains an open question why legal language is so complicated to understand–that is, why do lawyers write in such a convoluted manner in the first place? Answering this question is relevant not only to major questions in psycholinguistics but to legal doctrine and public policy as well.

Across modern civilization, societal norms and rules are established and communicated largely in the form of written laws. Because law is encoded in the form of natural language, it follows that an understanding of language is crucial to drafting, interpreting,

Author affiliations: [a]Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139; and [b]School of Informatics, University of Edinburgh, Edinburgh EH8 9AB, United Kingdom

Author contributions: E.M., F.M., and E.G. designed research; E.M. performed research; E.M. and F.M. analyzed data; E.G. supervised research; and E.M., F.M., and E.G. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: ericmart@mit.edu.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2302672120/-/DCSupplemental.

Published May 30, 2023.

and enforcing the rules and standards that comprise legal doctrine and underpin modern society. In particular, understanding why lawyers and lawmakers write in such a convoluted manner can help inform policy efforts to make laws more accessible–which have been advocated for decades (28–30), with little to no success (24). Such efforts are crucial to ensuring the comprehension and compliance of societal norms, as well as upholding the legitimacy of legal doctrines that either expressly assert or implicitly assume that legal documents are or ought to be easily interpretable to laypeople, such as *Ordinary Meaning* (31–33) and *Fair Notice* (34).

Here, we conducted two well-powered, preregistered experiments aimed at evaluating five hypotheses presented in the theoretical literature for why lawyers write the way that they do. In Experiment 1, we found that lawyers, like laypeople, were less able to recall and comprehend legal content drafted in a complex "legalese" register than content of equivalent meaning drafted in a simplified register. In Experiment 2, we found that lawyers rated simplified contracts as equally enforceable as legalese contracts, and rated simplified contracts as preferable to legalese contracts on several dimensions–including overall quality, appropriateness of style, and likelihood of being signed by a client. These results suggest that lawyers who write in a convoluted manner do so as a matter of convenience and tradition as opposed to an outright preference and that simplifying legal documents would be beneficial for lawyers and nonlawyers alike.

Hypotheses

In previous literature, scholars proposed several hypotheses for why lawyers write in a complicated manner. Here, we briefly present each of these hypotheses in turn, as well as the associated predictions of these hypotheses that we preregistered for our experiments.

Curse of Knowledge Hypothesis. Some scholars have speculated, in line with what has been dubbed the "curse of knowledge" in other disciplines (35, 36), that the reason legal language is so difficult to understand is because lawyers do not realize that they write in an esoteric manner (37). If this were true, one would predict that lawyers would not show the same degree of difficulty as laypeople in understanding complicated legal texts relative to simplified legal texts and that lawyers would underestimate how difficult legalese texts are for laypeople.

Copy-and-Paste Hypothesis. Some commentators have speculated that lawyers simply write in a complex register out of "habit, laziness" (16) or respect for "tradition" (38), that they "copy and paste" (39) from existing templates with old, complicated terms because that's the "quickest and cheapest way to produce a contract" (40). If this hypothesis were true, one would expect that lawyers would rate plain-English contracts as of equal quality as legalese contracts and that lawyers would be equally likely to agree to sign off on a contract written in a simpler register written by someone else as they would for a contract written in a legal register.

In-Group Signaling Hypothesis. Some commentators have hypothesized that lawyers write in legalese to be accepted by their peers, to sound more "lawyerly," to "mark themselves as members of the profession" (16). If so, one would predict that lawyers would rate contracts written in legalese as sounding more appropriate/suitable for a lawyer than those written in plain English, and would rate the author of that contract as more hirable than the author of a plain-English contract.

It's Just Business Hypothesis. Some commentators have hypothesized that lawyers write in legalese as a way of "preserving their monopoly" (41) on legal services and "justifying fees" (16). If this hypothesis were true, one would predict that lawyers would rate contracts written in legalese as being more likely to be signed by clients than contracts written in a simple register.

Complexity of Information Hypothesis. Some have speculated that legal language needs to be complex in order to satisfy certain communicative aims, such as conveying complex legal concepts in a way that "is far more precise than ordinary language" (38), to avoid ambiguity, and/or to ensure enforceability. To evaluate this hypothesis, we constructed a question that asked whether a given contract excerpt was enforceable. If this hypothesis were true, one would predict that lawyers would rate simplified contracts as unenforceable or lower quality than complicated contracts.

Results

Experiment 1. In Experiment 1, we evaluated the curse of knowledge hypothesis.*

To evaluate the predictions of this hypothesis, we conducted a preregistered experiment in which we evaluated lawyers' ($n = 105$) comprehension and recall of two types of legal contracts. The first set, "legalese" contracts, were written in a style containing linguistic features that have been shown to be disproportionately common in legal texts relative to nonlegal texts, and which have also been shown to inhibit recall and comprehension of legal content relative to contracts without these features. The second set, "plain-English" contracts, were of equivalent meaning drafted without these difficult-to-process features. We analyzed lawyers' performance alongside a reanalysis of Martinez, Mollica and Gibson's (23) experiment of laypeople ($n = 108$) that used an identical set of materials and procedure.

Results are visualized in Figs. 1 and 2. Contrary to the predictions of the curse of knowledge hypothesis, we observed a main effect of legal training and register on recall ($\beta = 0.353$, SE = 0.159, P = 0.026) and comprehension ($\beta = 0.808$, SE = 0.100, P < 0.001), but not an interaction between register and legal training on recall ($P = 0.360$) or comprehension ($P = 0.638$). That is, although lawyers were significantly better than laypeople at comprehending and recalling legal content overall in our materials, both lawyers and laypeople were better at comprehending ($\beta = 0.354$, SE = 0.088, $P < 0.001$) and recalling ($\beta = 0.360$, SE = 0.121, $P = 0.003$) plain-English texts than legalese texts, and there was no evidence that lawyers were disproportionately better than laypeople at comprehending ($P = 0.638$) or recalling ($P = 0.360$) legal content in legalese texts relative to plain English.

We observed converging results when comparing lawyer and layperson's subjective difficulty ratings of each text, as lawyer participants' predictions of how difficult a text would be for the average layperson did not significantly differ from those of lay participants. *SI Appendix* for details.

Experiment 2. In Experiment 2, we sought to evaluate the predictions associated with the four remaining hypotheses: the

*The preregistration for Experiment 1 can be viewed at the following link: https://osf.io/y8xjd/.

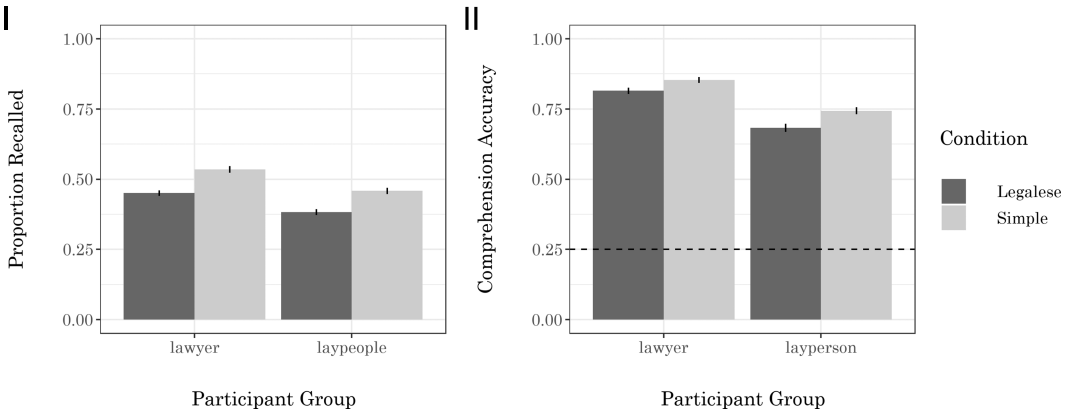


Fig. 1. Proportion of legal content recalled (*i*) and comprehended (*ii*) in legalese and simple contracts by lawyer and nonlawyer participants. Error bars represent 95% bootstrapped confidence intervals. The dotted line in (*ii*) represents chance performance in comprehension task.

in-group signaling hypothesis, the it's just business hypothesis, the complexity of information hypothesis, and the copy-and-paste hypothesis. To do so, we presented lawyers (n = 102) with the same set of contracts used in Experiment 1, and asked them to rate the contracts on a variety of dimensions, including overall quality and enforceability of the contract, hireability of the author who wrote the contract, willingness to sign off on the contract as written, and likelihood that a client would agree to the contract's terms.[†]

Results of Experiment 2 are visualized in Fig. 3. In line with all of the preregistered predictions of the copy-and-paste hypothesis and against all of the preregistered predictions of the in-group signaling, it's just business and complexity of information hypotheses, lawyers rated contracts written in plain-English as significantly higher quality (β =1.705, SE = 0.329, $P < 0.001$) and no less enforceable than legalese contracts (P = 0.717); rated the authors of plain-English contracts as significantly more hirable than those of legalese contracts (β = 1.835, SE = 0.318, $P < 0.001$); were significantly more likely to say that they would agree to use the contract as-written (β = 1.432, SE = 0.270, $P < 0.001$); and predicted that clients would be significantly more likely to sign plain-English contracts than legalese contracts (β = 1.232, SE = 0.338, $P < 0.001$).

The results of both experiments were robust to all measured demographic variables, including race, gender, age, years of practice experience, and "fanciness" of lawyer (see definition in *Methods*). These results are reported and visualized in *SI Appendix*.

Discussion

This study attempts to empirically investigates the long-puzzling question of why lawyers write the way that they do, undermining most prior accounts of the cognitive origins of legalese. For example, some commentators have maintained that lawyers prefer or are otherwise forced to write in a complex manner in order to satisfy certain communicative aims, to sound more lawyerly, or to justify exorbitant fees to clients. Others have speculated that lawyers simply do not realize they are writing in a complicated manner due to how easy it is for them to understand. In contrast, the fact that lawyers in our studies rated plain-English contracts as higher quality, even more likely to be signed by clients and no less enforceable than legalese contracts, and rated the authors of plain-English contracts as more hirable than authors of legalese contracts undermines both of these sets

[†]Preregistration for Experiment 2 can be viewed here: https://osf.io/b98j5/.

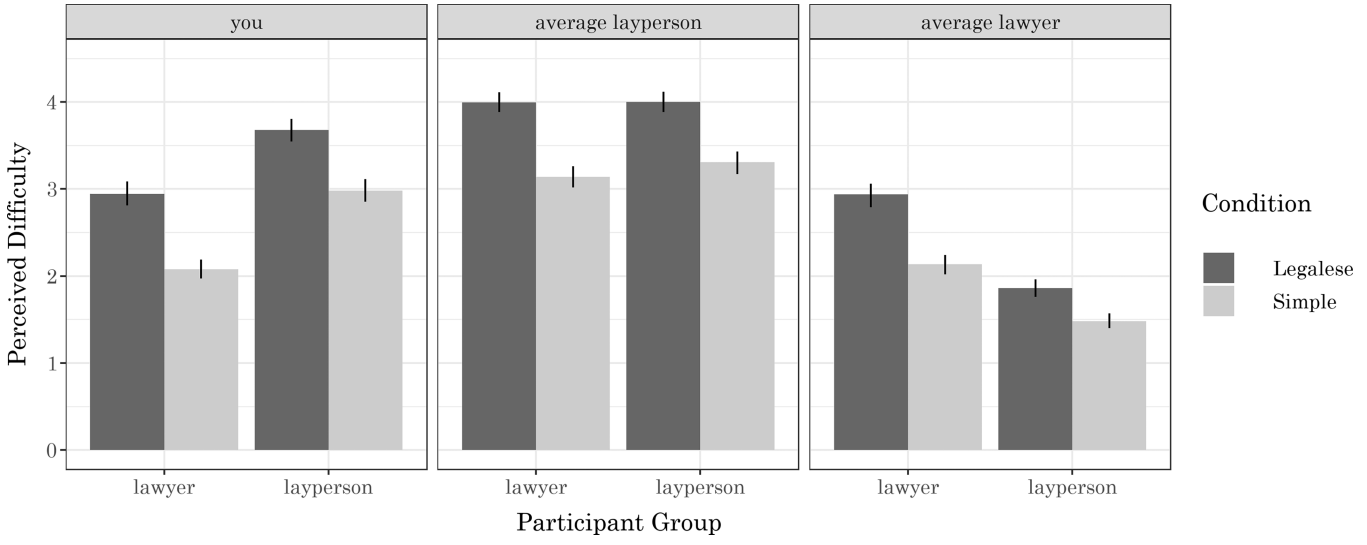


Fig. 2. Subjective difficulty ratings by lawyer and lay participants regarding how difficult participants found a given text (*A*) for themselves (*Left* panel); (*B*) for the average layperson (*Middle* panel); and (*C*) the average lawyer (*Right* panel).

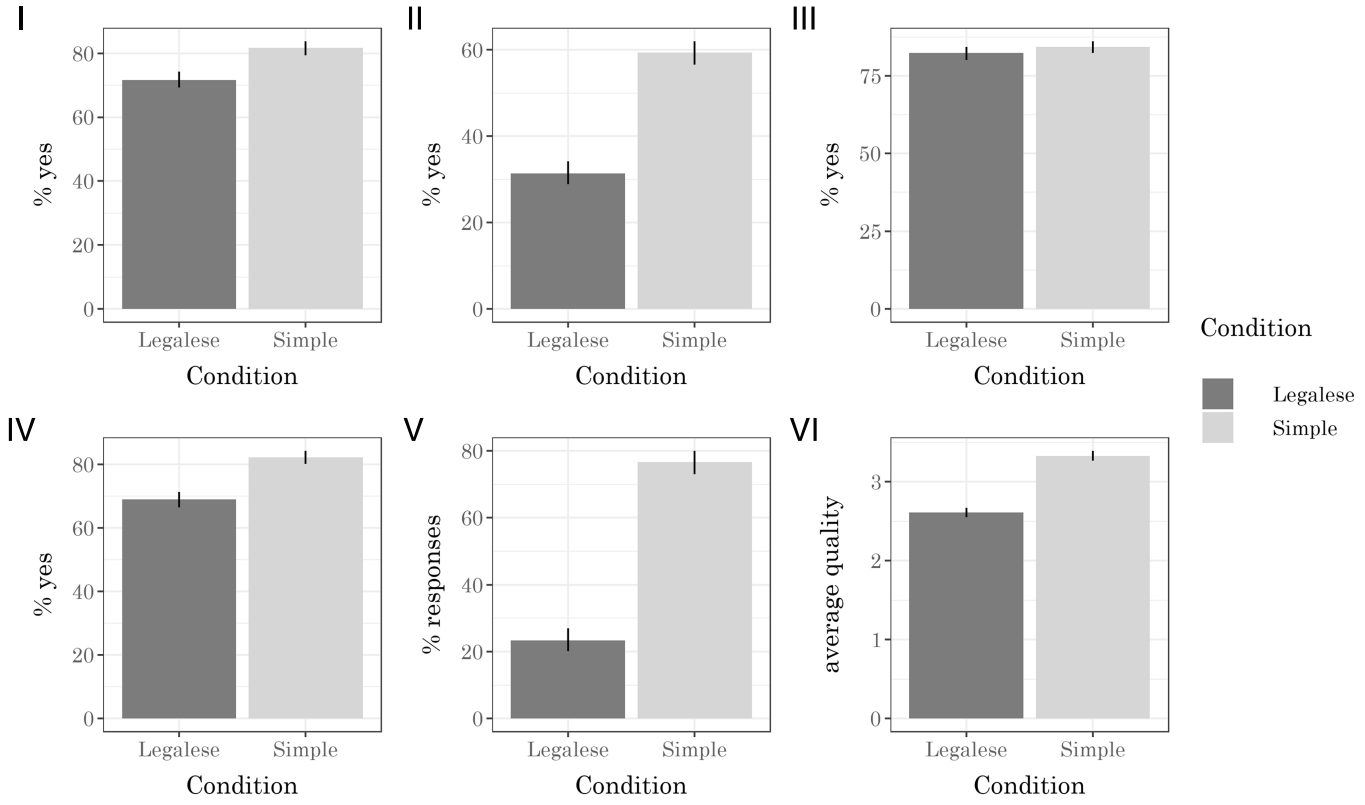


Fig. 3. Results of lawyer responses to questions regarding the quality of legalese and simple contracts according to a series of desiderata, including (*i*) appropriateness of style, (*ii*) hireability of author, (*iii*) enforceability of document, (*iv*) likelihood of document being signed by client, (*v*) willingness to use document as written, and (*vi*) overall quality of document.

of hypotheses, suggesting that in many instances lawyers both can and prefer to write in a more understandable manner as opposed to being bound by the nature of law, or engaging in a "conspiracy of gobbledygook." This study attempts empirically investigates the long-puzzling question of why lawyers write the way that they do, undermining most prior accounts of the cognitive origins of legalese. For example, some commentators have maintained that lawyers prefer or are otherwise forced to write in a complex manner in order to satisfy certain communicative aims, to sound more lawyerly, or to justify exorbitant fees to clients. Others have speculated that lawyers simply do not realize they are writing in a complicated manner due to how easy it is for them to understand. In contrast, the fact that lawyers in our studies rated plain-English contracts as higher quality, even more likely to be signed by clients and no less enforceable than legalese contracts, and rated the authors of plain-English contracts as more hirable than authors of legalese contracts undermines both of these sets of hypotheses, suggesting that in many instances lawyers both can and prefer to write in a more understandable manner as opposed to being bound by the nature of law, or engaging in a "conspiracy of gobbledygook."

Meanwhile, the fact that lawyers rated both contracts as enforceable and likely to be signed by clients but preferred plain-language contracts on several dimensions suggests, consistent with what we have dubbed the "copy-and-paste" hypothesis, that lawyers may simply draw from old, preexisting templates laden with arcane and convoluted language due to that being easier and cheaper to produce than drafting a simpler contract from scratch. This finding is consistent with recent empirical work indicating that lawyers rely heavily on templates in drafting contracts, with future agreements only rarely deviating from previous ones even when deviations would apparently benefit the involved parties(42). In addition to cost, said stickiness may also be borne out of lawyers' training in the importance of precedent, which overall might lead to an adherence to templates laden with old, archaic language by virtue of the fact (or assumption) that they worked before, and that the specific language may have been "defended in court" previously.

From a policy perspective, our results also provide insight into the longstanding question of how to make legal language more understandable. Although for decades, the US government has engaged in top–down efforts to simplify public legal documents for the benefit of society at large (Plain Language Action Information Network, 2011; USC 301, 2010), recent work has revealed these efforts to have failed, as laws, like contracts, remain laden with difficult to process features such as center-embedding and low-frequency words (24). While this failure may lead some to conclude that simplifying legal language is an intractable affair, our results paint a more optimistic picture, suggesting that lawyers a) believe legal documents can and should be simplified to better serve their communicative aims; and b) like laypeople, struggle to comprehend complex legal language relative to a simpler alternative. Our results further suggest that the processing difficulty of legal texts may be alleviated as lawyers and lawmakers become more aware of both the ways in which public legal documents tend to be complex, as well as the alternatives available to them in order to make them less complex.

It is worth noting that our results do not imply that legal documents can be simplified limitlessly without sacrificing communicative aims, nor do we discount the role of formality in legal writing. Like other professionals, lawyers may use a more formal tone in legal documents in order to, for example: a)

demonstrate their status of members of the legal community, which may require convergence on a style that is identifiable and replicable, and b) signal to a reader that a text should be taken seriously an official legal document as opposed to a form of casual, nonbinding communication.

Instead, our results indicate that such formality is not necessarily synonymous with complexity. That is, in many cases, lawyers can and should adopt a simpler register in order to achieve a level of formality that best aligns with their communicative aims as opposed to burdening clients and themselves with obfuscatory legalese.

Constraints on Generality. Examining the participant sample, the stimuli, and general experimental design suggests that the results of the present study would likely generalize to a broad array of relevant real-world scenarios.

With regard to participants, our sample included a large number of lawyers that, according to available estimates (43, 44), were broadly representative of the legal profession with regard to a number of demographic factors, including age, ethnicity, gender, years of legal experience, and type of legal employment. Analyses further revealed that our results were the same when controlling for these demographic variables in our analysis, such that we expect the results to generalize to the broader population of United States lawyers. It is unclear whether they would generalize to the legal profession in other countries.

With regard to materials, our stimuli consisted of a diverse array of contract excerpts whose content mirrored the most common types of clauses found in private legal documents in the United States (45, 46), and private legal documents are among the most common types of legal documents that people are exposed to. The linguistic features we looked at have been found to be disproportionately prevalent in both private legal documents (e.g., contracts) and public legal documents (e.g., laws) relative to other forms of written and spoken English (23, 47). Thus, we expect the results to generalize to other types of legal documents beyond those examined in the present study, though it is likely that some types of provisions will be less amenable to simplification than those used in the present study.

Regarding the ecological validity of the design, one might wonder whether lawyers' responses to questions in a hypothetical setting would generalize to real-world behavior. Given that an important role of a lawyer in the real world is to reason about hypothetical scenarios and engage in counterfactual reasoning, the fact that our experimental design asked lawyers to reason about hypothetical scenarios and engage in counterfactual reasoning would seem to imply that our study was well-aligned with the job of a lawyer in the real world. By extension, this would suggest that our design was an ecologically valid way to test our hypotheses.

A related concern relates to whether there was a performative element—if lawyers know they are subjects in an experiment and are being observed by scientists, maybe they will behave differently than in the real world. Although this is an important concern, we have no reason to expect that lawyers knew what result we were interested in, given that: a) we did not give away the specific research question we were interested in when recruiting lawyers for our study and b) we ensured that lawyers were unaware of register manipulation during the experiment. Supposing that lawyers did not know what result we were interested in, we also have no reason to expect that their behavior was systematically influenced to help the researchers get a desired result. Thus, we have no reason to expect that a potential performative element drove our results.

Experiment 1

Materials. The primary materials consisted of 12 pairs of short contract excerpts of roughly 150 words each (Fig. 4). Each pair contained of a) one excerpt drafted in a legalese register, containing features identified by previous studies to be strikingly more prevalent in legal texts relative to nonlegal texts, including center-embedded clauses, low-frequency jargon, nonstandard capitalization, and passive-voice structures; and b) one excerpt drafted in a simple register, identical in content to the other excerpt but without the above features.

For each contract pair, 12 to 15 comprehension questions were drafted in a "neutral" register. In addition to the main experimental materials, we also implemented the author recognition task [ART; (48, 49)] as a measure of individual differences in experience with language.

Participants and Procedure. US attorneys (n = 106) were recruited to participate as subjects in our experiment, through a combination of direct email invitations, word-of-mouth recruitment, and social media posts. Participants received $100 for their participation in the study. Participants were retained in our analysis as long as they were licensed to practice law in the United States. Participants were required to enter an official law school or law firm email, or provide their official bar number in order to help verify their attorney status. Of the 106 participants, 105 were verified to be attorneys and were retained in the final analysis.

With regard to demographics, the mean age of retained participants was 34 (median: 31). In total, 60.8% of participants identified as male, and 38.2% identified as non-White. Participants had a mean of 5.9 years of practice experience. Of note, 50.9% of the sample were coded as "fancy" lawyers, meaning that they either a) graduated from a top-25 law school according

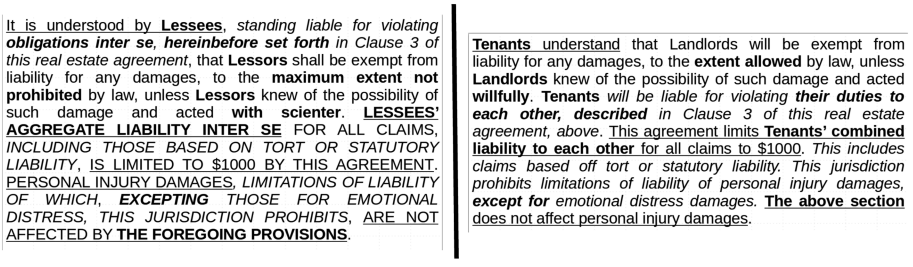


Fig. 4. An example stimulus pair in legalese (*Left*) and simple (*Right*) register. The differences in surface properties across registers are depicted by font style. Bold denotes word frequency. Italic denotes embedded clauses. Underlined denotes voice. Unfortunately, we have run out of font styles to make differences in capitalization more apparent. Image reprinted from ref. 23 *SI Appendix*.

to US News and World Report or b) worked at a top-200 law firm according to American Lawyer (AmLaw) magazine.[‡]

Retained participants were psuedorandomly assigned to six trials (3 legalese; 3 simple). Participants did not see the same contract in both a simple and legal register. Assignment of stimuli to participants was pseudorandom to ensure that across participants, every trial was administered with approximately the same frequency. The order of trials was randomized for each participant.

A trial consisted of a) reading an excerpt, b) a subset of the ART, c) recalling the excerpt, and d) answering comprehension questions. For the reading component, participants were presented with exactly one excerpt, written in either legalese or plain English. They were asked to carefully read the text twice and were given as much time as needed to do so. For the ART component, participants were given the names of 50 individuals and were asked to select which names corresponded to real authors. We expanded the ART task to 300 trials in order to keep the timing of a trial consistent. The original items from the published ART were presented first. For the remaining trials, the participants were administered items that looked virtually the same as authentic materials (half of the names corresponding to real authors, the other half corresponding to high-school track stars). We do not use these items in our analysis as they have not been validated (50). After being shown the ART materials, participants were asked to recall as much of the excerpt they had read as possible. They were told that they could use their own words, but that their version should stay true to the original. Finally, each trial ended with the comprehension questions corresponding to the excerpt.

Analysis Plan. Following Martinez et al. (23), two trained research assistants coded whether a proposition was successfully recalled (*SI Appendix* for details). Coders were unaware of whether a participant had seen or recalled the simple or legalese version of a text. Twenty percent of the retellings were coded by both coders so as to assess interrater reliability using Cohen's kappa coefficient (51, 52). For our regression analyses, we perform both a conservative analysis and an anticonservative analysis, with regard to ties. Our results do not qualitatively change, so we report only the conservative analysis in text (*SI Appendix* for anticonservative analysis).

Experiment 2

Materials. Our primary materials consisted of the same 12 pairs of short contract excerpts as those used in Study 1. In addition, we also constructed a series of questions aimed at testing specific hypotheses for why lawyers write the way that they do. Here, we discuss each of these questions in turn. The full list of questions, as well as the experimental interface, is provided in *SI Appendix*.

Copy-and-paste hypothesis. To test this hypothesis, we constructed a question that asked participants to rate the quality of a given contract excerpt (in plain English or legalese), as well as another question that asked participants whether they would agree to sign off on a given contract excerpt assuming it were written by someone else.

In-Group membership hypothesis. To test this hypothesis, we constructed two types of questions: one that asks whether the style of a particular excerpt sounds appropriate for a lawyer, and another that asks whether a participant would hire the author of the excerpt.

It's just business hypothesis. To evaluate this hypothesis, we constructed a question that asked participants to rate whether a client would be likely to sign a particular contract excerpt.

Complexity of information hypothesis. To evaluate this hypothesis, we constructed a question that asked whether a given contract excerpt was enforceable. We constructed a question that asked participants to rate the quality of a given contract excerpt (in plain English or legalese)

Participants and procedure.

US attorneys (n = 105) were recruited to participate as subjects in our experiment through similar means as Study 1. Participants received $40 for their participation in the study, and were retained in the analysis using the same criteria as Study 1.

With regard to demographics, the mean age of retained participants was 35.7 (median: 33). In total, 62.7% of participants identified as male, and 38.2% identified as non-White. Participants had a mean of 8.3 y of practice experience (median: 5.5), and 40.2% of the sample were coded as "fancy" lawyers.

With regard to procedure, retained participants were psuedorandomly assigned to six trials. Assignment of stimuli to participants was pseudorandom to ensure that across participants every trial was administered with approximately the same frequency. The order of trials was randomized for each participant. Within each trial, participants were first presented with one version of a contract excerpt in either legalese or plain English, and asked to answer several questions about it. Participants were then presented with the other version of the contract excerpt and asked to answer the same questions about it. Participants were then shown the two versions side-by-side and asked to answer several questions about the two versions in tandem.

Ethics Approval. Both experiments were approved by MIT's Committee on the Use of Humans as Experimental Subjects (COUHES), protocol number: 2107000425. Prior to completing each experiment, participants were shown a consent form, which provided further details about the experiment, including risks, as well as information about COUHES approval.

Data, Materials, and Software Availability. Anonymized data have been deposited in public (https://osf.io/dmkrx/).

ACKNOWLEDGMENTS. We thank Sofie Cheung, Sofia Zhang, Euphy Liu, and Anita Podrug for help with coding the recall responses. Thanks go to Nancy Kanwisher, Roger Levy, Rebecca Saxe, Ev Fedorenko, and Laura Schulz for helpful feedback and discussions regarding the hypotheses and experimental design. Thanks to audiences at TedLab and EvLab, the MIT Brain & Cognitive Sciences "Cognitive Lunch" talk series, and the Human Sentence Processing conference in Pittsburgh, PA, March 2023. And finally, thanks go to the editor, one anonymous reviewer, and Adele Goldberg for their comments on the submitted paper.

[‡]This was determined based on the email participants provided when taking the study.

1. E. Gibson *et al.*, How efficiency shapes human language. *Trends Cognit. Sci.* **23**, 389–407 (2019).
2. S. T. Piantadosi, H. Tily, E. Gibson, Word lengths are optimized for efficient communication. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 3526–3529 (2011).
3. S. T. Piantadosi, H. Tily, E. Gibson, The communicative function of ambiguity in language. *Cognition* **122**, 280–291 (2012).
4. S. T. Piantadosi, Zipf's word frequency law in natural language: A critical review and future directions. *Psychon. Bull. Rev.* **21**, 1112–1130 (2014).
5. K. Mahowald, E. Fedorenko, S. T. Piantadosi, E. Gibson, Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition* **126**, 313–318 (2013).
6. K. Mahowald, I. Dautriche, E. Gibson, S. T. Piantadosi, Word forms are structured for efficient use. *Cognit. Sci.* **42**, 3116–3134 (2018).

7. R. Futrell, E. Gibson, L2 processing as noisy channel language comprehension. *Bilingualism: Lang. Cognit.* **20**, 683–684 (2017).
8. R. Futrell, R. Levy, "Noisy-context surprisal as a human sentence processing cost model" in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers* (2017), vol. 1, pp. 688–698.
9. R. Ryskin, R. Futrell, S. Kiran, E. Gibson, Comprehenders model the nature of noise in the environment. *Cognition* **181**, 141–150 (2018).
10. Y. Zhang, R. Ryskin, E. Gibson, A noisy-channel approach to depth-charge illusions. *Cognition* **232**, 105346 (2023).
11. E. Gibson *et al*., A noisy-channel account of crosslinguistic word-order variation. *Psychol. Sci.* **24**, 1079–1088 (2013).
12. E. Gibson, L. Bergen, S. T. Piantadosi, Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 8051–8056 (2013).
13. G. K. Zipf, *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology* (Addison-Wesley Press, 1949).
14. R. Futrell, K. Mahowald, E. Gibson, Large-scale evidence of dependency length minimization in 37 languages. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 10336–10341 (2015).
15. M. E. Masson, M. A. Waldron, Comprehension of legal contracts by non-experts: Effectiveness of plain language redrafting. *Appl. Cognit. Psychol.* **8**, 67–85 (1994).
16. P. Tiersma, "The nature of legal language" in *Dimensions of Forensic Linguistics* (John Benjamins Publishing Company, 2008), pp. 7–25.
17. A. Elwork, B. D. Sales, J. J. Alfini, *Making Jury Instructions Understandable* (Michie Company, 1982).
18. R. Rogers, K. S. Harrison, D. W. Shuman, K. W. Sewell, L. L. Hazelwood, An analysis of Miranda warnings and waivers: Comprehension and coverage. *Law Hum. Behav.* **31**, 177–192 (2007).
19. L. Heuer, S. D. Penrod, Instructing jurors. *Law Hum. Behav.* **13**, 409–430 (1989).
20. N. E. S. Goldstein, L. O. Condie, R. Kalbeitzer, D. Osman, J. L. Geier, Juvenile offenders' Miranda rights comprehension and self-reported likelihood of offering false confessions. *Assessment* **10**, 359–369 (2003).
21. S. S. Diamond, B. Murphy, M. R. Rose, The kettleful of law in real jury deliberations: Successes, failures, and next steps. *Nw. U. L. Rev.* **106**, 1537 (2012).
22. R. P. Charrow, V. R. Charrow, Making legal language understandable: A psycholinguistic study of jury instructions. *C. Law Rev.* **79**, 1306–1374 (1979).
23. E. Martínez, F. Mollica, E. Gibson, Poor writing, not specialized concepts, drives processing difficulty in legal language. *Cognition* **224**, 105070 (2022).
24. E Martinez, F Mollica, E Gibson, So much for plain language: An analysis of the accessibility of United States Federal Laws over time. Available at SSRN: https://ssrn.com/abstract=4036863 (2022).
25. N. Chomsky, On certain formal properties of grammars. *Inf. Control* **2**, 137–167 (1959).
26. E. Gibson, Linguistic complexity: Locality of syntactic dependencies. *Cognition* **68**, 1–76 (1998).
27. C. B. Marks, M. J. Doctorow, M. C. Wittrock, Word frequency and reading comprehension. *J. Edu. Res.* **67**, 259–262 (1974).
28. Plain Language Action Information Network, *Federal Plain Language Guidelines* (2011).
29. Exec. Order No. 13648, 44 fr 69609 (1979).
30. Plain Writing Act of 2010, ("5 U.S.C. §301. 2010").
31. B. G. Slocum, "Ordinary meaning" in *Ordinary Meaning* (University of Chicago Press, 2015).
32. B. G. Slocum, J. Wong, The Vienna convention and ordinary meaning in international law. *Yale J. Int. L.* **46**, 191 (2021).
33. K. Tobia, "Law and the cognitive science of ordinary concepts in *Law and Mind: A Survey of Law and the Cognitive Sciences* (Cambridge University Press, 2021).
34. Mcboyle v. United States (1931).
35. R. S. Nickerson, How we know- and sometimes misjudge-what others know: Imputing one's own knowledge to others. *Psychol. Bull.* **125**, 737 (1999).
36. P. J. Hinds, The curse of expertise: The effects of expertise and debiasing methods on prediction of novice performance. *J. Exp. Psychol.: Appl.* **5**, 205 (1999).
37. S. Pinker, *The Sense of Style: The Thinking Person's Guide to Writing in the 21st Century* (Penguin Books, 2015).
38. P. Tiersma, Some myths about legal language. *J. Law, Cult. Hum.* Forthcoming, Loyola-LA Legal Studies Paper (2005).
39. K. A. Adams, *A Manual of Style for Contract Drafting* (American Bar Association, ed. 4, 2017).
40. C. A. Hill, A comment on language and norms in complex business contracting. *Chi.-Kent L. Rev.* **77**, 29 (2001).
41. D. Mellinkoff, *The Language of the Law* (Wipf and Stock Publishers, 2004).
42. J. Nyarko, Stickiness and incomplete contracts. *U. Chi. L. Rev.* **88**, 1 (2021).
43. AB Association, Aba profile of the legal profession (2020).
44. E. Martínez, K. Tobia, What do law professors believe about law and the legal academy? An empirical inquiry (2022).
45. A. L. Institute, *Restatement (Second) of Contracts* (American Law Institute, St. Paul, MN, 1981).
46. AL Institute, *NC of Commissioners on Uniform State Laws* (Uniform Commercial Code, 2020).
47. E. Martinez, F. Mollica, Y. Liu, A. Podrug, E. Gibson, "What did I sign? A study of the impenetrability of legalese in contracts" in *Proceedings of the Annual Meeting of the Cognitive Science Society* (2021), vol. 43.
48. K. E. Stanovich, R. F. West, Exposure to print and orthographic processing. *Reading Res. Q.* **24**, 402–433 (1989).
49. M. Moore, P. C. Gordon, Reading ability and print exposure: Item response theory analysis of the author recognition test. *Behav. Res. Methods* **47**, 1095–1109 (2015).
50. D. J. Acheson, J. B. Wells, M. C. MacDonald, New and updated tests of print exposure and reading abilities in college students. *Behav. Res. Methods* **40**, 278–289 (2008).
51. J. Cohen, A coefficient of agreement for nominal scales. *Ed. Psychol. Meas.* **20**, 37–46 (1960).
52. M. L. McHugh, Interrater reliability: The kappa statistic. *Biochem. Med.* **22**, 276–282 (2012).

NEUROSCIENCE

Reading dies in complexity: Online news consumers prefer simple writing

Hillary C. Shulman[1]*†, David M. Markowitz[2]†, Todd Rogers[3]

Over 30,000 field experiments with *The Washington Post* and *Upworthy* showed that readers prefer simpler headlines (e.g., more common words and more readable writing) over more complex ones. A follow-up mechanism experiment showed that readers from the general public paid more attention to, and processed more deeply, the simpler headlines compared to the complex headlines. That is, a signal detection study suggested readers were guided by a simpler-writing heuristic, such that they skipped over relatively complex headlines to focus their attention on the simpler headlines. Notably, a sample of professional writers, including journalists, did not show this pattern, suggesting that those writing the news may read it differently from those consuming it. Simplifying writing can help news outlets compete in the competitive online attention economy, and simple language can make news more approachable to online readers.

INTRODUCTION

How do people select what to read in competitive online news environments? Democratic societies prize a knowledgeable and engaged citizenry, which requires that people educate themselves on the most important and most credible news of the day. In reality, however, even though high-quality news has never been more available, so too is the competition for readers' attention (*1*). The competition for online attention is fierce. High-quality news must compete for reader attention with misinformation (*2*, *3*) and the proliferation of highly partisan content (*4*–*6*). Against this backdrop, we propose the simpler-writing heuristic as a way of explaining reading behavior in online news environments. Guided by the principle that people are economical with their attention (*7*), we propose that news headlines featuring simpler language will be clicked on, and consequently read, more than news headlines with more complex language. This research sheds light on how people navigate information-rich environments (*8*, *9*), with implications for how news ecosystems can better achieve democratic ideals (*10*).

Evidence across fields and approaches supports our prediction that people prefer simpler news headlines over more complex ones. Experimental evidence suggests simpler texts are rated more positively (*11*) and are engaged with more often (*12*, *13*) than complex texts. Correlational field studies show similar patterns for online engagement in the form of likes and views (*14*), although their implications typically lack strong causal confidence (*15*). Field experiments find that simpler documents (*16*), simpler disclosures (*17*), and simpler applications (*18*) can increase response rates and improve downstream outcomes like showing up for court appearances, signing up for insurance programs, and submitting federal forms. Given the strengths and limitations of this evidence, our understanding of simple writing's superiority would benefit from large-scale, ecologically valid experimental evidence in the wild.

The simpler-writing heuristic posits that in competitive information environments (e.g., websites with several headlines to select from), simpler writing is more likely to be selected and carefully read for further reading than complex writing. Across nearly 30,000 field experiments conducted with the news sites *The Washington Post* (study set 1) and *Upworthy* (study set 2), we find that readers are more likely to select simply written news headlines relative to complexly written news headlines. Study 3 uses a signal detection task (SDT) (*19*) to provide evidence that general news readers (e.g., people from the general public) more closely read simpler headlines when presented with a set of headlines of varied complexity. In addition to its theoretical implications, the finding that readers engage less deeply with complex writing has important practical implications. Specifically, writing simply can help news creators increase audience engagement even for stories that are themselves complicated.

To this end, in the final study, we test whether professional journalists read news headlines differently from the average, or general public news reader. In other domains, such as law, both professionals and nonprofessionals report disliking complex writing (e.g., "legalese") (*11*). Given that journalists produce both headlines and the news stories they connect to, understanding whether journalists exhibit similar reading patterns as their readers is of theoretical and practical importance. Thus, in study 4, professional journalists completed the same survey experiment as the general population sample in study 3. Crucially, we found that these professionals did not use the simpler-writing heuristic when reading headlines; they did not select the simpler headlines for further reading or read them more carefully. Apparently, those who write the news read it differently from those who merely consume it. As observed in many other areas, expertise may undermine effective perspective-taking (*20*). This suggests that those who produce high-quality news may not be well suited to effectively present it in competitive online news environments to general audiences. Although bad actors can also use heuristic-based strategies to vie for consumer attention (*21*), the normative goal of this work, which we adopt as well, is to examine whether credible news can benefit from these (verbal) strategies.

This research is supported by the simpler-is-better hypothesis (*14*), which suggests writing that requires less effort to read will tend to be approached, liked (*22*, *23*), and engaged with (*24*, *25*). Therefore, in accordance with this hypothesis, we preregistered the prediction that in study sets 1 and 2, simpler headlines will receive more clicks than more complex ones. We offer a heuristic-based explanation for

[1]School of Communication, Ohio State University, Columbus, OH 43210, USA. [2]Department of Communication, Michigan State University, East Lansing, MI 48824, USA. [3]John F. Kennedy School of Government, Harvard University, Cambridge, MA 02138, USA.
*Corresponding author. Email: shulman.36@osu.edu.
†These authors contributed equally to this work.

this, showing that attention is directed toward simple writing and away from complex writing. We test this explanation in studies 3 and 4. We hypothesize that recognition memory will be better for headlines written more simply than for headlines written more complexly. These studies are described below.

STUDY SET 1: RESULTS

We partnered with *The Washington Post* to obtain all headline experiments run between 3 March 2021, and 18 December 2022 ($N = 8972$ experiments and $N = 24{,}044$ headlines). From their data and analytics team, we received headline texts, engagement metrics from Chartbeat [e.g., click-through rate (CTR)], and metadata such as the status of the test (e.g., if a winner was found), the author of the headline, and the length of the A/B test. At no point in this project did the authors have access to, nor did we examine, user-level data. Instead, we received headline-level data. Our primary dependent variable was click-through rate (CTR), or "the percentage of visitors who click on a given trial headline" (*26*), which was predicted by the language patterns of each headline. Such internal data from *The Washington Post* was provided to the research team upon reaching a data use agreement, and these tests preceded the authors' involvement on this project. Our sample size was reduced to 7371 experiments ($n = 19{,}926$ headlines) after canceled tests by *The Washington Post* were excluded from the dataset. This study was preregistered (https://aspredicted.org/blind.php?x=253_PNN).

We used Linguistic Inquiry and Word Count (LIWC) to measure several linguistic properties of the headlines (*27*). LIWC identifies how often each word appears in its internal dictionary of social (e.g., family words), psychological (e.g., emotion terms), and part of speech categories (e.g., prepositions) as a percentage of the total word count. LIWC is a widely used text analysis program for dictionary-based evaluations of language and has been central to many psychology of language studies (*14*, *28*, *29*). The primary independent variable used to assess headline simplicity was a simplicity index we developed. This index was composed of four commonly used markers of linguistic simplicity, including common words, readability, analytic writing, and character count. Our creation of this index, along with details regarding the analysis plan, can be found in the Supplementary Materials. Sample headlines from three experiments, with simplicity index scores and CTRs, are presented in Fig. 1.

The data were evaluated in two ways. First, we took advantage of the A/B test design (described in Materials and Method below) from *The Washington Post* by extracting headlines from within each test that scored highest and lowest on the simplicity index. We also extracted such respective headlines' CTR. Within each test, we then created difference scores by subtracting the lowest simplicity score from the highest simplicity score, and the associated lowest CTR

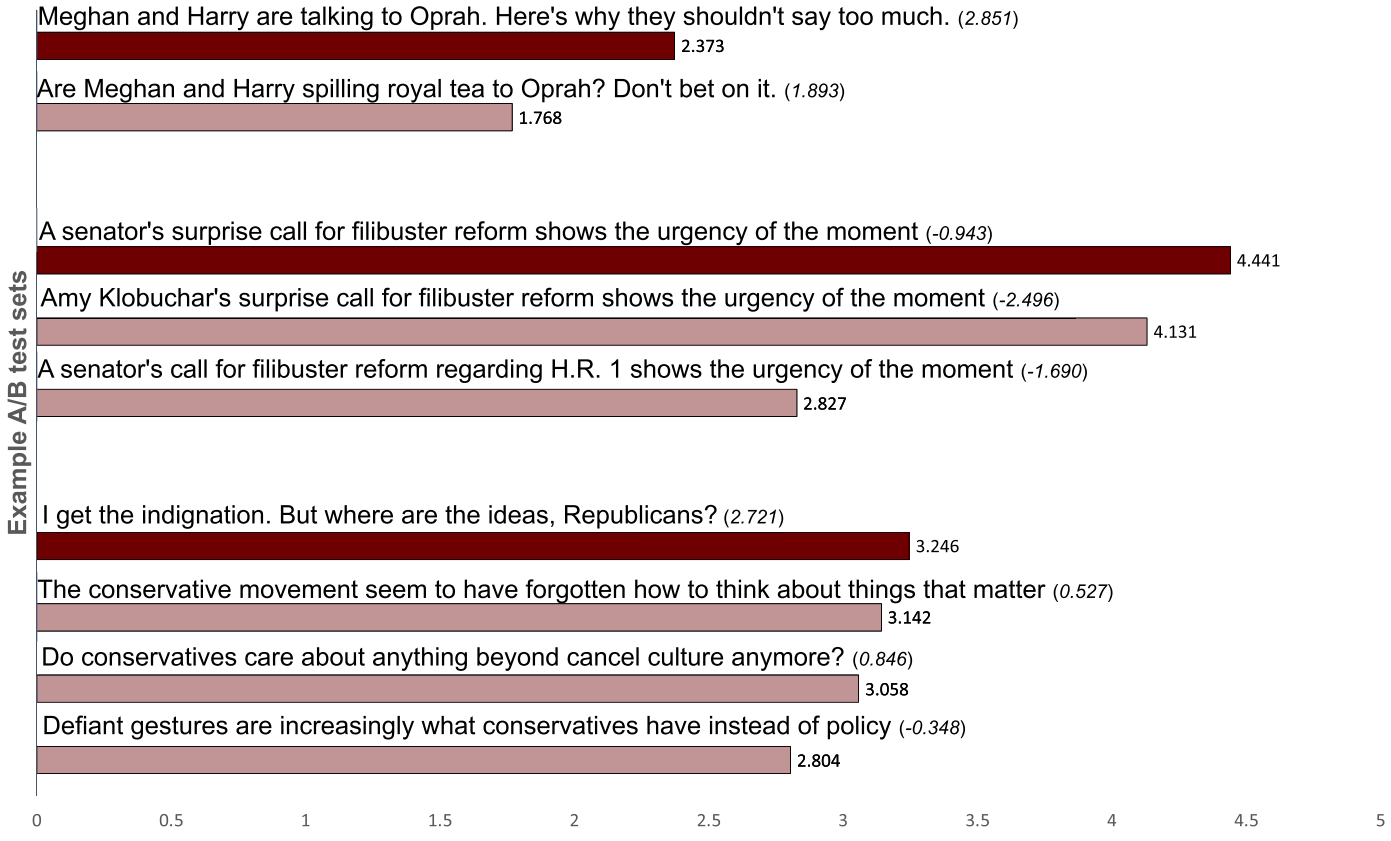


Fig. 1. Sample A/B tests and *CTR* from *The Washington Post*. These headline sets were selected to illustrate the range of headlines generated for a given story, and the direction of the simpler-writing heuristic hypothesis. Numbers in italics are scores on the simplicity index with higher scores indicating more simplicity. The dark red bar reflects the simplest version of the headline in a set whereas the light red bars reflect the more complex versions. Bars are presented in order of CTR within each example set.

(e.g., the CTR from the headline with the lowest simplicity score) from the highest CTR (e.g., the CTR from the headline with the highest simplicity score). We performed a simple bivariate correlation on these difference scores, which were natural log transformed due to skewness concerns (see Supplementary Text for more information). The second approach used a linear mixed model to evaluate the link between simplicity and CTR, controlling for within-test dependencies with a random intercept and other fixed effects described in the Supplementary Materials (e.g., the duration of the A/B test).

Consistent with our hypothesis, the difference in the simplicity index was positively associated with the difference in CTR [$r(7217) = 0.055$ and $P < 0.001$; bootstrapped 95% confidence interval (CI) with 5000 replicates (0.049 to 0.095); see fig. S1]. Using the linear mixed model approach, the simplicity index was also positively associated with CTR ($B = 0.008$, SE = 0.001, $t = 8.84$, $P < 0.001$, $R^2m = 0.092$, and $R^2c = 0.959$). Specifically, more common words, a simpler linguistic style, and more readable texts were associated with a higher CTR, although character count was not (see table S1). Together, we found evidence that simple writing is clicked on more than complex writing. As the evidence in Supplementary Text shows, these patterns were robust to content effects as well. Although these effect sizes are small in absolute magnitude, the size of the readership at *The Washington Post* is on the order of tens of millions (*30*). Thus, even a 1% difference could equate to tens of thousands of additional reads (see Discussion for additional commentary on effect sizes). Moreover, these effect sizes are consistent with other studies that have evaluated the impact of language effects on behavior in the wild (*31*).

The results of this first study set suggest that people engage with and click on linguistically simple headlines more than linguistically complex headlines. With these findings in place, we next explore whether these patterns replicate in other types of online writing by examining a similar set of experiments from a storytelling site that focuses on uplifting content, *Upworthy*.

STUDY SET 1: MATERIALS AND METHODS

Although the research team was not involved in data collection for these experiments, *The Washington Post* provided the team with a general primer about how they approached these experiments. The methodology for these field experiments can be considered an A/B test design. A/B testing refers to a method for comparing two (or more) versions of headline against one another to determine which version performs better [see (*32*) for more information on A/B testing in general]. *The Washington Post* field experiments were collected in collaboration with Chartbeat (*26*). According to this organization's website, Chartbeat randomly exposed users to one of the trial headlines using the Thompson sampling, or Bayesian bandit, algorithm. Headline exposure was then linked to cookies to ensure that users were exposed to the same headline during the life of the experiment. Some of these headline experiments tested CTR differences across two headline versions (approximately 50% of the tests in our sample), whereas some conducted experiments across more than two headline versions. The language on Chartbeat indicates that this approach constitutes a "live experiment" of headline effectiveness. While we echo this "field experiment" language here, we acknowledge that although these tests include some necessary requirements of experimentation, including manipulation on the independent variable, random assignment, and control (the same story was presented after the headline), there was never a true control group at least for our analytic purposes. After exposure to a trial, a CTR was calculated that assessed the proportion of clicks on a given headline relative to the number of users exposed to that headline version (the percentages provided to the research team). As the CTR for a particular headline began to conclusively favor a headline (with 95% confidence), this headline was determined the winner, the test would complete, and the winning headline would be presented 100% of the time. Notably, for some experiments, a winner was never conclusively determined. In these instances, after 20 min, the test would end and the headline variant with the highest CTR would be chosen (*30*).

Our assessment of linguistic simplicity/complexity considered four main variables of interest: (i) common words, (ii) analytic writing, (iii) readability, and (iv) character count. The rate of common words was measured with LIWC via the dictionary category, which considers the degree to which people use simple, everyday terms (*14*, *29*, *33*, *34*). Analytic writing is a measure of linguistic style composed of seven verbal categories (*35*). Texts that score high on analytic writing tend to be more formal and complex than texts that score low on analytic writing. Readability is a measure of structural complexity and accounts for the number of words per sentence and syllables per word. Using the Flesch Reading Ease metric (*36*, *37*), evidence suggests that texts with more words per sentence and syllables per word are more complex and less readable than text with fewer words per sentence and syllables per word. High scores on the Flesch Reading Ease metric are linguistically simpler (more readable) than low scores. We used the quanteda.textstats package in R to calculate readability (*37*). Last, character count is the raw frequency of characters per headline (including spaces). We evaluated character count instead of word count because word count is a basic component of readability, and this measure would therefore be tautological. All descriptive statistics (see table S2 and fig. S2) and correlations (see table S3) between these variables are provided.

STUDY SET 2: RESULTS

Upworthy data were obtained from prior research (*38*) and consisted of 22,664 unique experiments and 105,551 unique headlines from January 2013 to April 2015. Thus, as with the data from study set 1, the authors were not involved in the creation of the A/B tests and no user-level data were assessed. We used the same text analysis approaches and measures as study set 1, although our dependent variable was slightly different because of engagement metric availability. *Upworthy* provided two engagement metrics, impressions, and clicks. We created a click rate by dividing clicks by impressions (clicks per impression, or CPI), which is conceptually similar to the CTR measure from study set 1. Like in study set 1, we modeled the data by (i) correlating simplicity and CPI difference scores within each A/B test and (ii) as a linear mixed model, controlling for A/B test as a random intercept. This study was preregistered (https://aspredicted.org/blind.php?x=RQH_LML).

The simplicity index was positively associated with CPI [$r(22{,}662) = 0.022$ and $P < 0.001$; bootstrapped 95% CI with 5,000 replicates (0.021 to 0.054); see fig. S1]. In the linear mixed model, headlines with simpler language received more CPI than headlines with less simple language ($B = 0.002$, SE = 0.001, $t = 2.45$, $P = 0.014$, $R^2m = 0.00003$, and $R^2c = 0.830$). We replicated the common words, analytic writing, and readability effects from study set 1 (table S4), although texts with more characters had higher CPI. Consistent

with study set 1, the mixed model results were robust to content effects as well (see the Supplementary Materials for details).

Together, study sets 1 and 2 provided field-based evidence for the simpler-writing heuristic. Next, given the importance of attention in these spaces, we experimentally tested an explanation for the selection effects observed in study sets 1 and 2, namely, that people select headlines based on attention allocation.

STUDY SET 2: MATERIALS AND METHODS

The *Upworthy* Research Archive, and the methods therein, are discussed at length with the dataset (*38*) that introduced this resource. This archive reports the findings from multiple experiments that broadly examined how various features of news stories (headlines, images, previews, and content) affect a variety of outcomes, including selection. Given that these data were intended for academic use, the method, results, and rigor of the experimental methodology are articulated in the archive. Germane to the current investigation, the unit of analysis for each experiment was a web browsing session. Within this session, users were randomly assigned to receive a particular headline version using the RandomSample method. The outcome was whether participants selected the headline for reading. Editors determined when to conclude a test based on a host of custom calculations that determined the "significance" and relative success of each version of the test. Thus, although details may differ between the data collection or algorithms guiding headline exposure and test administration between study sets 1 and 2, they both sought answers to similar questions regarding which headline version was selected.

Given this similarity, our logic was that if these sites produced similar relationships, these findings support the generalizability of the claims made here. Thus, the goal with study set 2 was not pure replication but rather, generalizability across two different types of news websites that vary in interesting ways. For instance, study set 1 explored headline preferences using data from a legacy news outlet that has a national and international readership, prestige, and influence because millions of people visit this news site every month (*30*). The second study set, while also focused on news reading in the wild, leveraged the expertise of academics to create a news site that would contribute to the general store of knowledge. Thus, the pair of studies offered the opportunity to assess the generalizability of our claims in ecological, and methodologically sophisticated, ways.

Upworthy data (the confirmatory package) were obtained from prior research (*37*) and consisted of 22,664 experiments and 105,551 headlines from January 2013 to April 2015. We used the same text analysis approaches and measures as study set 1, although our dependent variable was slightly different because of engagement metric availability. *Upworthy* provided two engagement metrics, impressions and clicks. We created a click rate by dividing clicks by impressions (CPI), which is conceptually similar to the CTR measure from study set 1. CPI values were re-expressed using the formula $\ln(Y + 0.001)$ out of skewness concerns. As with study set 1, we provide all descriptive (table S2 and fig. S3) and correlational information (table S3) about these variables.

STUDY 3: RESULTS

This survey experiment had two aims. The first was to replicate in a more controlled environment whether a simple version of a headline chosen from *The Washington Post* received more clicks than a complex version of the same headline. Thus, participants read 10 headlines and indicated which headline they would be likely to select if they were reading the news. Information about the construction of these headlines is provided in the Supplementary Materials (table S5). The second aim was to understand the underlying cognitive process driving people to select simpler headlines. For this, we used a 24-item SDT paradigm (*19*), which is designed to assess recognition memory, or here, attention. The SDT was designed to assess whether people allocated more attention to simple texts relative to complex texts using a measure of sensitivity (d'). The higher the sensitivity score, the more attention and retention paid to the headline. This study was preregistered (https://aspredicted.org/7NV_PZR) and an a priori power analysis ensured that we had enough participants to detect a small effect at 95% power. Data, syntax, and output can be found on our Open Science Framework (https://osf.io/nwsgf/?view_only=55f56aec468c4d8699787b416b7afdee) page.

The results suggested that participants were significantly more likely to select simple headlines compared to more complex ones [$\chi^2(1, N = 524) = 32.25$, $P < 0.001$, and odds ratio = 2.83]. Specifically, when target headlines were written simply, they were selected more ($n = 177$, 34.8%) than the control headlines ($n = 78$, 15.3%). Alternatively, when the target headlines were written using complex language, they were selected less ($n = 113$, 22.2%) than the control headlines ($n = 141$, 27.7%; see Fig. 2).

The SDT outcome was also consistent with our hypothesis, as participants in the simple headline condition ($M = 1.23$ and SD = 0.81) demonstrated significantly better sensitivity compared to those in the complex headline condition [$M = 0.80$ and SD = 0.77; $t(483) = 6.01$, $P < 0.001$, and Cohen's $d = 0.55$]. This means that complex headlines were less likely to be selected, and that minutes later, the phrases in complex headlines were less likely to be recognized compared to simple headlines.

STUDY 3: MATERIALS AND METHODS

Participants in this survey experiment ($N = 524$) were recruited from Amazon Mechanical Turk via CloudResearch (*39*) from 2 to 8 May 2023. After being provided with screening measures to ensure that participants could pass a CAPTCHA and were at least 18 years old, participants were presented with an approved consent form (no. 2023E0422) from the lead author's institution. This sample identified as 53.8% male, 45.2% female, and 0.6% nonbinary, with 1% of participants' data either missing ($n = 3$) or preferred not to say ($n = 2$). The average age of the sample was 41.70 years old (SD = 12.03 and range = 20 to 77) and 76.4% of participants identified as white, 10.8% Black, 6.8% Asian, 1% American Indian or Alaska Native, and 2.8% other or multiracial. To obtain higher-quality data, participants were eligible to participate if they could pass a CAPTCHA, could respond to an open-ended prompt, and had at least a 95% completion rating on at least 500 human intelligence tasks. Participants were compensated $2.00 for their time.

To create the SDT measure, participants were presented with a three-word phrase and asked whether this phrase appeared (coded as 1) or did not appear (coded as 0) in the set of headlines they viewed. A "hit" was an instance in which participants accurately reported that they saw a phrase (true positive) or accurately reported that they did not see a phrase (true negative). Alternatively, "foils" were instances where participants incorrectly reported that they saw a phrase when they did not (false positive) or alternatively, when they stated that they did not see a phrase when they did (false negative).

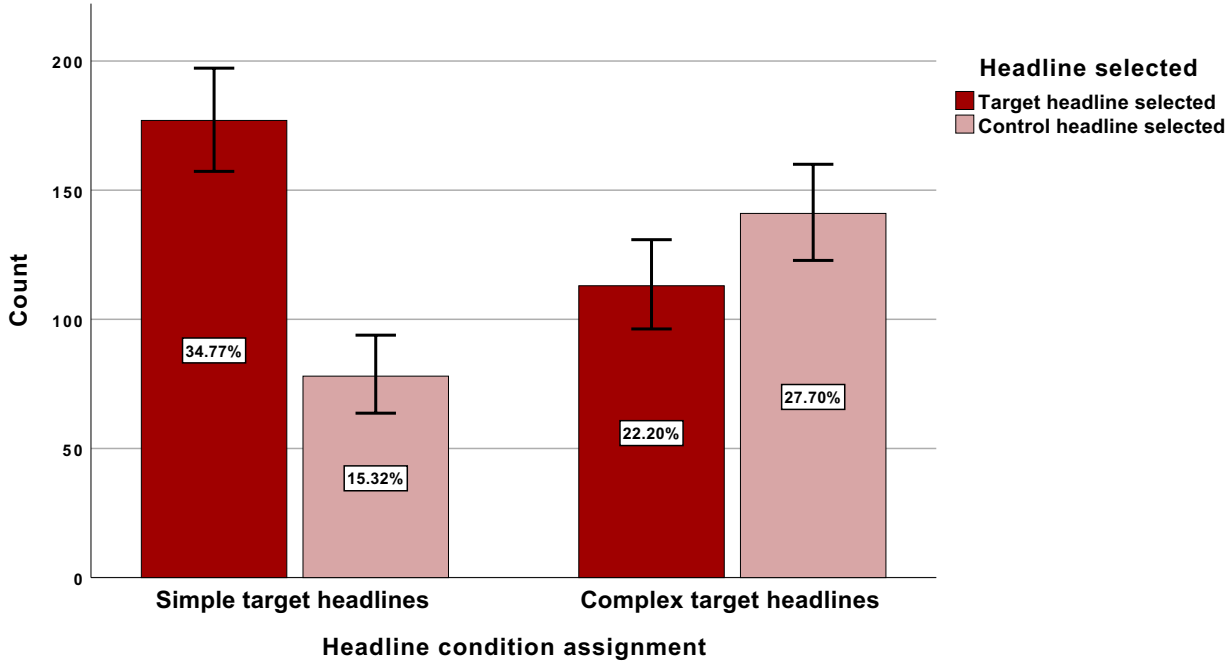


Fig. 2. Headline selections based on experimental condition. Error bars reflect 95% CIs surrounding the estimates presented. These estimates reflect the percentage of participants who selected a particular headline type.

STUDY 4: RESULTS

Study 4 included the same headline selection and SDT as study 3. The only difference was that this sample consisted of professional journalists and writers. Because this sample was different and the recruitment efforts were different, a new institutional review board protocol was used and approved (no. 2023E0883) by the lead author's institution. This study was preregistered (https://aspredicted.org/DQQ_Y6T) using the same hypotheses as study 3 and an a priori power analysis based on the study 3 effect size ensured enough participants were recruited.

The results from study 4 produced a notable departure from studies 1 to 3. First, the results of the headline selection task were not significant [$\chi^2(1, N = 225) = 0.36$ and $P = 0.549$]. Second, the results of the SDT were also not significant [$t(165) = -0.44$, $P = 0.660$, and Cohen's $d = 0.07$]. Together, these null findings suggest that, for journalists, headline simplicity does not affect selection, attention, or memory. One notable finding from the SDT was just how well journalists performed. This value of sensitivity ($M = 1.43$ and SD = 0.91) was significantly higher than the sensitivity observed in the general population sample [$M = 1.02$ and SD = 0.82; $t(523) = 3.01$, $P < 0.01$, and Cohen's $d = 0.47$], which suggests journalists appear to vigilantly read and remember what they read (see Fig. 3).

One final piece of evidence provides support for the idea that journalists read differently from the general population sample from study 3. Journalists were presented with six headline pairs from *The Washington Post* and were asked if they could correctly identify the winning headline. Here, journalists performed no better than chance [50% accuracy, $M = 3.09$, SD = 1.36, $t(146) = 0.79$, $P = 0.432$, and $d = 0.07$], suggesting a disconnect between what journalists think audiences will read and what they actually do. We consider this disconnect, along with the methodological, theoretical, and practical implications of these reading habits in General Discussion.

STUDY 4: MATERIALS AND METHODS

Participants in this survey experiment ($N = 249$) were recruited from a list of participants enrolled in a webinar on effective writing. This webinar included a presentation by one of the authors of this article. Before the webinar, attendees were asked to participate in a brief survey relevant to the presentation. They were told that the results of this survey would be of interest to them and would be shared during the presentation. The attendees of this webinar were a particularly interesting sample because they all identified as professional writers, including mostly current and former journalists (average of 13.86 years of experience, SD = 14.60, and $n = 122$, 47%), but also educators, communication directors, and government employees (see table S6 for occupational demographics). This sample identified as 32.5% male, 63.9% female, and 0.8% nonbinary, with 2.8% of participants' data either missing ($n = 4$) or preferred not to say ($n = 3$). The average age of the sample was 52.38 years old (SD = 15.38 and range = 18 to 93) and 73.1% of participants identified as white, 5.6% Black, 8.4% Asian, 0.8% American Indian, Alaska Native, or Pacific Islander, and 8.4% other or multiracial (2.8%).

GENERAL DISCUSSION

Thousands of field experiments across traditional (i.e., *The Washington Post*) and nontraditional news sites (i.e., *Upworthy*) showed that news readers are more likely to click on and engage with simple headlines than complex ones. General readers were also more likely

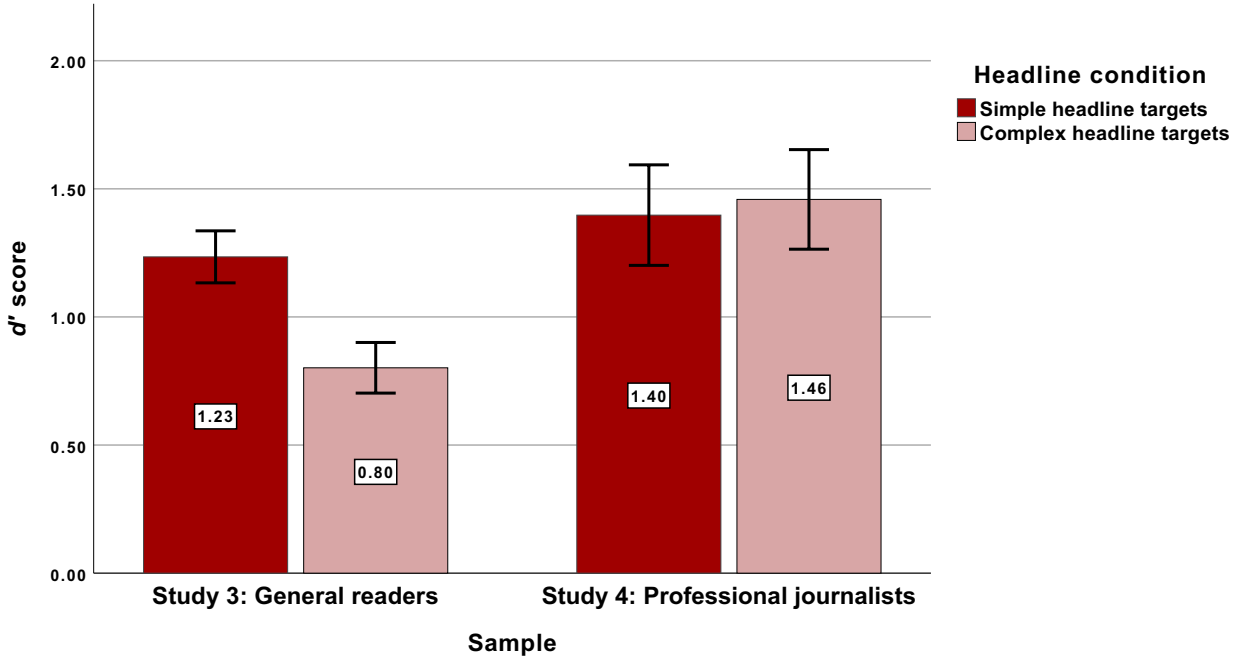


Fig. 3. Differences in task performance between studies 3 and 4. Error bars reflect 95% CIs surrounding the estimates. These sample estimates reflect data from crowdsourced workers in study 3 and the sample of professional writers in study 4. For more detailed occupational demographics for study 4, see the Supplementary Materials.

to recognize phrases from the simpler headlines than from complex headlines. These results are consistent with our theory that in crowded information environments, people are guided by a simpler-writing heuristic: People use the simplicity/complexity of the writing they encounter as a cue for what writing they will engage with and attend to.

There are several important takeaways from this package of studies. First, the consistencies across the first three studies suggest that, overwhelmingly, general readers are economical with their attention and that the simpler-writing heuristic provides a useful explanation for how people decide what to read online. Practically, this finding implies that small-scale efforts aimed at increasing the simplicity or fluency of language can increase the attention of casual readers. Second, the findings observed from the journalists' sample suggests that journalists may exhibit a different, and more thorough, approach to news reading. This thoroughness was evident in their news selection (no preference for the simpler titles) and in their high level of recognition memory across headlines. Notably, this observation presents a departure from other research that has found in professions like law, lawyers, and nonlawyers alike report a distaste for legalese (*11*).

There are a few potential explanations for the disconnect between the general readers' sample and the professional writer sample that merit further exploration. The first is methodological. It is possible that these two study samples approached the task differently. Specifically, journalists might have felt motivated to perform better because their performance reflected on their professional identity. General news readers, by contrast, might have approached the task more casually and as a result, underperformed relative to journalists. Although prior research has similarly found that motivation reduces the impact of heuristics, such as language complexity on topic engagement (*40*–*42*), this explanation does not quite square with all of the evidence obtained in our work. For instance, in the follow-up A/B test, where journalists' motivation should be quite high ("Can you guess the winning headline?"), journalists performed no better than chance at guessing *The Washington Post* headline that received the most clicks. Thus, when journalists were directly asked to perspective take about consumers' preferences, they were unable to do so accurately. Second, although demand characteristics may have been high for the journalists' sample, the benefit of using signal detection to assess attention allocation is that it uses a behavioral measure that, unlike self-report, is less susceptible to demand characteristics. Thus, even though there are some methodological differences between studies 3 and 4 due to sample recruitment, it is hard to imagine how these differences explain the entirety of the effects that we observed.

The different reading approaches of those who create the news and those who consume it may lead to consequential blind spots. The possibility that journalists are more motivated to carefully read and process the news, relative to general news readers, may suggest a disconnect between what journalists think audiences want to read, and what audiences actually read. Future work investigating this potential disconnect is important theoretically, to illuminate boundary conditions for the simpler-writing heuristic and, practically, to help news organizations understand where they can improve.

When looking at the findings across study sets 1 and 2, we note that the effect sizes obtained for simple writing are consistent with other language-based field experiments in the psychology of language area. For example, a paper by Kramer *et al.* (*31*) found that modifying rates of emotion in one's Facebook newsfeed changed rates of emotion in their subsequent posts, with a very small but systematic effect size (Cohen's $d = 0.02$, equivalent to $r = 0.01$). On the basis of Facebook's population size, language-based effects even at this magnitude can still lead to a nontrivial downstream behavioral impact at scale. The effect sizes that we observe are consistent

with prior work, and the scale of our news sources similarly suggests real-world consequences. To illustrate, on the basis of audience traffic data from *The Washington Post* (*30*), there was an average of about 70 million digital unique visitors to the site per month during the time period in question (March 2021 to December 2022). If we assume that each visitor reads three stories, small percentage differences at this scale matter greatly: A CTR difference of 0.10% (2.1% versus 2.0%) still equates to a difference of over 200,000 readers of the simpler stories based on headline simplicity alone.

Together, this work highlights the benefits of language simplicity as one of many elements that can increase demand for and attention to credible news. While many features can affect attention and selection of news headlines (*43*), one benefit of linguistic simplicity is its ease of implementation, even for otherwise complex stories. In online spaces where less credible (*21*) and highly polarized sources (*4*) already tend to use simpler writing, we suggest that the simpler-writing heuristic can increase demand for credible journalism in a competitive attention economy.

Supplementary Materials

This PDF file includes:

Supplementary Text
Figs. S1 to S3
Tables S1 to S8
References

REFERENCES AND NOTES

1. S. Vosoughi, D. Roy, S. Aral, The spread of true and false news online. *Science* **359**, 1146–1151 (2018).
2. J. Allen, B. Howland, M. Mobius, D. Rothschild, D. J. Watts, Evaluating the fake news problem at the scale of the information ecosystem. *Sci. Adv.* **6**, eaay3539 (2020).
3. V. Bélair-Gagnon, L. Graves, B. Kalsnes, S. Steensen, O. Westlund, Considering interinstitutional visibilities in combating misinformation. *Digit. Journal.* **10**, 669–678 (2022).
4. J. F. Sparks, J. D. Hmielowski, At the extremes: Assessing readability, grade level, sentiment, and tone in US media outlets. *Journal. Stud.* **24**, 24–44 (2023).
5. A. M. Guess, P. Barberá, B. Barberá, S. Munzert, J. Yang, The consequences of online partisan media. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2013464118 (2021).
6. M. Carlson, The information politics of journalism in a post-truth age. *Journal. Stud.* **19**, 1879–1888 (2018).
7. S. T. Fiske, S. E. Taylor, *Social Cognition* (Mcgraw-Hill Company, ed. 2nd, 1991).
8. P. Pirolli, S. Card, Information foraging. *Psychol. Rev.* **106**, 643–675 (1999).
9. T. Rogers, J. Lasky-Fink, *Writing for busy readers: Communicate more effectively in the real world* (Penguin, 2023).
10. T. P. Vos, R. J. Thomas, "They're making it more democratic": The normative construction of participatory journalism. *Digit. Journal.* **1-25**, 1–25 (2023).
11. E. Martínez, F. I. Mollica, E. Gibson, Even lawyers do not like legalese. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2302672120 (2023).
12. H. C. Shulman, G. N. Dixon, O. M. Bullock, D. Colón Amill, The effects of jargon on processing fluency, self-perceptions, and scientific engagement. *J. Lang. Soc. Psychol.* **39**, 579–597 (2020).
13. P. Tolochko, H. Song, H. Boomgaarden, "That looks hard!": Effects of objective and perceived textual complexity on factual and structural political knowledge. *Polit. Commun.* **36**, 609–628 (2019).
14. D. M. Markowitz, H. C. Shulman, The predictive utility of word familiarity for online engagements and funding. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2026045118 (2021).
15. C. E. Robertson, N. Pröllochs, K. Schwarzenegger, P. Pärnamets, J. J. Van Bavel, S. Feuerriegel, Negativity drives online news consumption. *Nat. Hum. Behav.* **7**, 812–822 (2023).
16. A. Fishbane, A. Ouss, A. K. Shah, Behavioral nudges reduce failure to appear for court. *Science* **370**, eabb6591 (2020).
17. J. R. Kling, S. Mullainathan, E. Shafir, L. C. Vermeulen, M. V. Wrobel, Comparison friction: Experimental evidence from Medicare drug plans. *Q. J. Econ.* **127**, 199–235 (2012).
18. E. P. Bettinger, B. T. Long, P. Oreopoulos, L. Sanbonmatsu, The role of application assistance and information in college decisions: Results from the H&R Block Fafsa experiment. *Q. J. Econ.* **127**, 1205–1242 (2012).
19. H. Stanislaw, N. Todorov, Calculation of signal detection theory measures. *Behav. Res. Methods. Instrum. Comput.* **31**, 137–149 (1999).
20. P. J. Hinds, M. Patterson, J. Pfeffer, Bothered by abstraction: The effect of expertise on knowledge transfer and subsequent novice performance. *J. Appl. Psychol.* **86**, 1232–1243 (2001).
21. P. Biyani, K. Tsioutsiouliklis, J. Blackmer, "'8 Amazing Secrets for Getting More Clicks': Detecting Clickbaits in News Streams Using Article Informality"; http://news.yahoo.com/justin-bieber-makes-huge-life-.
22. D. M. Oppenheimer, Consequences of erudite vernacular utilized irrespective of necessity: Problems with using long words needlessly. *Applied Cognitive Psychology* **20**, 139–156 (2006).
23. H. C. Shulman, M. D. Sweitzer, Advancing framing theory: Designing an equivalency frame to improve political information processing. *Hum. Commun. Res.* **44**, 155–175 (2018).
24. M. D. Sweitzer, H. C. Shulman, The effects of metacognition in survey research: Experimental, cross-sectional, and content-analytic evidence. *Public Opin. Q.* **82**, 745–768 (2018).
25. H. C. Shulman, M. D. Sweitzer, Varying metacognition through public opinion questions: How language can affect political engagement. *J. Lang. Soc. Psychol.* **37**, 224–237 (2018).
26. Guide to headline testing and image testing, *Chartbeat* (2023).
27. J. W. Pennebaker, R. L. Boyd, R. J. Booth, A. Ashokkumar, M. E. Francis, *Linguistic inquiry and word count: LIWC-22*. Pennebaker Conglomerates (2022).
28. R. L. Boyd, H. A. Schwartz, Natural language analysis and the psychology of verbal behavior: The past, present, and future states of the field. *J. Lang. Soc. Psychol.* **40**, 21–41 (2021).
29. D. M. Markowitz, Instrumental goal activation increases online petition support across languages. *J. Pers. Soc. Psychol.* **124**, 1133–1145 (2023).
30. WashPostPR, The Washington Post records more than 71 million unique digital visitors in December 2021, *The Washington Post* (2022). https://www.washingtonpost.com/pr/2022/01/25/washington-post-records-more-than-71-million-unique-digital-visitors-december-2021/.
31. A. D. I. Kramer, J. E. Guillory, J. T. Hancock, Experimental evidence of massive-scale emotional contagion through social networks. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 8788–8790 (2014).
32. A. Gallo, A refresher on A/B testing, *Harvard Business Review*, 28 June 2017.
33. D. M. Markowitz, J. T. Hancock, Linguistic obfuscation in fraudulent science. *J. Lang. Soc. Psychol.* **35**, 435–445 (2016).
34. Y. R. Tausczik, J. W. Pennebaker, *The psychological meaning of words: LIWC and computerized text analysis methods*, *J. Lang. Soc. Psychol.* **29**, 24–54 (2010).
35. J. W. Pennebaker, C. K. Chung, J. Frazee, G. M. Lavergne, D. I. Beaver, When small words foretell academic success: The case of college admissions essays. *PLOS ONE* **9**, e115844 (2014).
36. R. Flesch, A new readability yardstick. *J. Appl. Psychol.* **32**, 221–233 (1948).
37. K. Benoit, K. Watanabe, H. Wang, P. Nulty, A. Obeng, S. Müller, A. Matsuo, quanteda: An R package for the quantitative analysis of textual data. *J. Open Source Softw.* **3**, 774 (2018).
38. J. N. Matias, K. Munger, M. A. Le Quere, C. Ebersole, The Upworthy Research Archive, a time series of 32,487 experiments in U.S. media. *Sci Data* **8**, 195 (2021).
39. L. Litman, J. Robinson, T. Abberbock, TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behav. Res. Methods* **49**, 433–442 (2017).
40. H. C. Shulman, O. M. Bullock, Don't dumb it down: The effects of jargon in COVID-19 crisis communication. *PLOS ONE* **15**, e0239524 (2020).
41. H. C. Shulman, O. M. Bullock, E. E. Riggs, The interplay of jargon, motivation, and fatigue while processing COVID-19 crisis communication over time. *J. Lang. Soc. Psychol.* **40**, 546–573 (2021).
42. R. E. Petty, J. T. Cacioppo, "The elaboration likelihood model of persuasion" in *Advances in Experimental Social Psychology*, L. Berkowitz, Ed. (Academic Press, New York, 1986; 19, pp. 123–205).
43. K. Gligorić, G. Lifchits, R. West, A. Anderson, Linguistic effects on news headline success: Evidence from thousands of online field experiments (Registered Report). *PLOS ONE* **18**, e0281682 (2023).
44. D. Bates, M. Mächler, E. Zurich, B. M. Bolker, S. C. Walker, Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).
45. A. Kuznetsova, P. B. Brockhoff, R. H. B. Christensen, lmerTest package: Tests in linear mixed effects models. *J. Stat. Softw.* **82**, 1–26 (2017).
46. C. K. Chung, J. W. Pennebaker, Revealing dimensions of thinking in open-ended self-descriptions: An automated meaning extraction method for natural language. *J. Res. Pract.* **42**, 96–132 (2008).
47. D. M. Markowitz, The meaning extraction method: An approach to evaluate content patterns from large-scale language data. *Front. Commun.* **6**, (2021).
48. D. M. Blei, A. Y. Ng, M. I. Jordan, "*Latent Dirichlet Allocation*" (2002).
49. D. M. Markowitz, Self-presentation in medicine: How language patterns reflect physician impression management goals and affect perceptions. *Comput. Human Behav.* **143**, 107684 (2023).

50. D. M. Markowitz, M. Kouchaki, J. T. Hancock, F. Gino, The deception spiral: Corporate obfuscation leads to perceptions of immorality and cheating behavior. *J. Lang. Soc. Psychol.* **40**, 277–296 (2021).
51. M. Chmielewski, S. C. Kucker, An MTurk crisis? Shifts in data quality and the impact on study results. *Soc. Psychol. Personal. Sci.* **11**, 464–473 (2020).

Acknowledgments: We thank members of *The Washington Post*, including C. O'Lear and V. Varalu, for collaborating on this project. We also thank J. Lerner, K. Milkman, M. Norton, and A. Shah for the feedback and insights on various drafts of this paper. In Study set 1, the data use agreement was between D.M.M. and *The Washington Post*. At the time of this agreement, D.M.M. was at the University of Oregon, but changed institutions in the middle of this project. The project, and expectations communicated between parties prior to this, did not change despite the institution change. **Funding:** The authors acknowledge that they received no external funding in support of this research. **Author contributions:** Conceptualization: H.C.S., D.M.M., and T.R. Methodology: H.C.S., D.M.M., and T.R. Statistical analysis: D.M.M. and H.C.S. Data acquisition: H.C.S., D.M.M., and T.R. Project administration: H.C.S. Writing—original draft: H.C.S., D.M.M., and T.R. Writing—review and editing: H.C.S., D.M.M., and T.R. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. All data, code, and materials used in the analyses pertinent to studies 2 to 4 can be found on our Open Science Framework page (https://osf.io/nwsgf/?view_only=55f56aec468c4d8699787b416b7afdee). The data from Study set 1 cannot be made available because access to such data were contingent on them not being publicly released. Per the terms of this agreement, requests for the data should contact D.M.M., who will then refer interested parties to *The Washington Post*. The data for studies 2 to 4 are permanently stored on Harvard Dataverse (https://doi.org/10.7910/DVN/T8JTRP).

Submitted 30 November 2023
Accepted 1 May 2024
Published 5 June 2024
10.1126/sciadv.adn2555

Routledge
Taylor & Francis Group

Political Communication

The Role of Framing, Race, and Symbolic Racism in Policy Support

Submission ID	248519864
Article Type	Research Article
Keywords	Framing effects, experimental design, public opinion, race and prejudice, public policy

For any queries please contact:

UPCP-peerreview@journals.tandf.co.uk

Note for Reviewers:

To submit your review please visit https://mc.manuscriptcentral.com/upcp

SUPPLEMENTAL MATERIALS

Table of Contents

SUPPLEMENTAL MATERIALS

Out of Sample Pilot

The goal of this pilot was to assess the potency of the primary manipulations (valence, race, topic racialization) under investigation. We opted for an out of sample pilot test, or manipulation check, conducted after the two studies were already concluded, to obtain insights about the success of our manipulations without the drawbacks of risking participant sensitization regarding the purpose of this research and induce socially desirable responses in the process. As we present below, most of these manipulations were generally successful. That said, our results related to the success of the race manipulation – while in the correct direction – in some ways underscore the challenges with directly asking people about racial connotations. We present these ideas below.

Participants

Participants (N = 526) were recruited from CloudResearch (IRB# #2023E0658) from July 10, 2023 through July 14, 2023. The same eligibility requirements and compensation ($2.00) were used as the previous two studies to maintain consistency. Thus, we once again created separate HITS for participants identifying as White (n = 267) and participants identifying as Black (n = 253). The average age of this sample was 39.77 years old (SD = 16.25) and was 51.7% female (47.9% male and 0.4% identifying as non-binary). On average, this survey took approximately 8.5 minutes (SD = 9.11).

Procedure

Participants were informed that the purpose of the experiment was to read a ballot measure in support of [BALLOT CONDITION] in a specific US community. Participants were then randomly assigned to one of four ballot conditions from either Study 1 or Study 2. Thus, this was a mixed experimental design that included the following factors: Valence frame (

between subject: positive, negative), Race-cue frame (between subjects: Black, White), ballot topic (repeated measure where order was randomized: police reform first, election reform first), and the quasi-experimental factor of participant race (Black, White). Across both topics, immediately following the ballot manipulation participants were asked how they would vote on this issue, followed by manipulation check items described below.

Measures

Valence manipulation check. In this pilot, four items were used to assess ballot valence ($M = 3.67$, $SD = 0.84$, $\alpha = .94$). An example item includes, "How would you describe the tone of the language used in this proposal?" with response options ranging from (1) very negative to (5) very positive. The full scale is in the complete measures section of this document.

Race cue manipulation check. After exposure to ballot condition, participants were asked a single question that read "Which of the following best describes the racial composition of the community described in this proposal?" Response options included: majority Black community, majority White community, or unsure.

Topic comparison manipulation check. We used this pilot to validate our assumption that police reform is viewed as a more racially explicit, or racialized topic than election reform. Specifically, we asked three questions about how "racialized" participants perceived each political issue (police reform: $\alpha = .64$; election reform: $\alpha = .82$). An example item includes, "[Police/Election] reform is a racially sensitive issue."

Results

With regard to the success of the valence frame manipulation, the results of this pilot found that, for Study 1, as expected those in the positive condition rated the ballots as significantly more positive in tone ($M = 3.78$, $SD = 0.84$, $t\,[501] = 2.86$, $p < .01$, $d = .26$) than

those in the negative condition (M = 3.57, SD = 0.82). Thus, this manipulation was successful. For Study 2, however, the result of an independent samples t-test was not significant, t (502) = 0.17, p = .868, d = .02. Thus, unlike Study 1, the valence frame manipulation was not successful.

For the race-cue frame induction, in Study 1, the results of a Chi-Square analysis were significant, χ^2 (2, N = 504) = 64.42, p < .001. Specifically, participants in the Black frame condition correctly identified the majority race of the community as Black over White 74% of the time. Similarly, those in the White frame condition correctly identified the race as White over Black 70.5% of the time. That said, 50% of respondents overall stated they were unsure, but this option was chosen more in the White frame condition (60.4%). All told, while this induction was weak, people were more correct than incorrect in their perceptions indicating a somewhat successful manipulation, particularly for the Black frame condition. Within Study 2, the results of the pilot test similarly revealed that this manipulation was generally successful, χ^2 (2, N = 505) = 40.48, p < .001. Specifically, participants in the Black cue condition correctly identified the race as Black over White 74% of the time. Similarly, those in the White cue condition correctly identified the race as White over Black 68% of the time. Again, 54% of respondents overall stated they were unsure. Thus, while this induction was weak, people were once again more correct than incorrect.

Finally, we used our out of sample pilot study to validate our assumption that police reform was viewed as a more racially explicit topic than election reform. Specifically, we asked three questions about how "racialized" participants perceived each political issue. A paired-samples t-test revealed that police reform was perceived as a significantly more racially sensitive issue (M = 5.27, SD = 1.14) than election reform (M = 4.92, SD = 1.41), t (495) = 6.70, p < .001, d = .30. Thus, this set of topics was chosen effectively based on this result.

SUPPLEMENTAL MATERIALS

Study 1 Stimuli

	Defund the Police	**Reimagine Public Safety**
White-Caucasian	Should the city of Providence, Rhode Island charter be changed to eliminate the Police Department and replace it with a Department of Public Safety whose purpose would be to use a public health method to address suburban crime and improve community well-being? The Department of Public Safety's specific tasks would be decided together by the Mayor and City Council. Control of the creation, daily work, and leadership of this new department would not be exclusive to the Mayor. The general summary of changes being proposed are explained in the extra note below as a part of this ballot. This change would defund and remove the current Police Department, and its chief, and use these funds to finance a Department of Public Safety which combines public safety tasks through a broad public health method, as decided by the Mayor and Council. This method would include addressing suburban crime, reduction of illicit substances (i.e., opiates), and promoting social support among residents in the community. The department would be led by a Commissioner recommended by the Mayor and approved by the Council. **Word count: 179**	Should the city of Providence, Rhode Island charter be changed to remove the Police Department and replace it with a Department of Public Safety whose purpose would be to use a public health method to address suburban crime and improve community well-being? The Department of Public Safety's specific tasks would be decided together by the Mayor and City Council. Control of the creation, daily work, and leadership of this new department would not be exclusive to the Mayor. The general summary of changes being proposed are explained in the extra note below as a part of this ballot. This change would reallocate funds previously used for the Police Department, and chief, and use these funds to support a Department of Public Safety which combines public safety tasks through a broad public health method, as decided by the Mayor and Council. This method would include addressing suburban crime, reduction of illicit substances (i.e., opiates), and promoting social support among residents in the community. The department would be led by a Commissioner recommended by the Mayor and approved by the Council. **Word count: 179**
Black/African-American	Should the city of Detroit, Michigan charter be changed to eliminate the Police Department and replace it with a Department of Public Safety whose purpose would be to use a public health method to address urban crime and improve inner city welfare? The Department of Public Safety's specific tasks would be decided together by the Mayor and City Council. Control of the creation, daily work, and leadership of this new department would not be exclusive to the Mayor. The general summary of changes being proposed are explained in the extra note below as a part of this ballot.	Should the city of Detroit, Michigan charter be changed to remove the Police Department and replace it with a Department of Public Safety whose purpose would be to use a public health method to address urban crime and improve inner city welfare? The Department of Public Safety's specific tasks would be decided together by the Mayor and City Council. Control of the creation, daily work, and leadership of this new department would not be exclusive to the Mayor. The general summary of changes being proposed are explained in the extra note below as a part of this ballot.

SUPPLEMENTAL MATERIALS

	This change would defund and remove the Police Department, and its chief, and use these funds to finance a Department of Public Safety which combines public safety tasks through a broad public health method, as decided by the Mayor and Council. This method would include addressing urban crime, reduction of illicit substances (i.e., crack-cocaine), and promoting social support among people in the inner city. The department would be led by a Commissioner recommended by the Mayor and approved by the Council. **Word count: 179**	This change would reallocate funds previously used for the Police Department, and chief, and use these funds to support a Department of Public Safety which combines public safety tasks through a broad public health method decided by the Mayor and Council. This method would include addressing urban crime, reduction of illicit substances (i.e., crack-cocaine), and promoting social support among people in the inner city. The department would be led by a Commissioner recommended by the Mayor and approved by the Council. **Word count: 179**

SUPPLEMENTAL MATERIALS

Study 2 Stimuli

	Promote Election Safety Initiative	Prevent Election Fraud Initiative
White-Caucasian	North Dakotans need to have more confidence in the integrity of our elections to help ensure that registration rolls are adequately maintained, officials have distributed the correct ballots, and that current policies promote proper verification of an individual's identity and eligibility to vote in an election. The general summary of the changes being proposed are explained in the extra note below as a part of this ballot. To ensure that 'true North Dakotans' are able to vote in this election, only persons who can legally show they have resided in our State of North Dakota for one year and in the election district for 30 days preceding an election shall be allowed to vote. Additionally, polling places and ballot boxes in our rural farm communities will be observed to ensure that votes are properly counted in these rural communities. Voters who choose to vote in person must present government approved photographic identification before voting. For voting-by-mail, the individual must provide a valid government-issued identification card number and their signature must perfectly match the signature on-file with their voter registration. These changes will promote election safety and will restore confidence in election integrity. **Word count: 192 (3+ words are due to State name)**	North Dakotans have expressed a concern with fraud in our elections because our registration rolls were not adequately maintained. This occurred because officials distributed a significant number of ballots in error, and current policies did not require proper verification of an individual's identity and eligibility to vote in an elections. The general summary of the changes being proposed are explained in the extra note below as a part of this ballot. To ensure 'fake North Dakotans' do not vote in this election, persons who cannot legally show that they have resided in our State of North Dakota for one year and in the election district for 30 days preceding an election shall be prohibited from voting. Additionally, polling places and ballot boxes in our rural farm communities will be monitored to avoid votes getting miscounted in these rural communities. Voters who choose to vote in person must present government approved photographic identification before voting. For voting-by-mail, the individual must provide a valid government-issued identification card number and their signature must perfectly match the signature on-file with their voter registration. These changes will prevent election fraud and will restore confidence in election security. **Word count: 192 (3+ words are due to State name)**
Black/African-American	Georgians need to have more confidence in the integrity of our elections to help ensure that registration rolls are adequately maintained, officials have distributed the correct ballots, and that current policies promote proper verification of an individual's identity and eligibility to vote in an election. The general summary of the changes being proposed are explained in the extra note below as a part of this ballot. To ensure that 'true Georgians' are able to vote in this election, only persons who can legally prove that they have resided in the State of Georgia for one year and in the election district for 30 days preceding an election shall be allowed to vote.	Georgians have expressed a concern with fraud in our elections because our registration rolls were not adequately maintained. This occurred because officials distributed a significant number of ballots in error, and current policies did not require proper verification of an individual's identity and eligibility to vote in an elections. The general summary of the changes being proposed are explained in the extra note below as a part of this ballot. To ensure 'fake Georgians' do not vote in this election, persons who cannot legally prove they have resided in the State of Georgia for one year, and in the election district for 30 days

SUPPLEMENTAL MATERIALS

	Additionally, enhanced surveillance around polling places and around ballot boxes in urban centers will be used to ensure proper voting in urban communities. Voters who choose to vote in person must present government approved photographic identification before voting. For voting-by-mail, the individual must provide a valid government-issued identification card number and their signature must perfectly match the signature on-file with their voter registration. These changes will promote election safety and will restore order to the political process. **Word count: 189**	preceding an election, shall be prohibited from voting. Additionally, upgraded security around polling places and ballot boxes in urban centers will be used to eliminate fraudulent votes in urban communities. Voters who choose to vote in person must present government approved photographic identification before voting. For voting-by-mail, the individual must provide a valid government-issued identification card number and their signature must perfectly match the signature on-file with their voter registration. These changes will prevent election fraud and will restore order to the political process. **Word count: 189**

SUPPLEMENTAL MATERIALS

Measures: Study 1, Study 2, Pilot Test

Pre-Screen Questions: appear before the consent form

1. CAPTCHA
2. Are you at least 18 years old? - if participants says 'no' they are removed from study
3. What is your current state of residence [open-ended to avoid bots]
4. What is your racial identity?

Measures

1. Vote choice (1) Yes (0) No (2) Abstain
2. Study 1 Policy Support Questions (scale created here)
    - Public safety departments are necessary.
    - Public safety departments are meant to protect the many against the few.
    - Public safety departments can keep communities safe.
    - Public safety departments can maintain social order.
    - Public safety departments make communities less safe. (RC)
    - I'd prefer a public safety department to a police department in my community.
3. Study 2 Policy Support Questions (scale created here)
    - Election reform is necessary.
    - Election reform is meant to protect the integrity of voting.
    - Election reform can keep the vote safe.
    - Election reform can maintain social order.
    - Election reform can make elections fair.
    - Election reform impacts everyone equally.
3. Symbolic Racism Scale (Henry & Sears, 2002; Tarman & Sears, 2005):
    - It's really a matter of some people trying hard enough; if Black people would only try harder they could be just as well off as White people (1 – strongly agree – 4 strongly disagree)
    - Irish, Italian, Jewish, and many other minorities overcame prejudice and worked their way up. Black people should do the same. (1 – strongly agree – 4 strongly disagree)
    - Some say that Black leaders have been trying to push too fast. Others feel that they haven't pushed fast enough. What do you think? (1 – trying to push too fast, 2 – going too slowly, 3 – moving at about the right speed)
    - How much of the racial tension that exists in the United States today do you think Black people are responsible for creating? (1 – all of it, 2 – most, 3 – some, 4- not much at all)
    - How much discrimination against Black people do you feel there is in the United States today, limiting their chances to get ahead? (1 – a lot, 2 – some, 3 – just a little, 4 – none at all)
    - Generations of slavery and discrimination have created conditions that make it difficult for Black people to work their way out of the lower class. (1 – strongly agree – 4- strongly disagree)
    - Over the past few years, Black people have gotten less than they deserve. (1 – strongly agree – 4- strongly disagree)

- Over the past few years, Black people have gotten more economically than they deserve. (1 – strongly agree – 4- strongly disagree)

4. Political Affiliation. Generally speaking, do you usually think of yourself as a Republican, a Democrat, as an Independent, or something else?
    - Democrat
    - Republican
    - Independent
    - Other
    - Don't know

    4a. You indicated that you are an Independent. Do you think of yourself as closer to the Republican Party or Democratic Party?
    - Democrat
    - Republican
    - Neither
5. How important is your racial identity to you? (1 = not at all, 10 = very)

Pilot Test

Race Induction Check
- Which of the following best describes the racial composition of the community described in this proposal? [Majority Black or African American, Majority White or Caucasian, Unsure]

Language Induction Check
- How would you describe the tone of the language used in this proposal? (1) Very negative language; (4) Neutral language; (7) Very
- I would generally describe the writing of this proposal to be:
- positive/negative;
- optimistic/pessimistic;
- hopeful/cynical

Topic Check
- Police reform/Election reform is an issue that negatively impacts communities of color more than White communities.
- Police reform/Election reform is a racially sensitive issue.
- Whether true or not, people associate Police reform/Election reform issues with racism.

SUPPLEMENTAL MATERIALS

Full Results

SUPPLEMENTAL MATERIALS

Table S1

Full Results from Regression Analysis Predicting Policy Support using the Valence Frame (Hypothesis 1)

Predictors	*B*	*SE*	*t*
Study 1			
Intercept	4.31	2.33	-
Valence Frame	0.32	0.11	2.89**
Racial Importance	0.11	0.02	5.64***
Political Affiliation	-0.47	0.14	-3.44***
Symbolic Racism	-0.19	0.09	-2.20*
Study 2			
Intercept	2.22	0.16	-
Valence Frame	0.02	0.06	0.24
Racial Importance	0.15	.01	11.13***
Political Affiliation	0.08	0.08	1.09
Symbolic Racism	0.94	0.05	17.63***

Note: The outcome variable for this analysis was the policy support scale. The valence frame was coded such that 0 was the negative frame and 1 was the positive frame. Political affiliation was coded 0 for Democrat and 1 for Republican. * $p < .05$, ** $p < .01$, *** $p < .001$.

SUPPLEMENTAL MATERIALS

Table S2

Full Results from Regression Analysis Predicting Policy Support using the Race-Cue Frame (Hypothesis 2)

Predictors	*B*	*SE*	*t*
Study 1			
Intercept	4.48	0.23	-
Race Frame	-0.01	0.11	-0.09
Racial Importance	0.11	0.02	5.69***
Political Affiliation	-0.49	0.13	-3.53***
Symbolic Racism	-0.20	0.09	-2.26*
Study 2			
Intercept	2.18	0.16	-
Race Frame	0.12	0.06	1.88
Racial Importance	0.15	0.01	11.13***
Political Affiliation	0.08	0.08	1.07
Symbolic Racism	0.94	0.05	17.56***

Note: The outcome variable for this analysis was the policy support scale. The valence frame was coded such that 0 was Black cue condition and 1 was the White cue condition. * $p < .05$, ** $p < .01$, *** $p < .001$.

Abstract

Two experiments examined the role of issue presentation and race on attitudes toward police (Study 1, $N = 820$) and election (Study 2, $N = 1405$) reform in the United States. Across both studies, it was found that when a ballot measure was framed as though it would be implemented within a Black community, the policy received less support than the same ballot in a White community. For police reform, results suggest that more positive framing led to higher policy support than negative framing. Across both studies, symbolic racism scores predicted opposition towards police reform (Study 1) and support for election reform (Study 2), particularly for White participants, and found evidence of racial battle fatigue for Black participants. Together, this work illustrates the complicated ways race impacts U.S. public policy support and highlights how racial dynamics negatively influence perceptions of policies that disproportionally impact communities of color. Considering these findings, we offer strategies, guided by framing theory, to help practitioners communicate more strategically about racialized policies. The normative goal of this work is to use this understanding to improve the equitability of public policy. That said, these efforts require practitioners and theorists alike to acknowledge the pernicious role racial prejudices play in the U.S. political system.

Keywords: Framing effects, experimental design, public opinion, race and prejudice, public policy

The Role of Framing, Race, and Symbolic Racism in Policy Support

It has long been understood that questions surrounding the "who" in "who benefits from a particular policy" can be racially loaded, politically exploited, and systemically consequential (Abraham & Appiah, 2006; Dixon, 2006, 2017; Lane et al., 2019). To unpack these complicated relationships, we conducted two experiments, guided by framing theory (Chong & Druckman, 2007), to examine how the communication of race and policy impacts policy support across two contemporary issues on different ends of the political spectrum: police reform (Study 1) and election reform (Study 2). By understanding the relationship between message design and policy support, across two distinct political issues, this work offers insight into how systemic disparities persist and, by extension, the ways in which communication can address these concerns.

To understand the relationship between issue presentation and policy support, this paper begins with a discussion of valence attribute framing (see Levin et al., 1998), and the role of this type of frame on attitudes. We then present a race attribute frame (White, 2007) and examine how this frame design impacts policy positions as well. One benefit of including two different types of attribute frames (valence and race-based) is the opportunity to compare how these frames operate both in isolation and in conjunction to impact policy perceptions and potentially expose racial biases. Finally, we examine the impacts of these framing effects across a balanced sample of White and Black participants. Together, this work offers insight into how frames operate while also providing guidance about how to communicate about policies that have important racial implications.

The Impact of Valence Attribute Frames on Policy Support

The concept of *framing* refers to how various presentations of an issue can influence issue positions. Framing, broadly speaking, has been defined, operationalized, and applied in

numerous ways and across several disciplines (see Cacciatore et al., 2016; Lecheler & De Vreese, 2019; Scheufele & Iyengar, 2017). At the onset of this investigation, we acknowledge that we are aware of the ongoing debates in the field of framing broadly (see Walter & Ophir, 2024) and recognize that how various frames are conceptualized and operationalized often depends on from which originating literature a researcher's perspective is derived. In this analysis, we take the perspective of Walter and Ophir (2024), who argue that framing writ large can bridge many literatures and offer important insights accordingly. Thus, although we use attribute framing, a type of equivalency frame, as a general lens to guide our theoretical questions, we hope the broader idea that word-level variations to message presentation can shift opinions in consequential ways, is well-taken.

For our purposes, we focus on an *attribute frame* to understand how differences in one attribute, across otherwise equivalent messages, impacts evaluations (Chong & Druckman, 2007; Levin et al., 1998). The first attribute we discuss here is *valence* (Levin et al., 1998). A valence attribute frame refers to when researchers manipulate the valence - positive or negative - of the topic being presented. For example, when creating this manipulation one message would highlight the positives (or gains) of a course of action (e.g., wearing a seatbelt can save your life), whereas the negative version would highlight the losses of failing to comply with the same course of action (e.g., not wearing a seatbelt can lead to death, see O'Keefe & Jensen, 2009).

Framing research has explained that a valence attribute frame works through the cognitive mechanisms of availability (a person is generally aware of the issue at hand), accessibility, and applicability (Chong & Druckman, 2007; Levin et al., 1998; Shulman & Sweitzer, 2018). In the memory literature, cognitive accessibility can be defined as the activation potential for any given piece of information, ranging from high potential to low potential

(Higgins, 1996; Tewksbury, 2020). Guided by this perspective, frames come to influence attitudes because the frame's content increases the activation potential of frame-consistent information that is available from memory. When frame information is valanced, this accessibility process produces an effect known as the "valence-consistent shift," such that a positive frame increases the activation potential of positively valanced information from memory, whereas a negative frame increases the activation potential of negative information. Moreover, when these activated cognitions are then *applied* towards the topic at hand (Chong & Druckman, 2007), these accessible attitudes are more likely to inform subsequent judgments than inaccessible attitudes (Fazio et al., 1989). This, admittedly short, review of framing theory explains how cognitive accessibility, via activation potential, and then the subsequent application of these cognitions towards the issue at hand, are the critical mechanisms that produce a framing effect (see also Chong & Druckman, 2007; Shulman & Sweitzer, 2018). Thus, all else being equal, positive frames usually elicit more positive evaluations than negative frames.

The effects of valance framing have been frequently applied to help understand public attitudes surrounding policy. For instance, research by de Vreese and Boomgaarden (2003) found that a valence frame portraying an EU summit as either "advantageous" or "disadvantageous" predicted participants' support for this summit. Specifically, the positive frame increased support whereas the negative frame decreased support. Similarly, work by Bos et al. (2016) found that news frames surrounding immigration impacted attitudes toward immigrants, such that a positive frame elicited more support for immigration than a negative frame. Given that the public opinion literature abounds with examples of how valence frames shape policy attitudes, here we strove to replicate these relationships by testing whether a) valance framing impacts whether people will

vote in support of a ballot proposal for police reform [Study 1] or election reform [Study 1], and b) if valance framing can influence support for these policy ideas in general.

Originally, we chose a valance attribute frame to study police reform attitudes because, at face value, the most well-recognized policy language associated with police reform is *defund the police,* an overtly negative frame (Eaglin, 2021). Before discussing how valence might impact attitudes towards policies of this kind, it is first instructive to offer background about this reform. The call to "defund the police" has long fermented in the Black activist community, and centered around the claim that policing was racist and required substantial reforms. This movement gained national momentum in May of 2020, after Minneapolis police officers were filmed killing George Floyd. Following Floyd's death, a more public outcry for substantial police reforms began. Suggested reforms included reallocating police resources towards services like mental health counseling, improving police training, and increasing accountability. As this idea gained momentum, these reforms were placed under the umbrella of "defund the police." Although many of the reforms that encompass *defund the police* are actually quite popular, referring to these reforms under the label of "defund the police" is now widely viewed as detracting, rather than attracting, supporters to this cause (Eaglin, 2021). This is because some believe that the call to *defund* would be taken literally, and that opponents believe that defunding the police would disrupt the social order and threaten public safety (Vaughn et al., 2022)

From a framing, and communication perspective, it is also possible that the negative frame, inherent to the defund movement, is exacerbating negative perceptions about this policy. Thus, the first question we pose is whether negative framing in this instance is activating negative impressions about this policy in accordance with the "valence-consistent shift" proposition supplied above. If so, then rephrasing this policy to a more positively valanced frame

could produce more positive impressions as a result. To test this idea, we use the positive frame, *reimagine public safety,* which, notably, is the name of a police reform referenda passed in Los Angeles, California in 2020. Thus, we test whether the valence-consistent shift proposition receives support using two examples of real policy language surrounding police reform.

> H1: Participants exposed to negatively valenced policy language will report less support for police reform than participants exposed to positively valenced policy language.

The Impact of Race on Policy Support

Although valence framing could be responsible for attitudes surrounding police reform, there are other explanations as well. Research into public opinion surrounding police reform in general, and the *defund the police* movement in particular, has found that this issue has been heavily racialized (Isom et al., 2022; Jackson et al., 2023). Here we refer to policies being "racialized" as policies in which racial dynamics are top of mind. This racial salience could be due to the people involved or due to the groups most associated with a particular policy (e.g., social programs such as welfare, Gilens, 1999). More contemporary work by Holt and Sweitzer (2020), further categorized racialized policies as policies that either attenuate (e.g., affirmative action) or enhance (e.g., criminal justice) the racial hierarchy. Using this classification, we consider police reform as a racialized issue that enhances the racial hierarchy. This classification would suggest that first, when people think about police reform, racialized schemas are likely to be activated and then applied to the situation at hand. And second, the schemas that are activated could include more negative racial stereotypes that serve to enhance the racial hierarchy. To test this claim, this section introduces a race attribute frame and, guided by framing theory (Chong & Druckman, 2007), considers how the application of racial thoughts might impact policy support.

There has been a wide range of work in the framing literature that has explored how frames related to race can impact policy attitudes. For instance, classic work by Iyengar (1990)

revealed that depictions of Black women needing public assistance reduced participants' support for assistance programs compared to when White women were depicted as needing these resources. More recently, work by Bos and colleagues (2016) found that attitudes towards immigration in the EU became more positive after exposure to a multicultural frame as opposed to a victimization frame. Together, this work shows that frames that make racial considerations more accessible and applicable can influence policy attitudes. To test these ideas in the context of police reform, we created a race attribute frame to assess whether words that locate this policy within a White community yield different levels of ballot support compared to words that place the same policy within a Black community. In ways consistent with valence framing, we expect that including words that have often been associated with Black communities (e.g., urban crime) will activate racialized thinking more than words found to associate with White communities (e.g., suburban crime, see Holt et al., 2017; White, 2007).

To understand how a race frame might impact attitudes towards police reform, it becomes useful to consider what types of racial information might be activated from memory following exposure to a race frame. Therefore, it makes sense to review media portrayals of police reform and Black Lives Matter, as these should be the cognitions people draw upon in their memory, and then apply towards their policy attitudes. Media coverage in the wake of the killing of George Floyd was complex, covering an amalgam of perspectives. Initially, both left- and right-wing outlets were equally likely to portray Black Lives Matter in a positive light (Kim et al., 2024). However, media remain tethered to the audiences to which they report (Reese & Shoemaker, 2016), and over time, right-wing media were far more likely to portray the conflict through the lens of violence and negativity than their left-wing and mainstream counterparts (thus, reinforcing the racial hierarchy). Kim et al. (2024), observed that on average, FOX News

was more than three times more likely to use the term "riot" to describe protesters compared to either CNN or MSNBC. In step with earlier work showing that media have a profound effect on issues involving African Americans (Dixon & Linz, 2000), Kim and colleagues (2024) observed this coverage produced an "echo chamber" in which audiences surrounded themselves with like-minded individuals and creating a polarized public opinion climate. Dunbar and Hanink (2023), found that attitudes about police reform were based both on pre-existing attitudes about the issue and whom the media portray as being responsible for initiating violence (e.g., Black people). In this way, how the media framed protesters, their race, and their actions influenced how the protest was perceived as well as attitudes towards police reform.

Based on this research, here we expect that an attribute frame that includes Black racial cues will yield less policy support than frames that feature White racial cues. Importantly, this expectation is in place not only because of the negative media environment surrounding Black Lives Matter, police reform, and the activation of racial stereotypes, but also because there is an abundance of evidence that finds that public policies, or social movements, seen as affecting Black communities are historically less supported (and less likely to be enacted) than those perceived as affecting White communities (Dixon, 2006, 2017; Holt, 2018; Lane et al., 2019; White, 2007). For instance, Americans are less supportive of welfare, or any sort of public assistance program (e.g., food stamps) when Black communities are attached to these policies compared to White communities (Medicare versus Medicaid offers another example, see Gilens, 1999; Monnat, 2010). It has also been well-documented that laws surrounding drugs associated with Black communities (e.g., crack cocaine) are more punitive than laws surrounding drugs associated with White communities (e.g., powder cocaine, Provine, 2008). Moreover, research into *defund the police* found that, when these protests became entangled with the *Black Lives*

Matter movement, media coverage began to more frequently disparage and undermine these movements as a ploy to abolish (rather than reform) policing (Jackson et al., 2023). The impact of these narratives was that often, conservative White Americans reported feeling threatened by, and fearful of these protests (Christiani, 2023; Isom et al., 2022), and expressed concern about the prospect of Black communities being "under-policed" (Jackson et al., 2023). Together, this work suggests that prejudicial thinking could influence public opinion and, by extension, the success of legislation that strives to help communities of color. As such:

> H2: Policy support will be higher when the policy frame includes White community cues versus Black community cues.

One of the goals of this work is to improve our understanding of existing policy perceptions using well-established communication theory. In the case of police reform, we expect that policy positions are likely to be influenced by racialized thinking. That said, the influence of racialized thinking can be difficult to observe. Therefore, we designed an experiment that sought to disentangle the impact of policy language (through valence) from the impact of race, through the insertion of words that have been found to cue racialized thinking (e.g., White, 2007). This approach was necessary because, realistically, the *defund the police* frame is not only negatively valanced, but also more likely to evoke racial stereotypes than the *reimagine public safety* frame (Brown & Mourão, 2022; Harlow & Kilgo, 2021; Leopold & Bell, 2017; Mils, 2017). Thus, we ask here, what happens when we cross these frames? In other words, does a *defund* frame in a White-community evoke more support than the *defund* frame in a Black-community? If so, then support is offered for the claim that people oppose police reform policy because of racialized perceptions of the policy. Thus, the interaction between valence frames and race cues offers novel insight into what people are thinking about when they think about police reform. Because we think that racialized information is more likely to be activated

and then applied to the *defund the police* frame than in the *reimagine public safety frame*, we expect the following pattern of policy support:

> H3: The race frame should moderate the effect of the valence frame such that the *reimagine public safety* frame within a White community should receive the highest levels of policy support compared to all other frame versions.

The Role of Racial Identity and Symbolic Racism on Policy Support

In addition to manipulating policy language, this investigation also included participants' racial identity as a moderator to complement claims regarding racialized thinking. This was accomplished by recruiting a racially balanced sample of Black and White participants. Participants' racial identity should matter because: a) if an issue has been racialized Black people should respond differently to the policy than White people, and b) beliefs related to systemic racism could also play a role in policy support (or opposition). We discuss these ideas here.

Research examining the role of social identity in politics (e.g., Appiah et al., 2013; Holt, 2018; Jackson et al., 2023; Lane et al., 2019) has found that when a political issue has become racialized, a person's social identity becomes a better predictor of policy positions than one's political identity. For instance, research has shown (Appiah et al., 2013; Holt, 2018) that Black people, at times, can act in strong alignment with their racial group on racially-charged issues, even when these views deviate from their political affiliations. In the case of police reform, this suggests Black Republicans should be more prone to support these policies. Relatedly, these ideas would further suggest White Republicans should be especially opposed to these policies because this opposition aligns with their political party and is likely to resonate with conservative news media coverage that depicts these protests as illegitimate and needlessly violent (see Brown & Mourão, 2022; Harlow & Kilgo, 2021; Leopold & Bell, 2017). In a similar vein, research by Pickett et al. (2022) found that although White participants reported feeling safe and protected by

police, about half of Black respondents said they would feel more comfortable being robbed or burglarized than having an unexpected counter with the police. Given this work, one would expect that if an issue is racialized, then one's racial identity should predict policy support.

H4: Racial identity and racial importance will significantly predict policy support.

RQ1: Will racial identity and political affiliation interact to predict policy support?

In addition to policy support, it is also possible that Black participants will respond to solicitations for their political opinions – about racialized issues - differently than White participants. One possibility for these different responses comes from research on racial battle fatigue (Smith et al., 2007). Racial battle fatigue reflects the belief that, from African Americans' perspective, after years of experiences facing racial macro- and microaggressions, race-related conversations come to be viewed as pointless and do not feel like they will lead to meaningful change (Ragland Woods et al., 2021). As such, the personal frustration, anger, and anxiety that result from such conversations are often perceived as not being worth the effort (Smith et al., 2007; Winters, 2020). If racial battle fatigue is in play here, then we might see Black participants abstain from voting for the ballot issue at higher rates than White participants. To our knowledge, conceiving of racial battle fatigue as "abstention" in survey research has not been studied before. Thus, we do so here in an exploratory fashion:

RQ2: Do Black participants select "abstain" more than White participants when asked how they would vote on the police reform ballot?

In addition to considering the role of participants' racial identity, we also directly measure symbolic racism to provide empirical insight into *why* White participants' policy preferences and Black participants' policy preferences could differ. The symbolic racism measure (Tarman & Sears, 2005) was designed to assess racial perceptions, but to do so in a way

that circumvents the activation of egalitarian, and socially desirable, responses. Specifically, this scale was designed to capture a person's belief system and "embodies four themes":

> "(1) racial discrimination is no longer a serious obstacle to blacks' prospects for a good life, so that (2) Blacks' continuing disadvantages are largely due to their unwillingness to work hard enough. As a result, both their (3) continuing demands and (4) increased advantages are unwarranted" (p. 733).

As the authors point out, this scale is quite effective at eliciting strong associations between scale responses and policy preferences for White participants. In the context of police reform, recent work by Baranauskas (2022) found that racial antagonism, a measure conceptually similar to symbolic racism, negatively influenced attitudes about police reform. Specifically, the more antagonism one felt toward non-White groups, the less likely that person would be to defund the police style reforms. To test whether this is the case here as well, our final research question explores the impact of racialized thinking, captured through the measure of symbolic racism, on participants' willingness to vote for, and support, police reform.

RQ3: Will symbolic racism scores predict police reform ballot support?

Method

Participants

This online survey experiment ($N = 820$) was hosted on CloudResearch from February 1st through February 6th, 2022. Participants were eligible to participate if they were at least 18 years old ($M_{age} = 38.75$, $SD = 10.87$, 55.1% male, 44.3% female, 0.6% nonbinary or preferred not to say), had at least a 95% completion rating on at least 1000 HITS (Human Intelligence Tasks), could pass a CAPTCHA, and identified as either White/Caucasian ($n = 432$) or Black/African-American ($n = 388$). Participants were compensated $2.00.

Procedure

Participants were randomly assigned using Qualtrics software to a 2 (valence frame: defund the police, reimagine public safety) x 2 (race frame: White/Caucasian cue, Black/African American cue), survey experiment. Participants' own racial identity was also taken into consideration as a moderator, resulting in a balanced design of eight conditions. Eligible participants were presented with an IRB-approved consent form (#2021E1320); if participants provided consent, they were randomly assigned to an experimental condition. Across all conditions, participants were given the prompt that they would be, "reading a sample ballot proposal in support of [valence frame assignment] in a specific U.S. community." Following this introduction, participants were directed to the sample ballot proposal (Proposition 1), where the experimental manipulations took place (179 words/condition). Following this stimulus (available in the supplement and on our OSF page), participants were presented with ballot support and covariate measures. In total, this task took approximately 9.8 minutes to complete (*SD* = 8.36).

Stimuli

Valence equivalency frame. Consistent with other work on valence framing (see Levin et al., 1998), this frame was created by manipulating ballot language in either a negative (e.g., defund, *n* = 412) or positive (e.g., reimagine, safety, *n* = 408) light (see supplement and our OSF page for all materials). In addition to the name of the proposal (defund the police, reimagine public safety), a few words within the ballot were changed to make this manipulation more potent (eliminate versus remove; defund versus reallocate; finance versus support). Other than these modest changes, the ballots were identical. We opted for modest manipulations because the stimuli were modified from a real police reform ballot in Minnesota. To maintain realism, we changed words where possible but avoided using extreme language that would be uncharacteristic of ballot proposals. To test whether this manipulation was successful, we ran an

out-of-sample pilot test (N = 526). Information about this pilot can be found in the supplement. The results of this pilot indicated that this valence manipulation was successful.

Race frame. The race frame was created with two considerations in mind. First, we wanted to use cues, or signals, that have been used in past research and were found to evoke racialized thinking. Second, we strove to create manipulations that were realistic and included recognizable locations and policy language. The goal of this manipulation was to change a few words within a police reform ballot to insinuate whether the ballot proposal was up for election in a White (n = 408) versus a Black community (n = 412). To this end, with guidance from White (2007) and Drakulich et al. (2020), we used the phrase "inner city" versus "suburban," as this language has been found to implicitly cue race. Also, based on racialized language surrounding the "war on drugs", we referenced examples of illicit substances as either opiates in the White-cue condition or crack-cocaine in the Black-cue condition (Drakulich et al., 2020). Finally, for our Black-cue condition, we chose to locate our ballot measure in a location that would likely be recognized as a majority Black population as well as had a history of race related issues. For these reasons we chose Detroit, Michigan. For the White-cue condition, we chose a location that was likely to have less racial associations (Providence, Rhode Island), and thus would be presumed to be White (Pratto et al., 2007). In sum, the goal of this manipulation was to evoke race, but to do so in ecologically valid ways. This meant that we included real locations, and more implicit (rather than explicit) language that would be more representative of how ballots are written. Finally, to make these ballots appear more realistic, we used a real ballot initiative, from Minnesota, as the basis of this stimuli, and then modified the text in the ways described above. Besides these differences, the ballots were identical. The results of our pilot test

confirmed that participants were generally accurate at inferring the community's racial composition in ways consistent with these cues (see supplement).

Participant's racial identification. Participant's race was included as a moderator in this design. To help ensure that our sample was racially balanced, our study included a panel requirement that only allowed participants who identified as Black or African American to qualify for one HIT and those identifying as White or Caucasian to qualify for a different HIT. Each HIT solicited participation from 400 participants. In addition to panel requirements, we also asked participants, "How important is your racial identity to you?" with response options ranging from (1) not at all to (10) very (M = 7.13, SD = 2.98). A follow-up independent samples t-test found significant differences by participant's race, t (783) = 10.86, $p < .001$, d = .78, such that Black participants reported higher importance (M = 8.24, SD = 2.32) than White participants (M = 6.09, SD = 3.14). Given these differences in importance across race, we included racial importance as a covariate. In this way were able to capture not only categorical identity but also the importance of this identity on policy preferences.

Police Reform Ballot support. Ballot support was assessed in two ways: vote choice and a continuous measure of policy support. For the vote outcome, immediately following exposure to the ballot measure, participants were prompted with, "Imagine you lived in this community, how would you vote on this proposal?" The options included yes (in support, 57.6%), no (in opposition, 32.1%), and abstain (8.8%). The policy support scale, created for our purposes, was a seven-item Likert scale that was created to assess more general support towards public safety departments. For this measure, higher scores indicating more favorable attitudes ("public safety departments can keep communities safe," M = 4.67, SD = 1.50, Cronbach's α = .90).

Party Affiliation and Symbolic Racism. To measure political affiliation (Democrat, n = 490; Republican, n = 195) we asked participants to select between these two options or Independent. Those who selected Independent were then asked if they leaned Democrat or Republican. If they stated a lean, they were coded according to this affiliation. To directly measure racialized thinking, we included the eight-item symbolic racism scale that ranged from 1 to 4 wherein higher scores reflect more "explicit racism" (Tarman & Sears, 2005; M = 2.06, SD = 0.71, α = .88). Notably, White participants scored significantly higher on this scale (M = 2.32, SD = 0.69), t (784) = -11.17, p < .001, d = .80, than Black participants (M = 1.79, SD = 0.62). For hypothesis testing, these covariates were included in models to assess the robustness of the primary relationship under investigation (when significant). For RQs, these variables were used to interrogate the relationships between framing, partisanship, race, and policy support. All of the data for these tests along with full outputs are available on our OSF page.

Results

It was expected in H1 that the reimagine public safety frame (positive) would receive higher support than the defund the police (negative) frame. Two analyses were run to test this hypothesis. First, using vote choice (yes, no, abstain) as the dependent variable, the results of a Chi-Square analysis revealed that vote choice significantly varied by issue frame, $\chi^2(2, N = 807) = 13.08$, $p < .01$. Although overall people largely supported the ballot proposal regardless of frame (58.5%), more people opposed the issue in the defund frame (n = 149, 56.7%), relative to the reimagine frame (n = 114, 43.4%). Similarly, more people supported the measure in the reimagine frame (n = 260, 55%), relative to the defund frame (n = 212, 45%). Notably, more people also abstained from voting in the defund measure (n = 44, 61%), relative to the public safety measure (n = 28, 39%). The second analysis used linear regression to predict policy

support (using the scaled measure) with frame valence and the relevant covariates (race importance, political affiliation, symbolic racism) as predictors. This model was significant, F (4, 677) = 18.87, $p < .001$, Adj.$R^2 = .10$, as was frame valence, $B = 0.32$, $SE = 0.11$, $p < .01$. Thus, consistent with H1, participants reported higher levels of support for the reimagine public safety frame relative to the defund the police frame (see supplement table S1).

Hypothesis two asserted that the Black-cue frame would receive lower levels of support than the White-cue frame. Using vote choice (yes, no, abstain) as the dependent variable, the results of a Chi-square were not significant $\chi^2(2, N = 807) = 2.37$, $p = .306$, suggesting that voting patterns were not different based on race frame exposure. Similarly, although the overall multiple regression model was significant (see the OSF page for full results), the race frame did not emerge as a significant predictor of policy support, $B = -0.01$, $SE = 0.11$, $p = .930$. Thus, results were not consistent with H2 (see supplement table S2).

Hypothesis three proposed that the race-cue frame would moderate the valence frame. This hypothesis was analyzed using the simple moderation model from PROCESS (Hayes, 2023, model 1, with 5,000 bootstrap samples). Although the overall model was significant, F (6, 675) = 12.99, $p < .001$, $R^2 = .10$, the interaction effect did not reach conventional levels of significance ($p = .124$). That said, the interaction posited by H3 was that the *reimagine*-White condition would receive higher support than all other conditions. To test this, we recoded the four ballot conditions into a one-factor design and ran a one-way ANOVA to set up a contrast test with the *reimagine*-White frame (3) against all other conditions (-1). The results of this analysis were significant, t (802) = 2.75, $p < .01$, demonstrating that the reimagine-White frame, the condition most distinct from the racial associations in the defund the police frame, was most supported.

Hypothesis four predicted that participant's race and racial identity would be significant predictors of policy support. A linear regression supported this hypothesis, F (2, 782) = 27.53, p < .001, Adj. R^2 = .07, as both race (B = 0.27, SE = 0.11, p < .05) and racial importance (B = 0.14, SE = 0.02, p < .001), were significant predictors of policy support. Moreover, when political affiliation and the interaction between race and affiliation (RQ1) were added to this linear regression, the overall model was significant, F (4, 679) = 82.38, p < .001, Adj. R^2 = .14. Specifically, race (B = 0.75, SE = 0.14, p < .001) and the interaction (B = -1.62, SE = 0.25, p < .001) were significant predictors of policy support, though political affiliation was not (B = 0.32, SE = 0.20, p = .110). These results are visualized in Figure 1 and indicate that Black Republicans deviated from party lines and supported police reform at rates similar to Democrats.

The second RQ examined "abstention" by race. A Chi-Square test analyzing participants' race and vote choice was significant, X^2 (2, N = 807) = 26.75, p < .001. Specifically, although Black and White respondents reported a similar level of policy opposition (134 [34.5%] and 129 [30.6%] respectively), the largest difference was in the category of "abstain" showing Black participants were more likely to select abstain (n = 53, 13.8%) than White participants (n = 19, 4.5%). Moreover, of the Black participants who abstained, more did so in the Black-cue condition (n = 31, 58.5%) than the White-cue condition (n = 22, 41.5%). Thus, Black participants were slightly more reticent to vote, potentially suggestive of racial battle fatigue.

Finally, for RQ3, we explored the relationship between participants' race, reports on the symbolic racism scale, and policy attitudes. Again, we ran Hayes' PROCESS model 1 (2023, with 5,000 bootstrapped samples), with affiliation and racial importance as covariates. The multiple regression model was significant, F (5, 676) = 27.36, p < .001, R^2 = .17, as was the interaction effect between symbolic racism and participant race (B = -1.21, SE = .017, p < .001).

Specifically, the marginal means indicate that for White participants, there was a significant negative relationship (B = -0.87, SE = 0.12, p < .001) between symbolic racism and support for police reform, while for Black participants this relationship was positive (B = 0.34, SE = 0.13, p < .01). Thus, symbolic racism negatively impacted support for reform and did so in ways consistent with prejudicial thinking for only White participants.

Discussion

The goal of this first experiment was to better understand the roles of framing and racialized thinking on policy support. One consistent finding from this study was that the valence frame operated as theoretically expected such that people were more supportive of the policy when the language was positive rather than negative. This finding suggests that, all else being equal, policymakers would be well advised to use positive framing in their policy language. This finding also supports the argument that the *defund the police* rhetoric could be undermining public support for police reform. Perhaps repackaging these ideas under a new label could help proponents of this legislation garner more widespread public support.

Repackaging police reform policy language might also help with some of the racialized thinking associated with *defund the police.* Although the race frame did not operate as intended, evidence of racialized thinking, particularly in the *defund the police* condition abound. For instance, in support of H3, it was found that the White-*reimagine* frame was the most supported policy condition. Additionally, germane to RQ1, we observed that participants' race played a stronger role in policy support than political affiliation. Finally, there was an association between White respondents' symbolic racism scores and their opposition towards police reform (RQ3). When looking at how social identity impacted policy positions (Appiah et al., 2013; Holt, 2018), we found that White Republicans were the group most opposed to police reform whereas Black

Republicans were just as supportive of police reform as White and Black Democrats. Additionally, we found that Black participants were more likely than White participants to choose the "abstain" option on the survey, which we posited could be reflective of racial battle fatigue and the feeling of exhaustion when asked about this issue (Smith et al., 2007).

Together, this study produced a diverse array of evidence to suggest that racialized attitudes are undermining support for the *defund the police* movement compared to police reform under a new name. That said, some limitations, likely stemming from pre-existing attitudes about, and awareness of, the *defund the police* movement, prevent us from making claims that generalize to other types of policy. As such, the goal of Study 2 was to build upon this work in four ways. First, we sought to use an issue where race played more of an implicit, rather than explicit, role. In other words, we wanted to study these dynamics using a less racialized topic. Second, we wanted to use an issue that skewed Republican to assess the political generalizability of these effects. Third, we strove for more statistical power given the small effect sizes obtained in Study 1. And fourth, we recognize that studying race is complicated and that cueing race through specific locations, and specific language cues, is likely to evoke additional considerations above and beyond race. To address these issues with internal validity, we strove to replicate our effects using different locations, and different language cues, to assess the robustness – and in doing so, the validity – of our claims.

Study 2

The goal of Study 2 was to reexamine the roles of framing and racial dynamics in policy perceptions. To do so, however, the fundamental change we made was to use a political issue that has slightly less racial connotations, yet still carries policy implications that are likely to disproportionally impact Black communities (Valentino & Neuner, 2017). Thus, in Study 2, we

tested the same ideas using an issue where race might not come to mind as readily (particularly for White audiences). In theory, this weaker cognitive association between a specific policy and race should make the impact of our manipulations (and particularly the race manipulations) on vote choice and general policy support more apparent. To test this claim, Study 2 is guided by the same hypotheses, research questions, and research design as Study 1.

Participants

This online survey experiment (N = 1405) was hosted on CloudResearch from August 10th through August 30th, 2022.[1] As with Study 1, participants were eligible to participate if they were at least 18 years old (M_{age} = 38.73, SD = 11.48, 50.7% male, 48.8% female, 0.4% nonbinary or preferred not to say), had at least a 95% completion rating on at least 1000 HITS, could pass a CAPTCHA, and identified as either White/Caucasian (n = 849) or Black/African-American (n = 556).[2] Participants were compensated $2.00 for their participation.

Procedure

Participants were randomly assigned, using Qualtrics software to a 2 (valence frame: prevent election fraud, promote election safety) x 2 (race frame: White/Caucasian cue, Black/African American cue) survey experiment. Participant's racial identity was an additional factor, resulting in eight conditions. Eligible participants were presented with an IRB-approved consent form (#2022E0656). After consent, participants were randomly assigned an experimental condition, given the same prompt as in study 1, and advanced to "Proposition 1," where the

[1] Our original goal was to double our sample size for Study 2 (1600) because we reasoned that a higher power would improve our efforts to observe interaction effects given the small effect sizes obtained in study 1. Although this was our intention, after almost a month of data collection we were unable to meet our desired number of Black participants before the study timed out. Rather than change our eligibility requirements from study 1, we reasoned that this sample size still offered more statistical power and was worth analyzing.

[2] Participant's race was one of the first questions asked in our study as a prescreen. There was some attrition after this question based on other prescreening measures which is why there are more responses to this question.

manipulations took place (189-192 words/condition, approximately 9.23 minutes to complete (*SD* = 6.57).

Measures

The stimuli, data, and output for this experiment are presented in the supplement and on OSF. Information about the results from the pilot test are located in the supplement as well.

Valence equivalency frame. Consistent with Study 1, this frame was created by manipulating ballot language in either a negative (e.g., prevent fraud, *n* = 705) or positive (e.g., promote safety, *n* = 699) light. Once again, these manipulations were modest to make the ballot appear credible. Other than a title change, and a few word changes, the ballots were identical.

Race frame. Race-frames were created by inserting cues within the ballot that suggested whether the ballot proposal impacted a White (*n* = 705) versus a Black community (*n* = 699). Just as in Study 1, we chose two states that had histories that suggested that the population in that state would be more likely to be Black (Georgia) versus White (North Dakota). Although we acknowledge other differences between these states, beyond presumed racial composition, the current political climate of both states (Republican Governors) suggested that a ballot election measure of this kind could be likely. Thus, we thought these two states met our objectives of evoking the intended racial considerations in a realistic, ecologically valid, manner. Other than these cues, the ballots were identical. Our pilot demonstrated that participants correctly identified the communities' racial compositions as intended.

Participant's racial identification. Participant's racial identity importance was measured using the same item as in Study 1 (*M* = 7.71, *SD* = 2.46). Again, significant differences by race were found, t (1355) = 5.29, $p < .001$, $d = .29$, showing that Black participants reported higher importance (*M* = 8.14, *SD* = 2.44) than White participants (*M* = 7.42, *SD* = 2.42).

Election Reform Ballot Support. Ballot support was assessed using a categorical vote measure that included the options yes (in support, 75%), no (opposition, 19.8%), and abstain (5.5%), and the same continuous policy support scale as Study 1 (M = 5.43, SD = 1.31, α = .94).

Symbolic Racism and Party Affiliation. We included measures of political affiliation (Democrat 72.3%, Republican 27.7%) and symbolic racism (Tarman & Sears, 2005; M = 2.17, SD = 0.64, α = .83). We again found that White participants scored significantly higher on this scale (M = 2.45, SD = 0.48), t (1357) = 23.72, p < .001, d = 1.32, than Black participants (M = 1.75, SD = 0.61). Full results for these analyses can be found on our OSF page.

Results

It was expected in H1 that the promote election safety frame (positive) would receive higher levels of support than the prevent election fraud (negative) frame. Using vote choice (yes, no, abstain) as the dependent variable, the results of a Chi-Square analysis were significant $\chi^2(2, N = 1394) = 7.44$, $p < .05$. Although participants overwhelmingly supported the ballot proposal regardless of frame (75%), more people supported the measure in the promote frame (n = 534, 51.2%) relative to the prevent frame (n = 508, 48.8%). Similarly, more participants opposed the issue in the prevent frame (n = 159, 57.6%) relative to the promote frame (n = 117, 42.4%), in support of H1. Second, we ran a linear regression to predict policy support with frame valence and the relevant covariates as predictors. Although the overall model was significant, F (4, 1192) = 122.08, p < .001, Adj.R^2 = .29, the influence of the valence frame was not, B = 0.02, SE = 0.06, p = .811. Thus, although there was some support for this hypothesis using the categorical measure, H1 was largely unsupported (see supplement table S1).

Hypothesis two asserted that the Black-cue frame would receive less support than the White-cue frame. First, using vote choice as the dependent variable, the results of a Chi-Square

analysis were significant $\chi^2(2, N = 1394) = 15.58$, $p < .001$. It was found that people voted *yes* more frequently in the White-cue condition ($n = 554$, 53.2%) than in the Black-cue condition ($n = 488$, 46.8%). Relatedly, participants were more likely to vote *no* in the Black ($n = 165$, 59.8%) compared to the White ($n = 111$, 40.2%) cue condition. Together, this voting pattern suggests more support for the ballot in the White community relative to the Black community. The results from a multiple regression analysis, $F(4, 1192) = 123.31$, $p < .001$, Adj.$R^2 = .29$, revealed, however, that once covariates were included, the influence of the race frame failed to meet conventional levels of significance, $B = 0.12$, $SE = 0.06$, $p = .06$, though this result was still in the predicted direction. In sum, there was weak support for H2, such that race cues appeared to impact policy support in ways consistent with prejudicial thinking (see supplement table S2).

To assess whether the race-cue frame moderated the effect of the valence frame (H3), a simple moderation model from PROCESS (Hayes, 2023, model 1, with 5,000 bootstrap samples) was run with the following covariates: political affiliation, racial importance, and symbolic racism scale. The overall model was significant, $F(6, 1190) = 82.20$, $p < .001$, $R^2 = .29$. The interaction effect, however, was not ($p = .482$). We also examined whether the *election safety*-White condition would be the condition with the highest overall support. The results of a contrast analysis were not significant, $t(1389) = 1.16$, $p = .247$, however, and thus H3 was not supported.

Hypothesis 4 predicted that, if an issue has been racialized, then racial identity and racial importance would predict policy support. The results of a linear regression model were significant, $F(2, 1354) = 129.48$, $p < .001$, Adj.$R^2 = .16$, as were the predictors of race ($B = 0.80$, $SE = 0.07$, $p < .001$) and racial identity importance ($B = 0.17$, $SE = 0.01$, $p < .001$), providing strong support for this hypothesis. Additionally, germane to RQ1, we included the predictors of race, political affiliation, and their interaction, to test the influence of these variables on support.

The results of this test were again significant, F (4, 1195) = 68.32, $p < .001$, Adj.$R^2 = .18$, as were all three predictors, suggesting that both race and political affiliation impact policy support. To further investigate the relationship between participant race and policy support, an independent samples *t*-test revealed that White participants were far more supportive of election reform policy, t (1390) = -9.82, $p < .001$, $d = .54$, than Black participants. Together, these results indicate that attitudes towards election reform broke significantly by race such that White participants reported higher support than Black participants (see Figure 1).

The second RQ inquired whether abstention rates differed by participant race. First, a Chi-Square test revealed significant differences for ballot support based on participants' race, X^2 (2, N = 1394) = 119.68, $p < .001$. Specifically, White participants voted *yes* (n = 715, 85%) more frequently on the measure compared to Black participants (n = 327, 59%), and relatedly, Black participants were more likely to vote *no* (n = 173, 31%) relative to White participants (n =103, 12%). The differences between abstention across Black and White participants (53 [9.6%] versus 23 [2.7%]) again indicated that Black participants abstained at a higher rate. Moreover, as in Study 1, of the Black participants who did abstain, more did so in the Black-cue condition (n = 30, 56.6%) than in the White-cue condition (n = 23, 43.4%).

Finally, we examined the relationship between participants' race, symbolic racism, and policy support using the moderation model in PROCESS (model 1, with the covariates of political affiliation and racial importance). Overall, this multiple regression model was significant, F (5, 1191) = 99.68, $p < .001$, $R^2 = .30$, as was the interaction between race and symbolic racism (B = -0.30, SE = 0.13, $p < .05$). Interestingly, this interaction effect revealed that the relationship between symbolic racism and support for election reform was stronger for Black

participants (B = 1.04, SE = 0.09, p < .001) than White participants (B = 0.74, SE = 0.09, p < .001).

General Discussion

Across two experiments, this investigation set out to understand how issue presentation, and racial identity, impacted support for police and election reform. By conceptualizing racial prejudice through the lens of framing theory, the goal of this work was to understand *how* and *when* racialized thinking impacts policy decisions. To our knowledge, few studies have conceived of prejudicial thinking in ways akin to a framing effect, which is to say how prejudicial thinking can be made more (or less) accessible through issue presentation. To address these ideas, this discussion will present three takeaways that can help communicators understand, and potentially address, the impacts of racialized thinking on public policy.

The first key takeaway is that police reform, at least when labeled as *defund the police*, is racially loaded. Moreover, these racial associations appear to be undermining the success of this policy. We arrive at this conclusion because, first, the *reimagine* frame was more successful than the *defund* frame (H1). Although this perhaps could be due to the valance consistent shift (Levin et al., 1998), this finding, coupled with the result that the White-*reimagine* frame was more popular than the three other conditions combined (H3), reinforces this point. Specifically, this interaction implies that police reform policy is most likely to succeed when it is divorced from both the language of *defund the police* and the racial associations therein. Second, we found that racial identity, and the interaction between race and political affiliation, predicted police reform support, while political affiliation on its own did not. Given the political salience of this issue, the idea that racial identity can wipe out the impacts of affiliation is noteworthy. Furthermore, as we visualize in Figure 1, even Black Republicans supported police reform at levels similar to

Democrats (both Black and White), suggesting that Black audiences – even politically unreceptive audiences – recognize the negative impact current policing has on the Black community, and the need for reform. Third, to account for racialized thinking, we found that one's endorsement of symbolic racism predicted opposition to police reform for White participants (RQ3). Thus, altogether, our work finds that when a policy has been racialized, such that it becomes associated with the Black community, prejudicial thinking can undermine policy support. To combat these effects, repackaging issues of this kind through language or through shifting perspectives about who is likely to be impacted by this policy, may be one communication intervention strategy that lowers the activation potential of harmful beliefs.

To underscore this point using a real-world example, in 2020 in Los Angeles County an estimated 2 million voters approved a measure transitioning $900 million into community services and alternatives rather than incarceration under the title, "Reimagine L.A. County" (reimagine.la., n.d.). Meanwhile, 56 percent of voters rejected a measure using "defund the police" as the aim in Minneapolis, where the George Floyd killing occurred. Although we acknowledge that these examples vary in ways beyond the policy name, our experimental work seems to cohere with reality in these instances (see also Vaughn et al., 2021) and strengthen support for our overarching claim that language can matter.

The second key takeaway was how similar, and different, the findings were between Study 1 and Study 2. For context, we chose the topic of election reform because although this issue has been firmly linked to racial disparities (Valentino & Neuner, 2017), the public – and particularly the White public – may not be as conditioned to think about election reform along racial lines. Our pilot test supported this assumption in finding that participants perceived this issue as less racialized than police reform (see supplement), though notably scores were still

above the midpoint on this scale. We reasoned that an issue with fewer racial associations may offer a permission structure that allows race-related beliefs to play a stronger role in policy opinions. Indeed, our results for Study 2 revealed that race, racial importance, and symbolic racism were reliable and robust predictors of support for election reform across all experimental conditions. Specifically, White participants were more supportive of these policies than Black participants, with Black Democrats being *least* supportive of these reforms. We also found a stronger effect of the race frame than in Study 1. When the (highly influential) race covariates were removed from the regression model germane to H2, we did observe a significant effect of the race frame condition on policy support such that support was higher in the White cue condition. Moreover, White participants were more likely to support election reform than Black participants (RQ1). This suggests that Black participants were more likely to recognize the racial implications of this policy and oppose the policy accordingly. Without overt associations with race, however, White participants were more oblivious to these potential issues. This suggests that, when communicating about policy, there are times when explicit communication about the implications of these policies on communities of color is necessary.

The third takeaway is that, across both studies, Black participants abstained more than White participants (RQ2). Although the rates of abstention were low overall, Black participants were three times more likely to abstain. This type of response could be indicative of racial battle fatigue (Smith et al., 2007). Specifically, Black participants could have recognized that this study was assessing racial attitudes and, rather than report an opinion, opted to abstain due to feelings of exhaustion associated with being asked for an opinion on racially sensitive issues. Although this is speculative, future empirical research could benefit from insights from the racial battle fatigue literature. Currently, research in this area (e.g., Smith et al., 2007; Winters, 2020) finds

that minorities score higher on racial battle fatigue measures compared to White participants in contexts such as higher education (see Ragland Woods et al., 2021). From a communication perspective, it is important to investigate whether messaging aimed at minority groups, or research interested in minority groups, potentially activates this feeling of fatigue as well. Although it is of course important to continue work that improves racial equity, creating more burden on these communities may be an unintended consequence of these efforts that needs to be more fully understood and considered.

Limitations

Despite some provocative findings, it is important to recognize some limitations with this set of studies that merit future attention. For one, we recognize that our race frame manipulations were only able to skim the surface of how perceptions of race, political ideology, and other policy-relevant considerations affect voting on a given issue. Although we decided to locate our hypothetical ballots in real cities for generalizability purposes, we know that every city varies in its racial and political demographics, as well as historical and present-day issues related to race, policing, and electoral policies. We realize that these considerations may have influenced our outcomes in unmeasured ways. Moreover, we acknowledge that some aspects of our design that intended to manipulate racial perceptions may also impact political considerations as well. For example, the urban/rural divide is often cleaved by both race and political orientation. These factors, along with whether the policy was located in a red or blue state, or a battleground state, could have influenced participants' vote choice in this study as well. Although our studies admittedly sacrificed some internal validity for external validity, the significant findings we did obtain (despite this noise), suggest that further work – and more care towards experimental control - is warranted and necessary to better disentangle these relationships.

Additionally, although we did find important differences between Study 1 and 2, it is important to recognize that, overall, election reform received higher levels of baseline support than police reform. This could suggest that police reform is more controversial than election reform. Although we picked these two topics because they represented different levels of racialization, it is likely that these two issues maintain different levels of controversiality as well. What remains unclear from this investigation is whether issue controversiality might have impacted our ability to move opinions via a framing effect. While we did not account for issue controversiality here, the influence of this factor alongside framing effects should be explored.

Conclusion

Taken together, it is important to place the effects observed across these studies in perspective. In academia, the challenge in studying race is that while many acknowledge that racism exists at institutional levels, finding direct evidence of racism at the individual level, within the parameters of academic research, can be methodologically challenging. In this pair of studies, we strove to address this issue by using various approaches, guided by framing theory (Chong & Druckman, 2007) to understand where, and how, the evocation of race can impact policy support. Our results suggest that activating racial information, particularly for White audiences, can lead to more negative attitudes toward policies that could otherwise help communities of color. Thus, endeavoring to study the impacts of race and racism in response to policy, and striving to find communication solutions where possible, is a difficult but necessary pursuit.

References

Abraham, L., & Appiah, O. (2006). Framing news stories: The role of visual imagery in priming racial stereotypes. *Howard Journal of Communications*, *17*(3), 183–203. https://doi.org/10.1080/10646170600829584

Appiah, O., Knobloch-Westerwick, S., & Alter, S. (2013). Ingroup favoritism and outgroup derogation: Effects of news valence, character race, and recipient race on selective news reading. *Journal of Communication*, *63*(3), 517–534. https://doi.org/10.1111/jcom.12032

Baranauskas, A. J. (2022). Racial resentment, crime concerns, and public attitudes toward defunding the police. *Journal of Ethnicity in Criminal Justice*, *20*(1), 48–72. https://doi.org/10.1080/15377938.2022.2053626

Bos, L., Lecheler, S., Mewafi, M., & Vliegenthart, R. (2016). It's the frame that matters: Immigrant integration and media framing effects in the Netherlands. *International Journal of Intercultural Relations*, *55*, 97–108. https://doi.org/10.1016/j.ijintrel.2016.10.002

Brown, D. K., & Mourão, R. R. (2022). No Reckoning for the Right: How Political Ideology, Protest Tolerance and News Consumption Affect Support Black Lives Matter Protests. *Political Communication*, *39*(6), 737–754. https://doi.org/10.1080/10584609.2022.2121346

Cacciatore, M. A., Scheufele, D. A., & Iyengar, S. (2016). The end of framing as we know it … and the future of media effects. *Mass Communication and Society*, *19*(1), 7–23. https://doi.org/10.1080/15205436.2015.1068811

Chong, D., & Druckman, J. N. (2007). Framing theory. *Annual Review of Political Science*, *10*(1), 103–126. https://doi.org/10.1146/annurev.polisci.10.072805.103054

de Vreese, C., & Boomgaarden, H. (2003). Valenced news frames and public support for the EU. *Communications*, *28*, 361–381. https://www.degruyter.com/document/doi/10.1515/comm.2003.024/html

Dixon, T. L. (2006). Psychological reactions to crime news portrayals of Black criminals: Understanding the moderating roles of prior news viewing and stereotype endorsement. *Communication Monographs*, *73*(2), 162–187. https://doi.org/10.1080/03637750600690643

Dixon, T. L. (2017). Good Guys Are Still Always in White? Positive Change and Continued Misrepresentation of Race and Crime on Local Television News. *Communication Research*, *44*(6), 775–792. https://doi.org/10.1177/0093650215579223

Drakulich, K., Wozniak, K. H., Hagan, J., & Johnson, D. (2020). Race and policing in the 2016 presidential election: Black lives matter, the police, and dog whistle politics. *Criminology, 58*, 370-402. https://doi.org/10.1111/1745-9125.12239

Dunbar, A., & Hanink, P. A. (2023). Policing protests: an experimental evaluation of the impact of protester race on support for police reform. *Journal of Experimental Criminology*, *19*(2), 343–364. https://doi.org/10.1007/s11292-021-09499-2

Higgins, E. T. (1996). Knowledge activation: Accessibility, applicability, and salience. In E. Tory Higgins & A. W. Kruglanski (Eds.), *Social Psychology: Handbook of Basic Principles* (Issue January 1996, pp. 133–168). The Guilford Press.

Eaglin, J. M. (2021). To "Defund" the Police. *Stanford Law Review Online*, *73*, 120–140. https://perma.cc/N7VX-LLYV

Fazio, R. H., Powell, M. C., & Williams, C. J. (1989). The role of attitude accessibility in the attitude-to-behavior process. *Journal of Consumer Research*, *16*(3), 280. https://doi.org/10.1086/209214

Gilens, M. (1999). *Why Americans hate welfare: Race, media, and the politics of antipoverty society*. University of Chicago Press.

Harlow, S., & Kilgo, D. K. (2021). Protest News and Facebook Engagement: How the Hierarchy of Social Struggle Is Rebuilt on Social Media. *Journalism and Mass Communication Quarterly*, *98*(3), 665–691. https://doi.org/10.1177/10776990211017243

Hayes, A. F. (2018). *PROCESS (3.4.1)*. Software.

Holt, L. F. (2018). Using the Elaboration Likelihood Model to Explain to Whom "#Black Lives Matter" … And to Whom it Does Not. *Journalism Practice*, *12*(2), 146–161. https://doi.org/10.1080/17512786.2017.1370974

Holt, L. F., Ellithorpe, M. E., & Ralston, R. (2017). So Why Do You Think That Way?: Examining the Role Implicit Attitudes and Motivation Play in Audience's Perception of a Racially Charged Issue. *Media Psychology*, *20*(4), 584–606. https://doi.org/10.1080/15213269.2016.1227267

Holt, L. F., & Sweitzer, M. D. (2020). More than a black and white issue: ethnic identity, social dominance orientation, and support for the Black Lives Matter movement. *Self and Identity*, *19*(1), 16–31. https://doi.org/10.1080/15298868.2018.1524788

Isom, D. A., Boehme, H. M., Cann, D., & Wilson, A. (2022). The White Right: A Gendered Look at the Links Between "Victim" Ideology and Anti-Black Lives Matter Sentiments in the Era of Trump. *Critical Sociology*, *48*(3), 475–500. https://doi.org/10.1177/08969205211020396

Iyengar, S. (1990). Framing responsibility for political issues: The case of poverty. *Political Behavior*, *12*(1), 19–40. https://link.springer.com/content/pdf/10.1007/BF00992330.pdf

Jackson, J., Fine, A., Bradford, B., & Trinkner, R. (2023). Social identity and support for defunding the police in the aftermath of George Floyd's murder. *Group Processes and Intergroup Relations*, *26*(4), 833–858. https://doi.org/10.1177/13684302221128230

Kim, S.-H., Kotva, Z. R., Zain, A., Chen, Y., & Rusek Kotva, Z. (2024). Black Lives Matter and partisan media. *Journalism and Media*, *5*, 78–91. https://doi.org/10.3390/journalmedia

Lane, D. S., Coles, S. M., & Saleem, M. (2019). Solidarity effects in social movement messaging: How cueing dominant group identity can increase movement support. *Human Communication Research*, *45*(1), 1–26. https://doi.org/10.1093/hcr/hqy011

Lecheler, S., & De Vreese, C. H. (2019). News framing effects. In *Angewandte Chemie International Edition, 6(11), 951–952.* Routledge, Taylor & Francis.

Leopold, J., & Bell, M. P. (2017). News media and the racialization of protest: An analysis of Black Lives Matter articles. *Equality, Diversity and Inclusion*, *36*(8), 720–735. https://doi.org/10.1108/EDI-01-2017-0010

Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes*, *76*(2), 149–188. https://doi.org/10.1006/obhd.1998.2804

Mils, C. E. (2017). Framing ferguson: Fox news and the construction of US racism. *Race and Class*, *58*(4), 39–56. https://doi.org/10.1177/0306396816685030

Monnat, S. M. (2010). Toward a critical understanding of gendered color-blind racism within the u.s. welfare institution. *Journal of Black Studies*, *40*(4), 637–652. https://doi.org/10.1177/0021934708317739

O'Keefe, D. J., & Jensen, J. D. (2009). The relative persuasiveness of gain-framed and loss-framed messages for encouraging disease detection behaviors: A meta-analytic review.

Journal of Communication, *59*(2), 296–316. https://doi.org/10.1111/j.1460-2466.2009.01417.x

Pickett, J. T., Graham, A., & Cullen, F. T. (2022). The American racial divide in fear of the police. *Criminology*, *60*(2), 291–320. https://doi.org/10.1111/1745-9125.12298

Pratto, F., Korchmaros, J. D., & Hegarty, P. (2007). When race and gender go without saying. *Social Cognition, 25*(2), 221-247. https://doi.org/10.1521/soco.2007.25.2.221

Provine, D. M. (2008). *Unequal under law: Race in the war on drugs*. University of Chicago Press.

Reese, S. D., & Shoemaker, P. J. (2016). A media sociology for the networked public sphere: The hierarchy of influences model. *Mass Communication and Society*, *19*(4), 389–410. https://doi.org/10.1080/15205436.2016.1174268

Reimagine L.A. County (n.d.) RE LA. https://reimagine.la/

Scheufele, D. A., & Iyengar, S. (2017). The state of framing research: A call for new directions. In K. Kenski & K. H. Jamieson (Eds.), *The Oxford handbook of political communication theories* (pp. 1–27). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199793471.013.47

Shulman, H. C., & Sweitzer, M. D. (2018). Advancing framing theory: Designing an equivalency frame to improve political information processing. *Human Communication Research*, *44*(2), 155–175. https://doi.org/10.1093/hcr/hqx006

Smith, W. A., Allen, W. R., & Danley, L. L. (2007). "Assume the position...you fit the description": Psychosocial experiences and racial battle fatigue among African American male college students. *American Behavioral Scientist*, *51*(4), 551–579. https://doi.org/10.1177/0002764207307742

Tarman, C., & Sears, D. O. (2005). The conceptualization and measurement of symbolic racism. *Journal of Politics*, *67*(3), 731–761. https://doi.org/10.1111/j.1468-2508.2005.00337.x

Tewksbury, D. (2020). Accessibility-applicability model. In *International Encyclopedia of Media Psychology* (pp. 1–14). John Wiley & Sons. https://doi.org/10.1002/9781119011071

Valentino, N. A., & Neuner, F. G. (2017). Why the Sky Didn't Fall: Mobilizing Anger in Reaction to Voter ID Laws. *Political Psychology*, *38*(2), 331–350. https://doi.org/10.1111/pops.12332

Vaughn, P. E., Peyton, K., & Huber, G. A. (2022). Mass support for proposals to reshape policing depends on the implications for crime and safety. *Criminology and Public Policy*, *21*(1), 125–146. https://doi.org/10.1111/1745-9133.12572

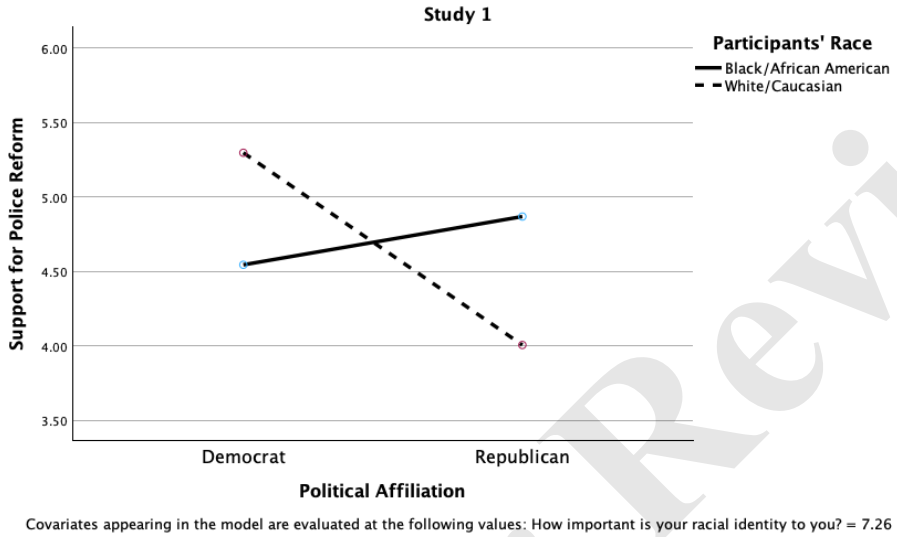
Walter, D., & Ophir, Y. (2024). Meta-theorizing framing in communication research (1992–2022): toward academic silos or professionalized specialization? *Journal of Communication*. https://doi.org/10.1093/joc/jqad043

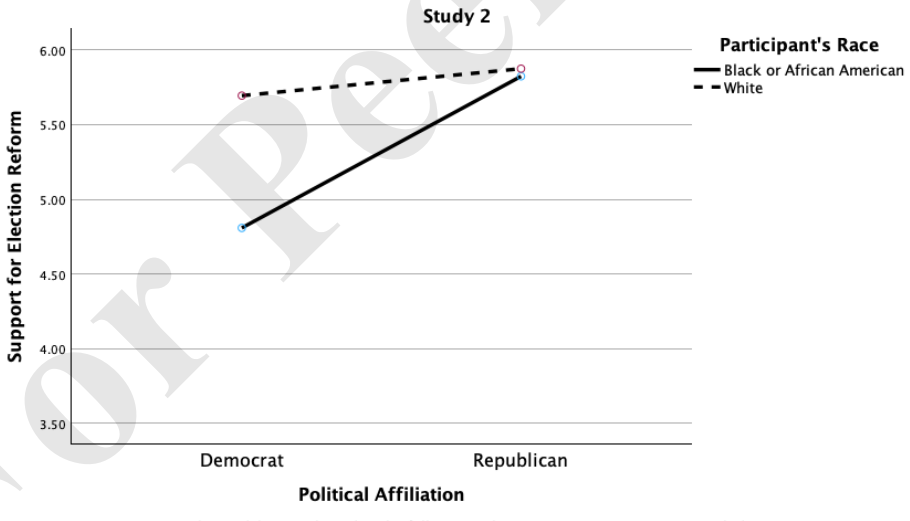
White, I. K. (2007). When race matters and when it doesn't: Racial group differences in response to racial cues. *American Political Science Review*, *101*(2), 339–354. https://doi.org/10.1057/9780230114340

Winters, M. (2020). *Black fatigue: How racism erodes the mind, body, and spirit*. Berrett-Koehler Publishers.

Figure 1

The Relationship Between Participant's Race, Party Affiliation, and Policy Support





Political Communication

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/upcp20

Predicting Vote Choice and Election Outcomes from Ballot Wording: The Role of Processing Fluency in Low Information Direct Democracy Elections

Hillary C. Shulman, Matthew D. Sweitzer, Olivia M. Bullock, Jason C. Coronel, Robert M. Bond & Shannon Poulsen

To cite this article: Hillary C. Shulman, Matthew D. Sweitzer, Olivia M. Bullock, Jason C. Coronel, Robert M. Bond & Shannon Poulsen (2022) Predicting Vote Choice and Election Outcomes from Ballot Wording: The Role of Processing Fluency in Low Information Direct Democracy Elections, Political Communication, 39:5, 652-673, DOI: 10.1080/10584609.2022.2092920

To link to this article: https://doi.org/10.1080/10584609.2022.2092920

POLITICAL COMMUNICATION
2022, VOL. 39, NO. 5, 652–673
https://doi.org/10.1080/10584609.2022.2092920

Routledge
Taylor & Francis Group

Predicting Vote Choice and Election Outcomes from Ballot Wording: The Role of Processing Fluency in Low Information Direct Democracy Elections

Hillary C. Shulman[a], Matthew D. Sweitzer[a], Olivia M. Bullock[b], Jason C. Coronel[a], Robert M. Bond[a], and Shannon Poulsen[a]

[a]School of Communication, Ohio State University, Columbus, Ohio, USA; [b]Department of Organizational Sciences and Communication, George Washington University, Washington, D.C USA

ABSTRACT

Two laboratory studies (N = 240) were designed to explain and predict how people make decisions in low-information political environments. Guided by feelings-as-information theory, it was argued that when direct democracy ballot issues do not receive any campaign expenditures and are not about moral/civic issues, voters are likely to encounter these ballots for the first time in the voting booth. And when this is the case, how these ballots are written should affect vote choice. In support of study hypotheses, it was found that the difficulty of the words on the ballot affected people's processing fluency, defined as the ease with which people processed the information presented. In turn, self-reports of processing fluency influenced vote choice. Specifically, easier texts were more likely to be supported and difficult texts were more likely to be opposed or abstained from voting on. As hypothesized, this relationship was mediated through self-reports of processing fluency. Additionally, to demonstrate the external validity of this process, it was found that the voting results obtained in the two laboratory studies replicated real-world election results 86% of the time. These results offer communicative and psychological insight into how communication affects information processing, and how these processing experiences inform political decisions of consequence to everyday life.

KEYWORDS

Direct democracy; ballot wording; information processing; decision aids; voting behavior

How voters arrive at decisions when they have little to no information about the decision at hand is a critical question for democracy. Indeed, a substantial amount of research in communication, political science, and beyond (e.g., Berinsky et al., 2020; Bowler, 2015; Downs, 1957; Lupia, 1994; Nicholson, 2005) has examined how under conditions of low information people use decision aids, such as partisan cues, endorsements, or the physical attributes of candidates, to guide vote choice. Notably, this line of research has helped researchers to better understand the power of partisanship as a social identity (Green et al., 2002) and the (ir)rationality of various voter decisions (e.g., Nai, 2015). Despite these advancements, the context of direct democracy is a different sort of instance where voters are asked to make decisions on policies that are often complicated, nonpartisan, and receive little media attention. As shown by Burnett (2019), voter knowledge of ballots about

CONTACT Hillary C. Shulman shulman.36@osu.edu School of Communication, Ohio State University, Columbus, OH
Supplemental data for this article can be accessed on the publisher's website at https://doi.org/10.1080/10584609.2022.2092920

nonsocial issues, that also tend to receive little media attention, lags behind all other forms of political knowledge. Thus, focusing on ballots that do not concern social/moral issues, and do not receive campaign expenditures (a proxy for media attention), offers the opportunity to study political decision-making in a very low, yet very common, information environment. We presume, and empirically test, our claim that in these low information environments, voters are first exposed to these ballots in the voting booth and, as such, the way these ballots are written should affect vote choice. Thus, identifying how ballot wording affects voting decisions is an important democratic pursuit and interesting communication inquiry.

Guided by the literature in metacognition (Petty et al., 2007; Shulman & Bullock, 2019) and feelings-as-information theory (Schwarz, 2011), we find that the metacognitive feelings associated with processing difficulty predict how people vote on ballots. We also offer support for the external validity of this prediction by demonstrating across two lab studies using registered voters, and 64 ballots, that the decisions arrived at in the lab correspond with past and future election outcomes. Taken together, the goals of this work are to offer communicative and psychological insight into how word choice affects information processing and how these processing experiences inform decisions of consequence to everyday life.

Low Information Direct Democracy Elections

Ballot referenda, also known as initiatives, measures, propositions, and amendments, are a form of democracy that allows voters to vote directly on a piece of policy or legislation. If the measure is passed, it goes into law, and if it fails, it doesn't. Direct democracy efforts are widespread: currently, in the U.S., all 50 states allow for some form of ballot measures (e.g., popular and/or legislative referendum; Bowler et al., 2020), as do 113 countries around the world (Ballotpedia, n.d.-a). Despite the proliferation of direct democracy voting, there is substantial variation in the ballot topics, media attention, and campaign spending directed toward these ballots. This variability presents a challenge for making broad inferences about how people engage with these sorts of decisions. Understandably, much of the existing work on direct democracy has focused on issues that have been able to capture media attention, citizen interest, deliberation (Suiter & Reidy, 2020), and noteworthy political endorsements (e.g., same-sex marriage, Shi, 2016). In other words, much of the research in this area has focused on ballots in which ballot-relevant attitudes are expected to exist prior to Election Day (e.g., climate change or taxing corporations, see, Goldberg & Carmichael, 2017). Consistent with this expectation, research in this vein has convincingly demonstrated that, much like high information elections, the partisan framing of ballot issues and/or campaign spending in support of, or in opposition of, various ballots are strongly predictive of election outcomes (Bowler, 2015; Bowler & Donovan, 1998; Branton, 2003; Burnett, 2019; Burnett & Kogan, 2015; Damore & Nicholson, 2014; Lupia, 1994; Nicholson, 2003, 2005). What is less understood, however, is how voters make decisions when they are likely to possess little to *no* ballot-relevant attitudes prior to entering the voting booth (e.g., non-attitudes, Converse, 2006). Here, we define contexts in which people are likely to hold weak or non-attitudes about the decision under consideration as *low information environments*. One common and reoccurring low information political environment is *low-salience* ballot voting. Here, low-salience ballots are defined as ballots that receive zero dollars in campaign expenditures (and thus receive little to no media attention), are *not* about social/moral

issues, nor are obviously partisan (Damore & Nicholson, 2014; Lupia, 1994). By focusing on how voters make decisions without the more commonly studied set of voting decision aids, new insights regarding political decision-making becomes possible.

Before reviewing research on direct democracy voting in general, our decision to focus on low-salience ballots merits investigation because, practically speaking, these types of initiatives are quite common. According to Ballotpedia (n.d.-a), a nonpartisan digital encyclopedia of American elections, in 2018, of the 167 state-certified ballots, 30.5% classify as low-salience under our criteria. And, if local ballots were included in this count, this percentage would be much higher as many local issues surrounding bonds, sales tax rates, and local health ballots, for example, would qualify as low salience. Thus, given the prevalence of this low-information environment, an opportunity arises for communication scholars to offer insight into how word choice informs a common, and consequential, type of political behavior.

Research on direct democracy generally finds that the media environment, ballot topics, and citizen deliberation substantially affect voter awareness and subsequent vote choice (e.g., Carlin & Carlin, 1989; Suiter & Reidy, 2020). Nicholson's (2003) work, for example, finds that campaign expenditures can increase awareness of a ballot measure by an average of 16%. This same work also finds that social/moral issues, on average, have awareness rates that are 18% higher than nonsocial issues (defined as revenues and taxes, environment and resources, business regulation, education, welfare, and health). Furthermore, Burnett's (2019) study similarly found that voter knowledge of ballot initiative facts was substantially higher for moral issues and ballots that were well-financed and, accordingly, better publicized. To better contextualize these results, foundational work from Lupia (1994) argued that voters use "shortcuts" including partisan cues, past behavior, and interest group campaigns, as a way to approximate real preferences when voting. Thus, taken together, this literature reveals that the type and amount of information voters hold about these ballots predictably impacts vote choice (see, Morisi et al., 2021). What is less understood, however, is how voters make decisions when the typical slate of informational cues is unavailable. As such, the goal of this study is to uncover a less considered source of influence within these environments: word difficulty.

Although research in direct democracy does not typically focus on low-salience ballots, there is some work that supports the claim that ballot wording could be a decision aid that guides vote choice under these conditions. Burnett and Kogan's (2015) survey experiment, for example, found that although the way ballots were framed initially influenced vote choice in favor of the frame, the presence of additional information – meant to emulate media attention in the real world – mitigated any framing effects. In addition to research on media effects, there has also been work examining the role of ballot language difficulty, given that many ballots contain words that are legalistic and likely unfamiliar to the average voter (Bowler & Donovan, 1998; Goldberg & Carmichael, 2017; Milita, 2015; Reilly & Richey, 2011; Shockley & Fairdosi, 2015). This research has found that language difficulty affects aggregate behavior, such as the likelihood of referenda passing (Milita, 2015), and individual behavior, such as voter roll-off and abstention (Reilly & Richey, 2011). Thus, taken together, experimental and cross-sectional work has shown that a) as media attention decreases, the influence of ballot wording increases, and b) ballot language difficulty can affect vote choice.

Though these findings inform the current investigation, to our knowledge no studies have examined these processes using real ballot measures and deploying methods that could isolate the communicative and psychological explanations that underlie these relationships. As such, the purpose of this investigation is to address these gaps in the literature and enhance our theoretical understanding of political decision-making in the process.

Word Difficulty and Processing Fluency

To enhance our understanding of political decision-making in these contexts, we set up a lab-based study that could test a psychological explanation for *why* vote choice is affected by word difficulty using metacognition. To theorize about this relationship, it is important to distinguish between two kinds of cognitions. The first, called *primary cognition*, refers to the declarative information and attitudes a person has in their memory (Petty et al., 2007; Schwarz, 2015; Shulman & Bullock, 2019). Primary cognitions affect our decision-making. When a ballot measure is introduced that has clear social or moral implications, people are able to access their existing value systems and decide whether to vote in support of the measure. So, for example, decision-making about ballots related to same-sex marriage should lead people to access primary cognitions such as religious beliefs, beliefs about civil rights, self-interests, etc. Declarative-based models of information processing state that people render a judgment based on their global assessment of their accessed primary cognitions. Although primary cognitions, or declarative-based models, are frequently used to explain why individuals arrive at particular decisions, in a low information environment, the availability of useful declarative information should be lacking. Under these circumstances, experience-based models may be more influential. Experience-based models are composed of *secondary cognitions*, also known as *metacognition* (Petty et al., 2007; Schwarz, 2015).

Metacognition refers to the thoughts people have about their thoughts or thought process (Schwarz, 2015). In other words, metacognition reflects how people feel about their information processing *experience*. Although there are many forms of metacognitive feelings, including emotions and moods while processing information (Schwarz, 2015), here we operationalize metacognition through the concept of *processing fluency*. Processing fluency can be defined as how easy or difficult information processing is experienced (Schwarz, 2015), or, put differently, how easy or difficult it feels to access primary cognitions. To convey how processing fluency is experienced, imagine being asked a difficult question. The feelings associated with something being hard (or easy) to understand and respond to reflects processing fluency.

Difficult or complex language has been reliably shown to affect processing fluency, such that as language gets more difficult, in terms of syntax, semantics, and/or word typicality (e.g., jargon), people report less fluent processing (Goldberg & Carmichael, 2017; Markowitz & Shulman, 2021; Shockley & Fairdosi, 2015; Shulman et al., 2020; Tolochko et al., 2019). For example, experimental work on the impact of scientific jargon revealed that the presentation of more specialized and less frequently used words produced lower processing fluency than when these words were replaced with more common or frequently used terms (Shulman et al., 2020). Within the political realm, three different papers by Sweitzer and Shulman (2018, Shulman & Sweitzer (2018b), 2018a) demonstrated that public opinion questions written with easier words produced higher reports of processing ease

than the difficult version of these questions. And finally, Shockley and Fairdosi's (2015) experimental work applied these relationships to ballot referenda by randomly assigning participants to evaluate a complex or simple version of a hypothetical ballot measure. Consistent with expectations, those in the complex condition reported lower levels of processing ease than those in the simple condition (see also, Goldberg & Carmichael, 2017). In this study, we aim to complement and advance existing research by replicating the results from these experiments using unmanipulated ballot measures. Specifically, given that the presence of difficult, uncommon, or unfamiliar words (Milita, 2015) is both common within ballot measures and also reliably shown to affect processing fluency, an opportunity is presented to assess whether the known association between word difficulty and processing fluency generalizes to this consequential, real-world context.

H1: The presence of easier words on ballots will be positively associated with processing fluency.

Feelings-as-Information Theory and Vote Choice

In a low information environment, the declarative content voters can draw upon should be limited. This is where experience-based models, and specifically processing fluency experiences, might offer a more useful explanation for how people make decisions in these circumstances. To this end, feelings-as-information theory (FIT; Schwarz, 2011) can guide predictions for how processing fluency experiences will influence whether people support, oppose, or abstain from voting on low-salience ballots.

The first proposition of FIT states that people use metacognitive information in the same way as declarative information and that "different types of [metacognitive] feelings provide different types of information" (Schwarz, 2011, p. 32). The second postulate of FIT states that, "People usually experience their feelings as being 'about' whatever is in the focus of attention; this fosters the perception that incidental feelings are relevant" (Schwarz, 2011, p. 32). Together, these postulates predict that processing fluency, induced through the presence of common or uncommon words, will be attributed toward the ballot at hand. As such, existing research about the outcomes produced by a (dis)fluent processing experience should be informative.

One of the most reliable findings in the literature on processing fluency is that easier processing feels good and a difficult experience feels bad (for reviews see, Petty et al., 2007; Schwarz, 2015). Guided by proposition two of FIT, people should misattribute these positive or negative feelings toward the subject of one's attention (Schwarz, 2011). To illustrate this Shulman and Sweitzer (2018a, 2018b) observed that participants randomly assigned to public opinion questions written with simpler language reported more interest, knowledge, and efficacy in politics than those randomly assigned to the difficult language condition. For the current study, theory and evidence suggests that language difficulty should affect one's processing experience, and the feelings compelled by this experience should be directed toward one's vote choice in the direction consistent with the hedonic

nature of processing fluency. Thus, one should expect that positive feelings, provoked by an easy experience, should engender support and negative feelings, provoked by a difficult experience, should lead to opposition (see also, Bowler & Donovan, 1998).

There is, however, an alternative way to interpret predictions guided by FIT. The second proposition of FIT further specifies that processing fluency provides information about how informed people are about the subject at hand (Schwarz, 2011). Specifically, an easy processing experience augments, and a difficult processing experience discounts, the perceived value of information. The perceived value of information refers to the perceived utility of the declarative information people are able to retrieve about a topic. An easier experience leads people to assume, with more confidence, that they possess adequate knowledge to proffer a decision (Schwarz, 2011). This is because an easy experience augments one's perceived understanding of the issue. Conversely, a difficult experience leads to the discounting of information. When this is the case, a person may conclude that they do not possess enough quality information on the topic (Schwarz, 2011), and should be more likely to abstain because they *don't know* which way to vote (Bowler & Donovan, 1998; Sweitzer & Shulman, 2018). Notably, this would lead to a slightly different expectation than the easy experience = support and difficult experience = oppose outcome. This work strives to offer a first step toward understanding the relationship between vote choice under conditions of easy, moderate, and difficult processing, and whether any of these conditions is more likely to result in a decision to *abstain*. Specifically, we assess whether an abstain vote better reflects a state between two affective poles (support and oppose) in ways similar to ambivalence or neutrality (H2, Bowler & Donovan, 1998; Song & Ewoldsen, 2015; Sweitzer & Shulman, 2018), or the feeling that one is not knowledgeable enough to vote (RQ1). Thus, because we are unsure whether low levels of processing fluency will result in a vote to oppose (H2), or a decision to abstain, the following hypothesis and research question is advanced:

H2: The relationship between word difficulty and vote choice will be mediated by processing fluency, such that an increase in the presence of easier words will lead to greater processing fluency which, in turn, will increase support for the ballot measure.

RQ1: What is the relationship between levels of processing fluency and outcomes of opposition and abstention?

Method

Participants

Participants for this study were 120 registered voters from Ohio.[1] Participants were recruited through ResearchMatch, a volunteer service sponsored by the National Institutes of Health, and their voter registration was verified using publicly available voter records (60 females; Age M = 34.99, SD = 16.19, Range = 18–79; Race: White = 97, Black = 9, Latina/o/Hispanic = 3, Asian = 4, Mixed = 6, Other = 1; Partisan Affiliation: Democrat = 71, Independent = 37, Republican = 12). All subjects were compensated $40 for taking part in this study.

Materials

The stimuli for this study consisted of 40 statewide ballots drawn from 21 states (see supplementary materials for examples). Several ballots were used in accordance with a message-sampling approach (see, Slater et al., 2015)[2] to better ensure that the scope of ballot topics (14 different topics were represented by this sample), and the range of word frequency scores, used in this sample was generalizable to the population of real low-salience ballots.

The 40 ballots chosen were selected from a larger data set of all statewide ballots (excluding ballots from Ohio) that were voted on during the 2012 general (n = 22) and 2014 midterm (n = 18) elections. These years were chosen to increase the likelihood that participants had not seen, remembered, nor previously voted on, these ballots. All stimuli were sourced from state boards of elections, the National Conference of State Legislatures (a nonpartisan advocacy organization), and Ballotpedia. This initial sampling frame included 307 statewide ballots from 43 states and was reduced to 40 ballots using the sampling procedure described below. All of the selected ballots are provided in the online supplement.

Stimuli Selection

The two goals of this sampling procedure were to (1) reduce the data set to only low salience ballots, operationalized as ballots that received zero dollars in campaign spending, and (2) to maximize variance in word difficulty.[3] Of the 307 ballots collected, 191 received zero dollars (62.21%), thus qualifying as low-salience issues. With the remaining ballots, we removed any ballots with a word count greater than 125 words (n = 48) to ensure that the full text of the measure could be shown on a single screen during the lab study procedure. This left 143 ballots that were included in the next stage of stimuli selection.

Word Difficulty

Although we acknowledge that the latent construct of language difficulty can be operationalized in a variety of ways, we opted to use word frequency measures, which assess how *commonly* or *frequently* a word is used in the English language using the 2012 Google Ngram English fiction corpus (Google, 2013; see also, Michel et al., 2011). For this study, we decided to use the Google Ngram measure for three reasons. The first was that recent work (e.g., Benoit et al., 2019b; Richey & Taylor, 2020) advocates for the content validity of this measure over other commonly used metrics (e.g., Flesch-based measures of reading ease, see, Benoit et al., 2019b). Second, given that we were interested in information processing experiences resulting from written, rather than spoken, word, we used a measure obtained from a sample of books rather than transcriptions from voice recordings. Finally, other work examining the effects of language difficulty (Milita, 2015; Shulman et al., 2020) has found that processing difficulty arises out of the usage of unfamiliar and technical terms (i.e., semantic difficulty) as opposed to syntactical difficulty (as indexed in Flesch-based measures, see, Tolochko et al., 2019). Thus, using a measure that assesses word frequency – as a proxy for word familiarity – is a valid, contemporary, and likely to be influential measure of language difficulty.[4]

A word's Ngram score is a count of the number of times that word appears in Google's entire corpus of electronic books. Words with lower Ngram scores appear less frequently in English literature and should thus be less familiar to readers than words with higher scores (see, Benoit et al., 2019b). To produce an Ngram score for each ballot, we calculated the median Ngram score for the title and text of each remaining ballot. To maximize variance in this independent variable, the 20 "easy" ballots used in this study had an Ngram score above the median Ngram score, whereas the 20 "difficult" ballots selected had Ngram scores that were below the median. We then recalculated the Ngram scores for each ballot after removing counts for stop words, such as "the" or "of," which have outlying Ngram scores, and which also contribute little to the meaning of the ballot.[5] The average Ngram score, rescaled to the order of millions of uses, for each ballot was used in the analyses that follow. Overall, *Average Ngram* scores varied considerably in our stimuli: $M = 2.00$ million uses, $SD = 1.16$, range = 0.83–5.59 (all ballots used in this investigation are available in supplementary materials).

Pretest Survey

Because the stimuli for this study were actual ballots whose content varied in ways beyond word difficulty, we conducted a pretest to assess these differences. To do so, we administered a survey to a separate set of 101 registered voters from Ohio , also recruited through ResearchMatch and compensated $10 for their participation. Findings revealed that easy and difficult ballots differed significantly from one another for perceived familiarity, F (39, 1877) = 8.04, $p < .001$, interest, F (39, 1874) = 9.50, $p < .001$, and importance, F (39, 1877) = 8.34, $p < .001$. Given that the independent variable, word frequency, covaried with these ballots, we opted to control for the effects of familiarity ($M = 2.30$, $SD = 0.54$), interest ($M = 3.22$, $SD = 0.53$), and importance ($M = 3.60$, $SD = 0.49$) by including the ballot-level means from this pretest as covariates for hypothesis testing.[6]

Procedure and Measures

This study employed a repeated measures research design with word frequency serving as the independent variable. Thus, each participant was exposed to all 40 ballots that varied from easy to difficult in the ways discussed above. The order in which each ballot appeared was randomized, thus controlling for ballot order effects on the aggregate. Participants in the in-person lab study were tested individually in a quiet room placed in front of a computer screen. Participants were told at the beginning of the study that they would be reading and voting on ballots that were currently being considered in Ohio to enhance the relevance of the task. Participants were instructed to imagine that they were in a voting booth, to read each initiative carefully, and then vote. Participants were shown ballot measures on the computer screen in front of them. They held a controller that enabled them to advance to the next screen, which was where they made their vote choice. The screen after the ballot measure contained the words "Support," "Oppose," and "Abstain." The order of these *Vote Choices* on the screen and their corresponding button locations were counterbalanced to ensure that observed effects were not due to a specific choice/

button configuration. After reading and voting on all of the ballots, participants were asked to complete a follow-up survey of self-report items. Participants read each ballot measure again and responded to a three-item *Processing Fluency* scale ($M = 4.86$, $SD = 1.57$, $\alpha = .91$; Shulman & Sweitzer, 2018b2018),[7] wherein higher scores reflect an easier processing experience. Participants then reported demographic information. All of our stimuli, measures, and data can be found on our open science framework page.

Study 1 Results

To test H1, H2, and RQ1, the structural equation models illustrated in Figure 1 were used. These mediation models were estimated using MPlus, which permits a number of specification options that pertain to this data. First, because our study uses repeated measures, we elected to cluster standard errors at the subject-level. Second, the dependent variable in tests of H2 and RQ1 (vote choice; oppose, abstain, or support) was treated as an ordinal categorical variable because the differences between abstain and oppose outcomes (and their relative ordering) was not yet known. Third, the mediating variable (processing fluency) was measured using three items; rather than creating a mean index of these items, structural equation models permit the construction of a latent processing fluency variable, allowing for mathematical inferences that reflect the abstraction of psychological processes (e.g., processing fluency).

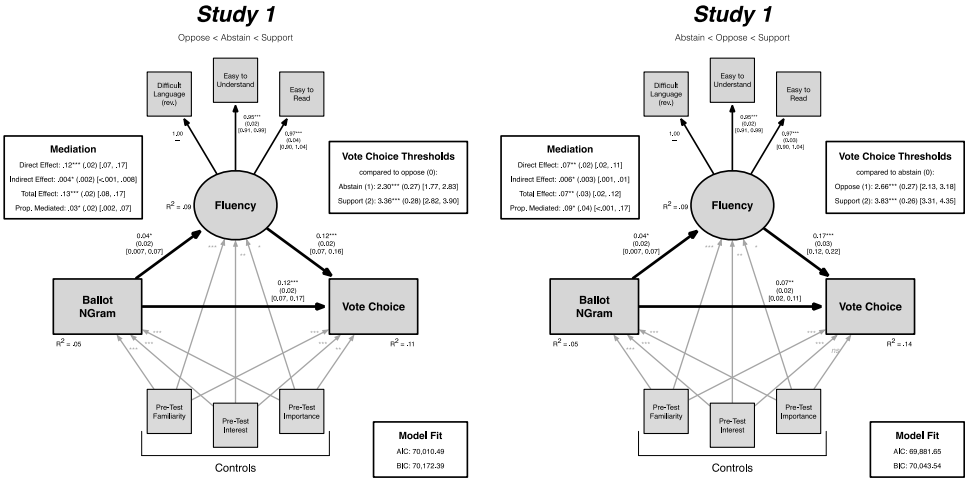


Figure 1. Structural equation models of in-lab voting decisions on ballot measures by average Ngram score and reported processing fluency – Study 1. Vote choice was coded ordinally in both models; the left panel shows the oppose (0), abstain (1), support (2) coding scheme, while the right panel shows the abstain (0), oppose (1), support (2) coding scheme. Estimates, standard errors, confidence intervals, and effect sizes were calculated with maximum likelihood estimation. Standard errors were clustered at the subject-level ($N = 120$). The path between NGram scores and Fluency supports the hypothesized relationship from H1 and that the indirect path between NGram scores and Vote Choice is consistent with H2 regardless of the ordering of the Vote Choice DV. *Note*: only significance levels of the paths from the control variables to the variables of interest are shown to conserve space. *$p < .05$, **$p < .01$, ***$p < .001$.

An Increase in Word Frequency Is Associated with an Increase in Processing Fluency (H1)

To test H1 – that word frequency and processing fluency would be positively associated – the path coefficient between word frequency and participants' self-reported *Processing Fluency* was estimated. More specifically, this analysis included *Average Ngram* values of the ballots, on the scale of millions of uses, as the independent variable, and mean *Familiarity, Interest*, and *Importance* scores were included as controls. As presented in both models within Figure 1, we found significant evidence in the predicted direction for the positive relationship between word frequency and processing fluency: $B = 0.04$, $SE = 0.02$, $p < .05$, $R^2 = .09$. Thus, when ballots included more frequently used words, self-reports of processing fluency increased as well.

Processing Fluency Mediates the Relationship between Word Frequency and Vote Choice (H2)

To address H2 – that processing fluency would mediate the relationship between word frequency and voting decisions – we estimated the mediation model depicted in the left panel of Figure 1. In this model, the dependent variable *Vote Choice* was coded as follows: support [2], abstain [1], and oppose [0]. Importantly, mediation effects (i.e., indirect effect, total effect, proportion of effects mediated) were estimated in the same structural equation modeling framework in MPlus. As shown in the left-panel of this figure, the indirect effect was significant and positive ($B = .004$, $SE = .002$, $p < .05$, 95% CI = [<.001, .008], $R^2 = .11$), suggesting that word frequency is associated with a vote in support of a ballot as a result of increased processing fluency. The direct effect was also significant and positive in this model, indicating that the effect is not exclusively mediated through processing fluency. In sum, these findings support H2.

Word Frequency Affects Vote Choice Primarily through Processing Fluency at the Level of Support (RQ1)

To address research question one, we estimated the same mediation model as above, but recoded the ordinal dependent variable *Vote Choice* (support [2], oppose [1], and abstain [0]). These results are depicted in the right panel of Figure 2. The results here are substantively similar: the indirect effects were significant and positive among lab self-reports ($B = .006$, $SE = .003$, $p < .05$, 95% CI = [.001, .01], $R^2 = .14$). Moreover, in both models, the confidence intervals of the indirect effects overlap the estimates in the other model. These converging results may indicate that word frequency affects voting decisions through processing fluency at the level of support, but that the decision between abstention and opposition may depend on some as-yet-unmeasured characteristic of the ballot or the voter.

Aggregate Votes in the Lab Predicted Real-World Aggregate Votes

Finally, to understand whether these results replicate, and in turn generalize, to the real-world elections these ballots were chosen from, we ran two analyses. The first provided a global assessment of whether the "election" results from our in-lab study produced the

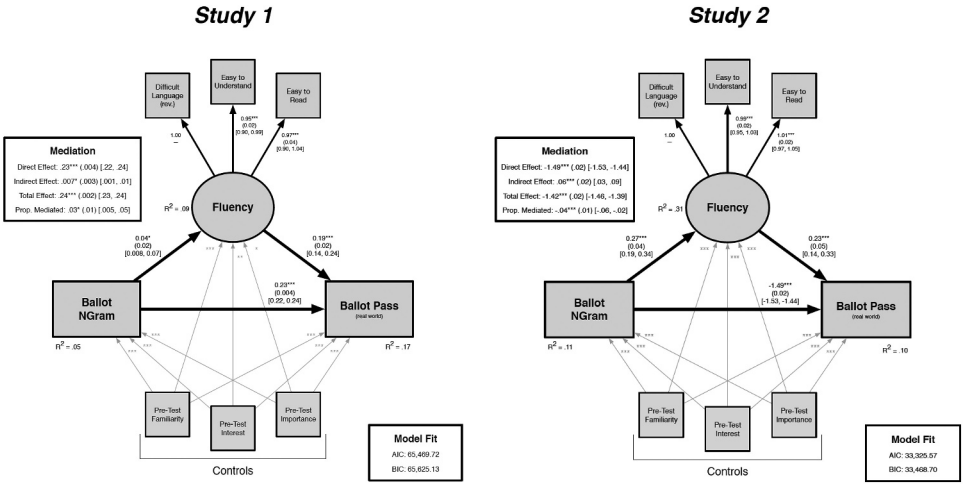


Figure 2. Structural equation modeling results with real-world voting outcomes for Studies 1 (left panel) and 2 (right panel). Processing fluency was measured among lab participants in these models. Estimates shown with standard errors in parentheses and 95% confidence intervals in brackets. As these figures indicate, the indirect path between ballot wording and real-world passage rates, mediated through fluency, was supported across both studies. *Note*: to save space, only the significance level is shown for the paths from the pre-test control variables to the variables of interest. *$p < .05$, **$p < .01$, ***$p < .001$.

same result as the real-world election using the appropriate state-level election rules (Ballotpedia, n.d.-b). To do so, we converted support rates into a pass/fail categorical variable for both in-lab responses and real-world elections.[8] We then conducted a chi-squared test to determine if the observed rates of similar results (i.e., ballots that pass/fail in real-world elections also pass/fail among lab participants) occur above and beyond chance. The results of this test were significant, $\chi^2(1) = 10.21$, $p < .01$. Of the 40 ballots, 34 (85%) had the same result (28 pass, 6 fail) in both the lab "election" and in their real-world election. Just 6 ballots (15%) had different outcomes: 4 passed in the lab but failed in the real-world, and 2 failed in the lab but passed in the real-world. In addition to this analysis, we also estimated the same mediational model used to test H1, H2, and the RQ1 (described in detail below) but replaced laboratory outcomes with real-world ballot passage rates (0: fail, 1: pass) as the dependent variable. This allowed us to test whether a ballot's NGram score and aggregated processing fluency score (obtained using the lab participants), could be used to *explain* ballot passage rates in the real-world election. In sum, the results from this analysis revealed that word frequency, with processing fluency as the mediator, predicted real-world ballot passage rates, $B = .007$, $SE = .003$, $p < .05$, 95% CI = [.001, .01]. Together, these analyses demonstrate the generalizability of these relationships.

Study 1 Discussion

The purpose of study 1 was to understand whether word difficulty could affect vote choice through processing fluency. The results from our laboratory study revealed that ballots written with more frequently used words were also easier to process. And, when ballots were easy to process, participants were more likely to support the measure. Furthermore, we

provide evidence that the processes observed in the lab appear to emulate the processes that occur in real elections. Together, this evidence is both practically important and theoretically compelling.

Theoretically, examining political decision-making within this context offered a broader opportunity to study the decision aids people use to make decisions when little prior, or outside, information is available. Some research on direct democracy has revealed that endorsements (e.g., Lupia, 1994), the presence of scientific evidence (e.g., Stucki et al., 2018), or perceptions of argument quality (e.g., Nai, 2015) affect voting decisions. The findings obtained here contribute to this literature by adding a theoretical explanation for why word difficulty can be another influential decision aid within low-information environments.

Finally, it is noteworthy that the influence of word difficulty on vote choice (through fluency) persisted across 40 different ballots that varied in ways beyond word difficulty. Thus, despite an abundance of alternative explanations for why a person would support or oppose a ballot (that should manifest as error in our design), word difficulty, and the mediating effect of processing fluency, still produced a statistically significant relationship and strong to medium sized effects ($.09 < R^2 < .14$). Given the important practical and theoretical utility of this finding, Study 2 was conducted to replicate this effect.

Study 2

Study 1 provided initial evidence that subjective experiences while processing difficult versus easily worded information informs decision-making. The aims of Study 2 were to try and replicate this effect using different ballots, and to shed more light on the processes that underlie an abstain versus opposition vote choice. Thus, hypotheses one and two, along with research question one were tested again using a different sample of ballots from an upcoming election using a new sample of participants.

Method

The method for Study 2 was identical to Study 1, with the exceptions discussed below.

Participants

Participants for Study 2 were 120 registered voters from Ohio recruited through ResearchMatch. Participants from Study 1 and pretest survey participants were ineligible to participate in Study 2 (60 females; Age M = 33.98, SD = 18, Range = 18–73; Race: White = 103, Black = 9, Asian = 4, Mixed = 3, Other = 1; Partisan Affiliation: Democrat = 61, Independent = 29, Republican = 30).

Materials

The stimuli for this study consisted of 24 statewide ballots from 11 states (see Appendix A for an example). These 24 ballots were selected from a larger data set, sourced by the researchers using the same procedure described in Study 1. In total, there were 142 statewide ballots from 35 states that were available to the public at the time of this data

collection (Summer 2018) and set to be voted on during the 2018 midterm election. These 142 ballots were reduced to 24 in accordance with the sampling procedure described in Study 1. Ballots that received financial support or that contained more than 125 words were removed first, leaving 58 ballots. A median Ngram score was determined for the entire data set and then 12 ballots which had a median Ngram score below the set's median were selected for use as difficult ballots, and 12 ballots whose median Ngram score was above the set's median were selected for use as easy ballots. As in Study 1, we then recalculated *Average Ngram* scores – again, on the scale of millions of uses – after removing stop words. Notably, although the ballots used in Study 2 contained somewhat more obscure words compared to the stimuli in Study 1, $M = 1.68$ million uses, $SD = 0.61$, range = 0.51–2.74, Ngram scores were not significantly different between Studies 1 and 2, $t\ (62) = 1.26$, $p = .21$, $r^2 = .02$.

Pretest

Once again, a pretest survey was used to assess whether these ballots varied in ways beyond word difficulty. To make this assessment, 102 registered voters in Ohio assessed the selected 24 ballots (see supplement). Participants were compensated $10 for their participation. Survey findings revealed that the easy ballots scored significantly higher in perceived familiarity, $F\ (23, 2315) = 17.59$, $p < .001$, interest, $F\ (23, 2317) = 19.13$, $p < .001$, and importance, $F\ (23, 2312) = 16.79$, $p < .001$, than difficult ballots. As such, we elected to include ballot-level means for familiarity ($M = 2.90$, $SD = 0.55$), interest ($M = 3.69$, $SD = 0.51$), and importance ($M = 3.99$, $SD = 0.46$), from our pretest as covariates in hypothesis testing.

Procedure and Processing Fluency

The design and procedures from study 1 were replicated in study 2. Once again, this study used word frequency as the independent variable across 24 ballots. Participants read each ballot on a screen and were asked to make a *Vote Choice* between support, oppose, or abstain. Ballot order and decision position were counterbalanced between subjects. Following voting decisions for all 24 ballots, participants read each ballot again on a separate survey and responded to the same three-item *Processing Fluency* scale ($M = 4.57$, $SD = 1.73$, $\alpha = .94$).

Study 2 Results

An Increase in Word Frequency Is Associated with an Increase in Processing Fluency (H1)

The statistical tests for Study 2 were the same as those from Study 1. To test H1, we estimated a structural equation model in which (among the other modeled paths; see, Figure 3) participants' mean *Processing Fluency* evaluation of each initiative served as the dependent variable, and *Average Ngram* values on the scale of millions of uses served as the independent variable. Familiarity, importance, and interest were included as controls, and a subject identifier was used to cluster standard errors. Consistent with expectations, there

was a significant and positive relationship between word frequency and processing fluency; $B = 0.27$, $SE = 0.04$, $p < .001$, 95% CI = [0.19, 0.34], $R^2 = .31$. These results offer support for H1 and replicate the findings obtained from Study 1.

Processing Fluency Mediates the Relationship between Word Frequency and Vote Choice (H2)

Hypothesis two was tested using the same full structural equation model used in Study 1. Importantly, the dependent variable – *Vote Choice* – was coded ordinally with opposition assigned the lower value (0) relative to abstention (1). The results of this model are shown in the left-panel of Figure 3. The indirect effect was positive and significant, supporting our hypothesis of mediated effects ($B = .06$, $SE = .01$, $p < .001$, 95% CI = [.03, .09], $R^2 = .05$). Interestingly, the direct effect of *Average Ngram* on *Vote Choice* was significant and negative after accounting for the mediated effect through *Processing Fluency*. This could be the biproduct of a few important differences between Study 1 and Study 2. Namely, the indirect effect is much stronger in Study 2 than in Study 1, and this difference should influence the direct effect between word frequency and vote choice. Moreover, ballot word frequency was slightly more difficult in Study 2 than Study 1, which may also account for these differences. Taken together, however, the relationship posited by H2 was supported and replicates the finding from Study 1.

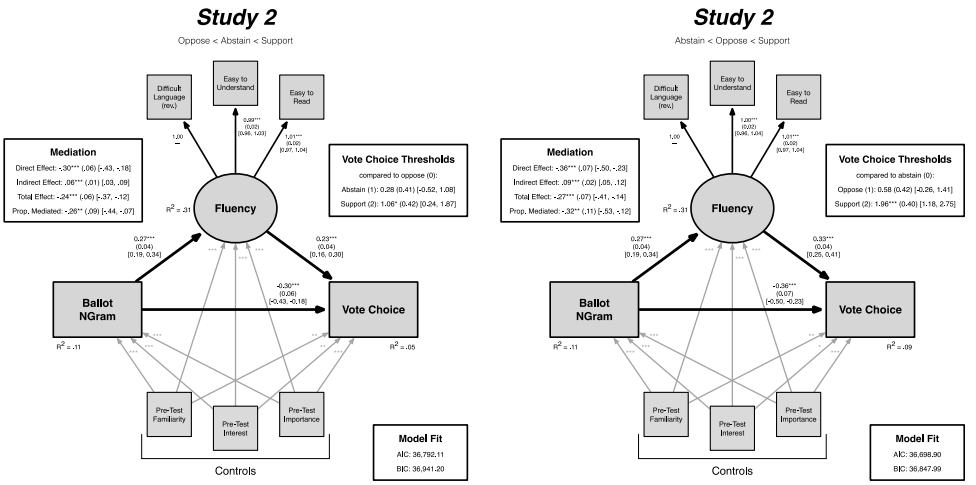


Figure 3. Structural equation models of in-lab voting decisions on ballot measures by average Ngram score and reported processing fluency – Study 2. Vote choice was coded ordinally in both models; the left panel shows the oppose (0), abstain (1), support (2) coding scheme, while the right panel shows the abstain (0), oppose (1), support (2) coding scheme. Estimates, standard errors, confidence intervals, and effect sizes were calculated with maximum likelihood estimation. Standard errors were clustered at the subject-level ($N = 120$). The path between NGram scores and Fluency supports the hypothesized relationship from H1 and that the indirect path between NGram scores and Vote Choice is consistent with H2 regardless of the ordering of the Vote Choice DV. *Note*: only significance levels of the paths from the control variables to the variables of interest are shown to conserve space. *$p < .05$, **$p < .01$, ***$p < .001$.

Word Frequency Affects Vote Choice Primarily through Processing Fluency at the Level of Support (RQ1)

To address RQ1, we again reversed the order of the abstain (0) and oppose (1) *Vote Choices* and estimated the equivalent models to those used in the test of H2. The results are presented in the right panel of Figure 3. Once more, the results of these tests are substantively similar to our tests of H2 and show a significant indirect effect in the expected direction (B = .09, SE = .02, p < .001, 95% CI = [.05, .12], R^2 = .09). In fact, all of the confidence intervals of the estimates in the model for RQ1 overlap with the estimates in the H2 model, indicating no significant differences in the indirect effects between dependent variable specifications. This bolsters the results from Study 1: the effects of word frequency and processing fluency on voting decisions are likely at the level of support and not at the level of abstain versus oppose.

Aggregate Votes in the Lab Predicted Real-World Aggregate Votes

And finally, to understand the generalizability of these relationships we again examined whether the lab results replicate real-world election results by converting voting percentages into a pass/fail categorical measure for both real-world elections and our lab participants. Again, we found that the rates of co-occurring results (e.g., pass/pass and fail/fail) significantly differed from chance: $\chi^2(1) = 5.05$, $p < .05$. Specifically, among the 24 ballots in Study 2, 21 (87.5%) had the same outcome in both the real-world election and in the lab election: 18 passed and 3 failed. Of the 3 ballots (12.5%) which had differing results, 2 passed in the lab and failed in the real-world, and 1 passed in the real-world while failing to pass in the lab. In addition to these analyses, we also estimated a mediational model with real-world ballot passage rates (0: fail, 1: pass) as the dependent variable. The results from this analysis revealed support for the positive indirect effect between word frequency and real-world ballot passage rates, B = .06, SE = .02, p < .001, 95% CI = [.03, .09], with processing fluency as the mediator. These results are presented in full in the right panel of Figure 2. Together, these results support the generalizability of the focal relationships such that the same ballots that tend to pass (fail) in our study also tend to pass (fail) in the real world, and these pass and failure rates can be explained by ballot word difficulty and in-lab reports of processing fluency.

General Discussion

The purpose of these studies was to apply theory in practice and, in doing so, better understand how people make decisions in low information political environments. In addition to providing communicative and psychological explanations for how these types of decisions are made, we also tested these ideas within the context of a consequential and commonly occurring low information environment – direct democracy voting. The results from two carefully controlled laboratory studies, and across 64 ballots, offer compelling evidence for the role of experience-based models of decision-making within these contexts. Notably, we find that feelings of difficulty or ease, evoked by word difficulty, offer incidental information that is attributed toward the ballot being read. Across two studies, we found support for the notion that ballots written with more frequently used words (i.e., easier)

compelled an easier experience that led to a higher likelihood of ballot support. Moreover, the decisions arrived at using this experience-based model produced decisions identical to both past and future elections 86% of the time, despite our use of a small, non-representative sample. These findings provide support for the role of communication and processing fluency in explaining how people make decisions.

Of note, the obtained associations between word difficulty, processing fluency, and voting, were supported across 64 ballots that spanned 11 different issue topics that varied greatly in substance and content. In fact, pretests revealed that participants reported that the substance of these ballots were differentially interesting, familiar, and important. Yet nevertheless, even when controlling for these important informational differences, a consistent pattern of participants relying on their experience was still observed. This pattern supports FIT (Schwarz, 2011) in theoretically important ways. Specifically, the notion that feelings are privileged over declarative information provides a powerful demonstration of the guiding role of our experiences when informational value is discounted. Recall that in FIT the discounting effect occurs when participants discount the utility of their declarative information. Under these conditions, people feel less certain, less confident, and more ambivalent about their opinions (Schwarz, 2011), whereas under conditions of easy processing people are more likely to use their positive affect as information. The results from our experiments support these processes. Namely, under conditions of easy processing, people offered greater support for ballot measures than under difficult processing.

Although this work was tested within the context of direct democracy ballots, it merits mentioning that the processes observed here should generalize to other low-information contexts, political and beyond. We found here that word difficulty evokes a predictable experience: An easy experience promotes a more positive experience, and a difficult experience engenders negativity (all else being equal). When a person is asked to render a judgment (political or otherwise) and does not possess much information about the subject of this judgment, our data (and FIT) propose that people use their information processing experience to guide their decision. There is a rich literature on political decision making aids (see, Downs, 1957; Green et al., 2002; Lupia, 1994) because scholars have long been fascinated with the cues people use to make complicated decisions. Although decision making aids have been well-studied, an examination of the types of cues people rely upon in low-information contexts is not as well understood. Specifically, in high-profile elections, and thus high-information political environments, people rely upon common cues such as endorsements, source cues, partisan cues, and status quo biases (Branton, 2003; Damore & Nicholson, 2014; Lupia, 1994; Morisi et al., 2021; Nicholson, 2005). When these cues are unavailable, however, we found that experience-based feelings, evoked by word frequency, could guide decisions instead. Although this work suggests that word difficulty serves as a cue that guides behavior for low-salience ballots, moving forward, it is important to contextualize the direction and magnitude of this effect in contrast to more well-known processes such as voting behavior in high-salience elections. Given that we only tested low-salience ballots, the degree of difference between these contexts remains unknown and thus merits future investigation.

Another benefit of more direct comparisons between voting behavior under conditions of high versus low information, or strong versus weak preexisting attitudes, is the ability to understand how the findings obtained here comport with related work. Goldberg and Carmichael's (2017) experiment, for instance, similarly examined the role of ballot language

complexity on processing ease and policy favorability. Unlike the current study, however, Goldberg and Carmichael (2017) considered the moderating role of preference-consistency on the relationship between language complexity and policy support. Similar to the results obtained here, Goldberg and Carmichael (2017) found that when participants expressed no preference (e.g., non-attitudes, Converse, 2006), simply worded ballots were more likely to be supported than complex ballots (see also, Bowler & Donovan, 1998). Interestingly, however, when the impact of one's initial policy preferences were taken into consideration, the effect of language complexity and processing ease became more nuanced. Specifically, in the preference-consistent condition it was found that the simply worded version of the ballot measure was more supported than the complex version. In the preference-inconsistent condition, however, the simply worded ballot was rated more negatively than the complex version. Notably, this pattern of results can also be explained by FIT (Schwarz, 2011) which proposes that the value of information (i.e., the ballot information under consideration) is *augmented* under conditions of fluent processing and *discounted* under conditions of disfluent processing. Thus, simpler forms of language can produce more "accurate," or more preference-consistent, decisions in ways similar to other, more commonly studied, political heuristics (see also, Lupia, 1994; Morisi et al., 2021; Shulman & Sweitzer, 2018a).

When integrating these ideas, existing work suggests that in low-information environments the impact of word difficulty seems to predict, through fluency, whether ballots pass or fail, and in high-information environments, word difficulty can positively impact people's ability to vote their preferences (Bowler & Donovan, 1998). Thus, word difficulty can affect political decision-making in a variety of ways, some of which are normatively positive for democracy, some of which might not be. Although work on language complexity as a decision-aid is relatively new, hopefully these ideas inspire future thinking on the role of communication, and more specifically word difficulty, in consequential decision-making environments.

Despite these contributions, there were some limitations and areas that could benefit from further research. First, like all observational studies, endogeneity is a concern that prevents us from having a strong causal interpretation of our results. For example, it is possible that unmeasured ballot characteristics such as an ideological lean (see, Branton, 2003; Damore & Nicholson, 2014), or unmeasured environmental characteristics, such as notable endorsements or general support, could have been driving the relationships under investigation. Although we attempted to minimize these possibilities methodologically, by limiting our sample to low-salience ballots from across the country, and statistically by including control variables reflecting general support, our inability to control for these possibilities reflects the trade-off inherent in our decision to use real, as opposed to hypothetical or manipulated, ballot measures. This decision could reflect why, despite the fact we found support for our mediated voting models as predicted, direct effects were observed as well. This suggests that there are still unmeasured properties of these ballots that account for vote choice beyond word difficulty, processing fluency, familiarity, interest, importance, and party identification and future researchers may wish to use alternative designs to investigate these processes.

Second, as previously mentioned, the antecedents for an oppositional vote and an abstention vote remain unclear. Future research would be well served to better understand what communication and psychological phenomena account for these different behavioral

decisions. Third, we acknowledge that the causal effects we propose remain conjectural due to our lack of a between-subjects experimental method and our methodological decision to have participants offer a vote before reporting on processing fluency. Although the causal relationship we hypothesized was predicated upon theory, our pretest data, and an abundance of experimental work in this domain (for a review see, Shulman & Bullock, 2019), future work should utilize different study designs to offer stronger methodological support for these causal claims. Fourth, these studies relied on only one, semantic-based, measure of language difficulty. Although recent work (e.g., Tolochko et al., 2019) has observed that semantic-based measures, such as Ngram, are more influential than syntactic-based measures (e.g., Flesch-based measures), future work would be benefited by a more robust understanding of how different features of language (including other operationalizations of difficulty) affect the relationship between ballot language and vote choice. And finally, we acknowledge that some of the methodological decisions guiding this study hindered our ability to make broadly generalizable claims. Specifically, our decision to use low-salience ballots impaired our ability to make broader claims about how people make decisions for high-salience ballots. Additionally, we did not take any steps toward ensuring that our lab sample was politically representative of the U.S. as a whole. Still, we found that our lab results mirrored election results even without a representative sample and without modeling influential political variables such as partisanship, political knowledge and participation. Nevertheless, future work should endeavor to replicate and extend our understanding of these processes by using a more politically representative sample and by modeling potentially important individual differences.

In conclusion, this research strove to integrate ideas within the communication, metacognition, and direct democracy literatures to better understand how people behave in low information environments. We found compelling support for the notion that metacognitive feelings guide decision-making within this domain and that communication features, independent of communication content, can guide these experiences. We hope these findings can be used to inspire future work into how communication and processing fluency can explain how people make decisions that affect everyday life.

Notes

1. This study is part of a larger project examining the effect of ballot language on voting decisions. This is the second manuscript from this project. Other data are not reported here and are reported in a separate article (Coronel et al., 2021).
2. The message-sampling approach refers to technique in which researchers use messages sampled from the population of real messages as stimuli for their study (see, Slater et al., 2015). As argued by authors, the goal of this approach is to address generalizability concerns in communication research, while also being able to infer with more confidence how a particular feature of message design impacts audiences using a variety of instantiations of this feature and within a variety of settings. Consistent with this approach, the current study selected a sample of real low-salience ballots from the population of ballot proposals of this kind. Please see the cited work for more information on the utility of this methodological approach.
3. Because the ballots chosen varied by topic (11 subtopics were represented) alongside word difficulty, we ran a chi-square to assess whether our primary variable (word difficulty: easy, hard) covaried with topic in ways that could impact our primary relationship. The results of

this analysis found that there was no significant association between topic category and word difficulty (Study 1: F (6, 33) = .93, p = .49; Study 2: F (9, 14) = 1.28, p = .33), indicating that topic categories were randomly distributed across our word difficulty measure.

4. We ran all of our primary analyses using Flesch reading ease, Flesch-Kincaid grade level, and SUBTLEX-US measures as well (results from the reading ease analyses are presented in the online supplement). However, comparing and contrasting results across these analyses is not straight-forward due to our removal of stop-words for the Ngram measure (Benoit et al., 2019a), and the lack of availability of words in the SUBTLEX-US dictionary. In sum, though the results from analyses using SUTBLEX-US as the independent variable replicate the findings presented here, we did not find support for our hypotheses using reading ease and grade level measures as the independent variable. These divergent results, however, are not particularly surprising given that Ngram can be considered a semantic-measure of language difficulty whereas Flesch-based measures are syntactically based, and these metrics have been shown to perform differently in studies on political information processing (see, Tolochko et al., 2019).
5. We used the SMART stop word dictionary available in the "stopwords" package in the R statistical software (Benoit et al., 2019a).
6. More information about this pretest is included in the supplemental materials.
7. For consistency with the survey responses, we also elected to remove the item "A lot of the information presented was new to me." from the processing fluency scale. Doing so provided a more reliable scale of processing fluency (α = .91) compared to the same scale with this item included (α = .81). The three retained items were: 1) The ballot measure I just read was easy to read; 2) Overall, I found the language used in this ballot measure to be difficult (reverse-coded) ; 3) It was easy for me to understand the information presented.
8. In most cases, "pass" equates to greater than 50% of the votes cast in support of a measure. However, ballots in Colorado (n = 1) require >55% of support to pass, while ballots in both Florida (n = 2) and Illinois (n = 1) require >60% of support to pass (Ballotpedia, n.d.-b). These ballots were coded accordingly. Results of the same test in which all ballots are considered to have passed if more than 50% of the votes support the measure produced substantively similar results.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

The data for this study was made possible through funds awarded by the School of Communication's annual Miller Research Award.

Notes on contributors

Hillary C. Shulman, Ph.D. is an Associate Professor in the School of Communication at The Ohio State University. Her work examines how communication can be used to stimulate engagement in the areas of politics, health, and science.

Matthew D. Sweitzer, Ph.D. is a recent graduate of the Ohio State University School of Communication and a postdoctoral fellow at Sandia National Laboratories. His research explores the spread of information in social networks through a variety of computational methods. Note: this paper and associated research are in no way affiliated with Sandia National Laboratories; any views expressed here are solely those of the authors.

Olivia M. Bullock, Ph.D. is an Assistant Professor in the Department of Organizational Sciences and Communication at George Washington University. Her research focuses on message design strategies that can reduce the ideologically motivated processing of political, science, and health information.

Jason C. Coronel, Ph.D. is an Associate Professor in the School of Communication at The Ohio State University. His work examines the psychological processes that underlie political decision making.

Robert M. Bond, Ph.D. is an Associate Professor in the School of Communication at The Ohio State University. His research interest is in political behavior and attitudes with a specific interest in how our social networks influence our political behavior and communication.

Shannon Poulsen is a doctoral candidate in the School of Communication at The Ohio State University. She is broadly interested in how people become, and stay, politically misinformed and the role of humor as a source of, and solution to, misinformation.

ORCID

Hillary C. Shulman http://orcid.org/0000-0001-7525-8119
Matthew D. Sweitzer http://orcid.org/0000-0002-2075-6177
Olivia M. Bullock http://orcid.org/0000-0001-5403-7149
Jason C. Coronel http://orcid.org/0000-0001-6397-3860
Shannon Poulsen http://orcid.org/0000-0002-7680-0253

Data availability statement

The data described in this article are openly available in the Open Science Framework at https://osf.io/2q7cw/?view_only=15f95ed1a4a476099d0244b7dfba919.

Open Scholarship

This article has earned the Center for Open Science badge for Open Materials. The materials are openly accessible at https://osf.io/2q7cw/?view_only=15f95ed1a4a476099d0244b7dfba919.

References

Ballotpedia. (n.d.-a). *Ballot measure campaign finance, 2018*. Retrieved October 14, 2019, from https://ballotpedia.org/Ballot_measure_campaign_finance,_2018

Ballotpedia. (n.d.-b). *Supermajority requirement*. Retrieved July 15, 2019, from https://ballotpedia.org/Supermajority_requirement

Benoit, K., Muhr, D., & Watanabe, K. R Foundation for Statistical Computing, Vienna, Austria. (2019a Stopwords: The R package). (). https://stopwords.quanteda.io/

Benoit, K., Munger, K., & Spirling, A. (2019b). Measuring and explaining political sophistication through textual complexity. *American Journal of Political Science*, *63*(2), 491–508. https://doi.org/10.1111/ajps.12423

Berinsky, A. J., de Benedictis-Kessner, J., Goldberg, M. E., & Margolis, M. F. (2020). The effect of associative racial cues in elections. *Political Communication*, *37*(4), 512–529. https://doi.org/10.1080/10584609.2020.1723750

Bowler, S., & Donovan, T. (1998). *Demanding choices: Opinion, voting, and direct democracy*. University of Michigan Press.

Bowler, S. (2015). Information availability and information use in ballot proposition contests: Are voters over-burdened? *Electoral Studies*, *38*, 183–191. https://doi.org/10.1016/j.electstud.2015.02.002

Bowler, S., Dobbs, R., & Nicholson, S. (2020). Direct democracy and political decision making. In *Oxford research encyclopedias*. Oxford University Press.
Branton, R. P. (2003). Examining individual-level voting behavior on state ballot propositions. *Political Research Quarterly*, *56*(3), 367–377. https://doi.org/10.1177/106591290305600311
Burnett, C. M., & Kogan, V. (2015). When does ballot language influence voter choices? Evidence from a survey experiment. *Political Communication*, *32*(1), 109–126. https://doi.org/10.1080/10584609.2014.894160
Burnett, C. M. (2019). Information and direct democracy: What voters learn about ballot measures and how it affects their votes. *Electoral Studies*, *57*(November), 223–244. https://doi.org/10.1016/j.electstud.2018.12.001
Carlin, D. P., & Carlin, J. (1989). A typology of communication functions in ballot issue campaigns. *Political Communication*, *6*(4), 229–248. https://doi.org/10.1080/10584609.1989.9962877
Converse, P. E. (2006). The nature of belief systems in mass publics (1964). *Critical Review*, *18*(1–3), 1–74. https://doi.org/10.1080/08913810608443650
Coronel, J. C., Bullock, O. M., Shulman, H. C., Sweitzer, M. D., Bond, R. M., & Poulsen, S. (2021). Eye movements predict large-scale voting decisions. *Psychological Science*, *32*(6), 838–848. https://doi.org/10.1177/0956797621991142
Damore, D. F., & Nicholson, S. P. (2014). Mobilizing interests: Gorup participation and competition in direct democracy elections. *Political Behavior*, *36*(3), 535–552. https://doi.org/10.1007/s11109-013-9252-2
Downs, A. (1957). *An economic theory of democracy*. Addison Wesley.
Goldberg, M. H., & Carmichael, C. L. (2017). Language complexity, belief-consistency, and the evaluation of policies. *Comprehensive Results in Social Psychology*, *2*(2–3), 1–17. https://doi.org/10.1080/23743603.2017.1404427
Google. (2013). *Google Books: Ngram Viewer*. http://storage.googleapis.com/books/ngrams/books/datasetsv2.html
Green, D., Palmquist, B., & Schickler, E. (2002). *Partisan hearts & minds: Political parties and the social identities of voters*. Yale University Press.
Lupia, A. (1994). Shortcuts versus encyclopedias : Information and voting behavior in California insurance reform elections. *American Political Science Review*, *88*(1), 63–76. https://doi.org/10.2307/2944882
Markowitz, D. M., & Shulman, H. C. (2021). The predictive utility of word familiarity for online engagements and funding. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(18), 1–7. https://doi.org/10.1073/pnas.2026045118
Michel, J. B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Team, T. G. B., Aiden, E. L., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M. A., & Aiden, E. L. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, *331*(6014), 176–182. https://doi.org/10.1126/science.1199644
Milita, K. (2015). Election laws and agenda setting: How election law restrictiveness shapes the complexity of state ballot measures. *State Politics and Policy Quarterly*, *15*(2), 119–146. https://doi.org/10.1177/1532440015575901
Morisi, D., Colombo, C., & De Angelis, A. (2021). Who is afraid of a change? Ideological differences in support for the status quo in direct democracy. *Journal of Elections, Public Opinion and Parties*, *31*(3), 309–328. https://doi.org/10.1080/17457289.2019.1698048
Nai, A. (2015). The maze and the mirror: Voting correctly in direct democracy. *Social Science Quarterly*, *96*(2), 465–486. https://doi.org/10.1111/ssqu.12154
Nicholson, S. P. (2003). The political environment and ballot proposition awareness. *American Journal of Political Science*, *47*(3), 403–410. https://doi.org/10.1111/1540-5907.00029
Nicholson, S. P. (2005). *Voting the agenda: Candidates, elections, and ballot propositions*. Princeton University Press.
Petty, R. E., Brinol, P., Tormala, Z. L., & Wegener, D. T. (2007). The role of metacognition in social judgment. *Social Psychology: Handbook of Basic Principles*, *2*(1994), 254–284. https://doi.org/10.1016/j.eplepsyres.2008.03.002

Reilly, S., & Richey, S. (2011). Ballot question readability and roll-off: The impact of language complexity. *Political Research Quarterly*, *64*(1), 59–67. https://doi.org/10.1177/1065912909349629

Richey, S., & Taylor, J. B. (2020). Google Books Ngrams and political science: Two validity tests for a novel data source. *PS - Political Science and Politics*, *53*(1), 72–78 doi:https://doi.org/10.1017/S1049096519001318.

Schwarz, N. (2011). Feelings-as-information theory. In P. Van Lange, A. Kruglanski, & E. T. Higgins (Eds.), *Handbook of Theories of Social Psychology: Volume 1* (pp. 289–308). Sage Publications. https://doi.org/10.4135/9781446249215.n15

Schwarz, N. (2015). Metacognition M. Mikulincer, P. R. Shaver, E. Borgida, J. A. Bargh . In *APA Handbook of Personality and Social Psychology: Vol. 1 Attitudes and Social Cognition* (American Psychological Association) (pp. 203–229). https://doi.org/10.1037/14341-006

Shi, Y. (2016). Cross-cutting messages and voter turnout: Evidence from a same-sex marriage amendment. *Political Communication*, *33*(3), 433–459. https://doi.org/10.1080/10584609.2015.1076091

Shockley, E., & Fairdosi, A. S. (2015). Power to the people? Psychological mechanisms of disengagement from direct democracy. *Social Psychological and Personality Science*, *6*(5), 579–586. https://doi.org/10.1177/1948550614568159

Shulman, H. C., & Sweitzer, M. D. (2018a). Advancing framing theory: Designing an equivalency frame to improve political information processing. *Human Communication Research*, *44*(2), 155–175. https://doi.org/10.1093/hcr/hqx006

Shulman, H. C., & Sweitzer, M. D. (2018b). Varying metacognition through public opinion questions: How language can affect political engagement. *Journal of Language and Social Psychology*, *37*(2), 224–237. https://doi.org/10.1177/0261927X17707557

Shulman, H. C., & Bullock, O. M. (2019). Using metacognitive cues to amplify message content: A new direction in strategic communication. *Annals of the International Communication Association*, *43*(1), 24–39. https://doi.org/10.1080/23808985.2019.1570472

Shulman, H. C., Dixon, G. N., Bullock, O. M., & Colón Amill, D. (2020). The effects of jargon on processing fluency, self-perceptions, and scientific engagement. *Journal of Language and Social Psychology*, *39*(5–6), 579–597. https://doi.org/10.1177/0261927X20902177

Slater, M. D., Peter, J., & Valkenberg, P. (2015). Message variability and heterogeneity: A core challenge for communication research. *Annals of the International Communication Association*, *39*(1), 3–31. https://doi.org/10.1080/23808985.2015.11679170

song, H., & Ewoldsen, D. R. (2015). Metacognitive model of ambivalence: The role of multiple beliefs and metacognitions in creating attitude ambivalence. *Communication Theory*, *25*(1), 23–45. https://doi.org/10.1111/comt.12050

Stucki, I., Pleger, L. E., & Sager, F. (2018). The making of the informed voter: A split-ballot survey on the use of scientific evidence in direct-democratic campaigns. *Swiss Political Science Review*, *24*(2), 115–139. https://doi.org/10.1111/spsr.12290

Suiter, J., & Reidy, T. (2020). Does deliberation help deliver informed electorates: Evidence from Irish referendum votes. *Representation*, *56*(4), 539–557. https://doi.org/10.1080/00344893.2019.1704848

Sweitzer, M. D., & Shulman, H. C. (2018). The effects of metacognition in survey research: Experimental, cross-sectional, and content-analytic evidence. *Public Opinion Quarterly*, *82*(4), 745–768. https://doi.org/10.1093/poq/nfy034

Tolochko, P., Song, H., & Boomgaarden, H. (2019). "That looks hard!": Effects of objective and perceived textual complexity on factual and structural political knowledge. *Political Communication*, *36*(4), 609–628. https://doi.org/10.1080/10584609.2019.1631919

HUMAN COMMUNICATION RESEARCH

Human Communication Research ISSN 0360-3989

ORIGINAL ARTICLE

The Influence of Episodic and Thematic Frames on Policy and Group Attitudes: Mediational Analysis

Sheila A. Springer & Jake Harwood

Department of Communication, University of Arizona, Tucson, AZ 85721-0025, USA

An experiment examined how episodic and thematic political message frames affect attitudes toward older adults and Social Security. When exposed to messages about abolishing Social Security, participants exposed to episodic frames were significantly more likely to endorse message-consistent attitudes than participants exposed to a thematic frame. In mediation analyses, an episodic frame featuring a counterstereotypical exemplar increased endorsement of individual responsibility for retirement planning, which then led to more negative attitudes toward Social Security. These effects did not occur with a stereotypical exemplar in an episodic frame. The same mediated pathway influenced attitudes toward older adults in a more complex manner. Results provide support for individual responsibility attributions as a mediating mechanism underlying the effects of certain episodic frames.

Keywords: Episodic and Thematic Frames, Framing, Media Exemplars, Political Messages, Policy Attitudes, Attitudes Toward Aging, Group Attitudes.

doi:10.1111/hcre.12045

This study examines the intended and unintended effects of political messages on attitudes, utilizing framing theory. We aim to understand the nature of the effects and to explore some of the mechanisms by which message frames influence different types of attitudes. We compare episodic and thematic frames. Thematic frames emphasize broader trends or background information on a topic (Iyengar, 2011). Episodic frames utilize a particular individual's experience or a specific event to illustrate the issue (Iyengar, 1991). The human interest details in the episodic frame allow receivers to put a real face on the presentation of a problem (Semetko & Valkenburg, 2000). We believe that this exposure to a specific individual has implications for the intended persuasive intent of the message, and unintended effects on attitudes toward other similar individuals.

Corresponding author: Sheila A. Springer; e-mail: sspringe@email.arizona.edu

The study examines the effects of anti-Social Security messages on young adults. Social Security is a U.S. government benefit program for older adults, which is primarily funded through payroll tax revenues that go into a trust fund. The trust fund is projected to be depleted by 2036, resulting in substantial challenges in funding the program (Social Security Administration, 2011). These challenges could significantly impact how young adults prepare for retirement.

We consider two broad classes of attitudinal consequences of these messages: the intended effects on attitudes about the policy, and the (presumably unintended) effects on attitudes toward older adults—the policy's prime beneficiaries. Attitudes are important because they not only potentially predict behaviors (Fishbein & Ajzen, 1975), but also allow us to understand how people view the world (Fazio, 1986). Over the past several decades, social, political, and demographic change in the United States has polarized public attitudes toward large government programs that support vulnerable groups in the population (e.g., older adults, children, people with disabilities, etc.; Marmor, Cook, & Scher, 1997). Thus, understanding what sorts of messages lead to attitude change toward such policies is important.

We manipulated the message frame as either episodic or thematic; the former features an older adult supporting abolishment, whereas the latter features statistics and broader arguments for the same outcome. Using this design, we made two key contributions to the literature: We test the idea that (a) episodic frames affect attitudes through enhancing perceptions of individual responsibility and accountability and (b) episodic frames have unintended effects on attitudes about groups, by integrating message effects work with research on intergroup contact. Next, we present a brief review of the relevant framing literature and related hypotheses.

Message framing effects 1: Effects on attitudes about social policy

One way that media messages give meaning to issues and connect them with the larger political environment is through framing. A frame "suggests how the issue should be thought about and understood" (Nelson & Kinder, 1996, p. 1057). News frames are an efficient way for journalists to classify information to communicate to audiences (Gitlin, 1978). These are often combined with audience frames which "guide individuals' processing of information" (Entman, 1993, p. 53).

News framing of political issues is often divided into two types: episodic and thematic. As defined by Iyengar (2011), "a thematic news frame … usually takes the form of in-depth background," and tends to use statistics. In contrast, an episodic news frame "depicts issues in terms of individual instances or specific events," often an individual's personal narrative (p. 253). Episodic frames are more engaging than thematic frames (Gross, 2008; Iyengar, 1991). This appeal has been explained as being a result of the personal connection to the individual in the episodic frame (Gross, 2008), the narrative arc of an episodic presentation (Iyengar, 1991), the ability of narrative to reduce reactance and counterarguing (Slater & Rouner, 2002), and the relatively digestible nature of the content (i.e., personal stories as compared to numerical information and broader generalizations; Iyengar, 1991).

Beyond these mechanisms, the presence of individual people in episodic frames may permit yet another mechanism to operate: a shifting of attributions of responsibility to the individual (Gamson & Modigliani, 1987; Iyengar, 1991). Episodic framing focuses on a particular individual's story, and hence frames the problem (and plausibly the solution) as an individual one (Iyengar, 1991). For messages advocating elimination of government social programs, framing at the individual level should enhance perceptions that the individual-level solutions are most relevant and, hence, should reduce favorability toward the collective-level solutions—i.e., large government programs. On the other hand, thematic framing of messages about social welfare focuses more on breadth and background information and frames the problem as a collective issue—one that has its locus and perhaps its solution at the governmental or societal level (Iyengar, 1991). Such messages should be less successful in gaining support for eliminating government programs.

This prediction is somewhat supported by prior framing research. Iyengar's (1987, 1991) series of experiments on crime, poverty, and unemployment found that when the issue was described in thematic terms, respondents assigned responsibility to societal factors such as failed governmental programs, political climate, or economic conditions. In contrast, when news coverage of poverty used episodic terms and dwelled on particular instances of poor people, respondents were more apt to hold the poor causally responsible. However, although those findings were directionally consistent with the hypotheses, they failed to achieve statistical significance.

Research to extend and clarify Iyengar's (1987, 1991) claims proposed that this distinction between episodic and thematic frames was overly simplistic and did not account for the multiple dimensions involved in the cognitive processing of frames. Shah, Kwak, Schmierbach, and Zubric (2004) declared that Iyengar's (1987, 1991) inconsistent findings may be the result of confounding two frame dimensions: An episodic message "favors specific instances over enduring problems," and "emphasizes individual situations over societal conditions" (p. 104). Therefore, a sense of shared responsibility to solve societal problems may be lessened as audiences encounter stories about *individuals* dealing with their personal struggles (Shah et al., 2004). Shah et al. extend Iyengar's work by testing the effects of mixtures of frames, showing that a combination of frame categories (societal gain/individual loss) led to a more detailed description of the issue.

Research also has expanded on the effects of *exemplars*, an important feature of episodic frames (Brosius & Bathelt, 1994; Lefevere, Swert, & Walgrave, 2012; Perry & Gonzenbach, 1997; Tran, 2012). Brosius and Bathelt proposed that exemplars significantly affect how the message is perceived by the audience and found that the perception of the majority opinion was strongly influenced by exemplars. Exemplification theory proposes that the vividness of exemplars makes the message more memorable and triggers heuristic processing, giving exemplars an advantage over base-rate information in influencing perceptions of an issue (Tran, 2012; Zillmann

& Brosius, 2000). Heuristic processing tendencies are supported such that the vividness and quality (extremity of opinion) of the exemplar was more memorable than the base-rate information suggesting audiences "computed the perceived opinions by the number of pro- and con-exemplars" (Brosius & Bathelt, 1994, p. 73). Extending this line of research, Perry and Gonzenbach (1997) compared national and local exemplars, finding that opinion change followed the direction of the exemplars, and local exemplars were more effective at influencing the perception of public opinion about a local controversial issue. Thus, exemplars are tools that can be used to sway public opinion in a particular direction, generate perceptions of majority public opinion, and affect opinions about the future of an issue (Perry & Gonzenbach, 1997).

Newer research supported the idea that exemplar vividness influences issue perceptions. Popular exemplars were perceived as trustworthy and their accounts were seen as vivid; they were seen as representing public opinion and their message was taken more seriously than expert exemplars (Lefevere et al., 2012). Thus, popular exemplars had a greater effect on perceptions of the issue than experts. Tran (2012) found that exemplar vividness moderated the relationship between message valence and opinions such that valence has stronger effects with a vivid exemplar. Research consistently supported the notion that messages using exemplars have stronger influence on opinion than messages using base-rate information (Zillmann, 1999; Zillmann & Brosius, 2000).

Beyond discussions of individual exemplars, however, the broader issue of *how* exemplars influence processing of messages (relative to a lack of exemplars) remains somewhat unresolved. Iyengar suggested that exemplars and episodic frames led to more individual-level thinking than broader thematic frames, but his data did not fully support the idea, and he never explicitly examined the mediated pathway implied by that argument. In this study, we argue that attributions of responsibility at the individual (versus societal) level are an outcome of framing, and that they have attitudinal consequences. We predict that episodic frames should lead to individual attributions of responsibility, whereas thematic frames should lead to government/social attribution of responsibility. Thus, in the case of a message advocating abolition of Social Security:

H1: An episodic frame will lead people to be more supportive of abolishing Social Security than a thematic frame.

H2a: The effect in H1 will be mediated by attributions of individual responsibility such that an episodic frame will lead to higher attribution of individual responsibility for older adults' financial security in retirement than a thematic frame, and attributions of individual responsibility will increase desire to abolish Social Security.

H2b: The effect in H1 will be mediated by attributions of government responsibility such that a thematic frame will lead to higher attribution of government responsibility for older adults' financial security in retirement than an episodic frame, and attributions of government responsibility will decrease desire to abolish Social Security.

Message framing effects 2: Effects on attitudes about older adults

In addition to the effects of the anti-Social Security messages on attitudes about the policy, these messages also have the potential to impact attitudes about older adults. Given that Social Security is a benefit program for older adults, it places older people in a position of dependency (i.e., as in need of government assistance), and given the financial problems perceived to stem from Social Security it also raises the possibility of older people being held responsible for broader financial challenges to the nation. Hence, we extend research on episodic and thematic framing effects by examining the unintended effects of messages on attitudes toward *groups* embedded in the messages. Specifically, we propose that the framing of political messages about cuts to Social Security may affect young adults' attitudes toward older adults in general.

Allport's (1954) intergroup contact theory suggests that contact between members of different social groups will increase positive attitudes toward outgroup members. *Direct* contact (e.g., face-to-face conversation) allows people the opportunity to connect, develop mutual understanding and trust, and discover shared interests with outgroup members (Dovidio, Eller, & Hewstone, 2011). *Indirect* contact via media messages may also influence attitudes toward outgroup members (Ortiz & Harwood, 2007; Schiappa, Gregg, & Hewes, 2005). Media messages about aging often support preexisting stereotypes about older adults, with few messages portraying positive or even adequate older occupational role models (Signorielli, 2004); such messages reinforce perceptions that older people are unproductive and do not contribute to society (Signorielli, 2004). Globally, media consumption has been shown to contribute to attitudes about aging (Levy, 2003).

Our episodic frames, which featured articulate and independent older people, offer an opportunity for indirect intergroup contact, an opportunity that is not present (nor would be expected) in the thematic frame. Through this indirect contact, young adults could gain personal knowledge and information about the particular older adult, broadening the scope for empathy with older adults in general and hence positive attitudes about the group. An individual experiencing indirect contact via an episodic frame has an opportunity to develop more positive attitudes toward older adults in general that is unavailable in the thematic frame.

H3: An episodic frame will lead to more positive attitudes toward older adults than a thematic frame.

The older adult's narrative in our episodic message supported abolishing Social Security by discussing the values of individual hard work, independence, and self-reliance. Our society places a strong value on an individual's ability to work and maintain independence; an older adult who espouses such values and disdains government support should be perceived by young adults as accepting individual responsibility. Meta-analyses of attitudes toward younger and older adults have indicated there are generalized negative attitudes toward older adults across a wide range of measures: competence, attractiveness, health, hearing ability, demeanor (e.g., grouchy, critical, miserly), participation in activities, happiness, and desirability as company (Kite & Johnson, 1988; Kite, Stockdale, Whitely, & Johnson, 2005). Given

there are widely accepted negative stereotypes of older adults as incompetent (Cuddy, Norton, & Fiske, 2005), exposure to responsible and independent older people should reduce endorsement of the stereotype and improve perceptions of older people in general. In other words, effects of our episodic message on more positive attitudes about older people should be mediated by perceptions of individual responsibility.

H4: The effect in H3 will be mediated by attributions of individual responsibility such that an episodic frame will lead to higher attribution of individual responsibility for older adults' financial security in retirement than a thematic frame, and attributions of individual responsibility will lead to more positive attitudes toward older adults.

Thus, we aim to understand the impact of episodic versus thematic frames in influencing the attitudes targeted by the message *and* attitudes concerning the social group of which the episodic condition features an exemplar. Understanding how message frames can affect attitudes toward older adults and public policy programs like Social Security helps us (a) better understand the tensions implicit in messages about social programs in a political environment that favors self-reliance over government intervention, and (b) appreciate the effects of messages about social programs that are associated with specific social groups who may be present in messages about the programs (Silverstein & Parrott, 2001). Although we focus specifically on older adults and Social Security, our research has implications for messages about welfare that might feature poor people or disadvantaged ethnic groups, messages about education policy featuring children or college students, messages about health policy featuring people suffering from stigmatized medical problems, and myriad other situations in which social policies and social groups are intertwined.

Method

Participants and procedure

In total, 218 participants were recruited from communication classes at a large southwestern U.S. university to participate in a study about "evaluating news stories." Five participants over the age of 30 were excluded from data analysis (counterstereotypical episodic [1], control [2], and stereotypical episodic [3]), and an additional 16 participants were removed for failing a validation check (described later). Of the remaining 197 participants, 72.1% were female (27.9% male); ages ranged from 18 to 27 ($M = 20.77$, $SD = 1.32$), and 71.1% of the participants were White (11.2% Latino/Hispanic; 6.1% African American; 6.1% Asian American; and 5.6% Other).

Participants were randomly assigned to one of four online experimental conditions (stereotypical episodic, counterstereotypical episodic, thematic, or control) via an e-mail link provided by the investigator. All groups except the control read an article advocating the abolishment of Social Security. For the thematic group the article contained general background information supported by statistical figures. For the stereotypical episodic group the article was written from the perspective of a typical older adult. For the counterstereotypical episodic group, it was written from the perspective of an atypical older adult. The differentiation between the stereotypical

and counterstereotypical conditions is discussed in the messages section that follows; these two conditions were included primarily to provide a degree of message replication within the episodic condition, given the potential for massive variation in the nature of exemplars that might be present in such messages. The control group read an article about personality genes in bees. All participants completed a pre- and posttest questionnaire-containing measures described below.

Messages

Given that our message advocated abolition of Social Security, we elected to include exemplars that were broadly supportive of the theme that older adults are active, independent, self-supporting, and self-reliant, and hence not necessarily *in need of* programs like Social Security. We used two variations on this theme, one of which was more dramatic than the other: the stereotypical episodic ($n = 47$) and counterstereotypical episodic ($n = 51$) conditions. Both versions told the story of an individual retiree who supports the abolishment of Social Security. The conditions were distinguished using photos and information about the older adult's age, education, and reason for retirement.

In the *counterstereotypical episodic* condition, the older adult was a 73-year-old CEO retiring by choice after 47 years in the workforce. Counterstereotypicality in the character was represented through age (older than the average American retiree), title of CEO (high income bracket), education (MBA graduate degree), and retiring by choice (experienced financial success). The photo portrayed an older adult in suit and tie comfortably reclining in a high back, leather business chair. His chin is resting on his hand and he is looking at the camera with a wry smile. The appearance of the counterstereotypical older adult indicates power, status, and competence which often are not attributed to older adults in general.

The older adult in the *stereotypical episodic* condition was a 65-year-old being forced to retire due to memory problems. This character is more in line with aging stereotypes: He is retiring at an age which most of the population would connect with retirement, and having memory problems which are often attributed to age. Stereotypicality in the character is further represented by education (high school graduate) placing him in a lower income bracket when compared to the counterstereotypical older adult. The photo of the stereotypical older adult is a head shot of a man dressed in a casual golf shirt; he has a slight furrow in his brow which might indicate some confusion. The manipulation check for this distinction failed—the two older characters were viewed as equally typical of older people in general. Nonetheless, we retained the distinct groups in our analysis.

The thematic framed article ($n = 53$) used background information about Social Security to support the argument for abolition. It included statistics such as cost to maintain the program and contributing factors such as the increasing dependence of Americans on government programs. The control group read an Associated Press article about personality genes in bees that was comparable in length to the experimental articles and contained a photo ($n = 46$).

All experimental versions were comparable in length and provided the same background information on Social Security. They featured the same number of arguments against Social Security, and the language used was similar in all versions (e.g., *abolishment*, *trillion dollar liability*, *rapidly bankrupting our government*, and *unmanageable burden*). In addition, all versions contained a visual aid to enhance the manipulation (a statistical chart in the thematic condition, a photo in the episodic conditions). All articles were created using portions of existing editorials representative of the kind of coverage people might actually read online.

Our approach to framing emphasized precision such that facts presented about Social Security were maintained across conditions using similar language. By holding content constant, internal validity is increased by "restricting framing very narrowly to an effect of presentation and modality" (Scheufele & Tewksbury, 2007, p. 10). However, external validity may be limited because effects are more likely due to a combination of both content and framing (Scheufele, 2000).

Measures

All measures were completed after exposure to the message, except as noted.

Subject screening

To ensure participants read the articles, there were two multiple choice questions regarding the content of the articles. Sixteen respondents answered one or both of the validity questions incorrectly and were removed from data analysis (thematic (1), counterstereotypical episodic (2), stereotypical episodic (5), and control (8)). A review of these responses in the control condition showed incorrect responses varied and indicated no systematic bias. Therefore, we concluded that these participants were unmotivated to participate mindfully in the study, perhaps due to a lack of interest in the control message topic.

Attitudes toward older adults

These attitudes were measured using the General Evaluation Scale (Wright, Aron, McLaughlin-Volpe, & Ropp, 1997). Participants rated their feelings about older adults (people 65 and older) on six items (*negative*, *warmth*, *suspicious*, *friendly*, *respect*, and *disgust*) using a 5-point Likert-type scale where higher scores indicate more positive attitudes toward older adults (Harwood, Hewstone, Paolini & Voci, 2005; $\alpha = .76$).

Attitudes toward Social Security

Respondents were provided basic information to define Social Security from the Administration's Website (www.ssa.gov) prior to completing a single-item pretest measure of attitudes toward Social Security embedded in other measures (single item, 1–10, disfavor vs. favor the Social Security system). In the posttest, attitudes were measured using five items on a 5-point Likert-type scale from 1 (*strongly disagree*) to 5 (*strongly agree*) (e.g., "Social Security is not important for future generations" (reverse coded); "older adults should have access to Social Security"). Existing measures could

not be located so items were created for this study ($\alpha = .73$); high scores indicate positive attitudes about Social Security.

Attribution of responsibility

Participants responded to two single-item questions on Likert-type scales from 1 (*none*) to 5 (*a great deal*): "How much responsibility do older adults have for financing their own retirement needs?" and "How much responsibility does the government have for supplementing older adults' retirement needs?"

Data analysis

Descriptive statistics and information about statistical transformations are provided in Table 1. Pretest attitudes toward Social Security, race (coded White/not White), and political affiliation (coded Democrat/not Democrat) were used as covariates in all data analyses. To test hypotheses of mean differences, one-way analyses of variance with covariates (ANCOVAs) were performed to compare the four conditions. Post-hoc Tukey's least significant difference (LSD) analyses were performed to test pairwise differences. Mediator hypotheses were tested via Hayes' PROCESS macro for SPSS (Hayes, 2013) incorporating dummy codes for *both* the stereotypical and counterstereotypical (relative to thematic) conditions as predictors, attributions of individual responsibility and attributions of government responsibility as mediators, and attitudes toward Social Security or attitudes toward older adults as outcome variables. Each outcome variable model was run twice including one of the dummy codes as a predictor and the other as a covariate in order to assess the indirect effects for each. Additionally, the attitude measures were included as potential mediators (i.e., attitudes toward older adults was used as a mediator when Social Security attitudes was the outcome and vice versa) which (a) controls for any intercorrelation between the attitude measures, and (b) allows us to consider the direction of causality. The control group was not included in tests of mediator effects.

Results

For analyses involving attitudes toward older adults, covariates of pretest attitudes about Social Security, race, and political affiliation were not significant predictors. For analyses involving Social Security attitudes, political affiliation was typically a significant predictor (explaining about 7% of the variance) as were preexisting attitudes about Social Security (explaining about 12% of the variance). Race was not a significant predictor. In the subsequent reports, these covariates are not reported. Differences across conditions and descriptive statistics for the major study variables are reported in Table 1, and correlations related to the hypotheses are presented in Table 2. All confidence intervals are 95% bias corrected.

Hypothesis 1 predicted that the episodic frames would lead people to be less supportive of Social Security than a thematic frame. The four groups differed significantly

Table 1 Raw Means, Standard Deviations, and Descriptive Statistics for the Major Study Variables

Variable	CST-Episodic *M* (*SD*)	ST-Episodic *M* (*SD*)	Thematic *M* (*SD*)	Control *M* (*SD*)	Theoretical Range	Observed Range
Attitudes toward Social Security	3.58 (.60)[a]	3.48 (.52)[a]	3.81 (.63)[b]	3.74 (.59)[b]	1.00 – 5.00	2.00 – 5.00
Attitudes toward older adults	0.65 (.11)[a]	0.67 (.10)[ab]	0.70 (.12)[b]	0.67 (.12)[ab]	0.33 – 1.00	0.33 – 1.00
Individual attribution mediator	4.12 (.79)[a]	3.87 (.99)[ab]	3.62 (.79)[b]	3.96 (.87)[a]	1.00 – 5.00	2.00 – 5.00
Government attribution mediator	3.08 (.94)[a]	3.43 (.93)[b]	3.43 (.89)[b]	3.39 (.95)[ab]	1.00 – 5.00	1.00 – 5.00

Note: $N = 197$. CST = counterstereotypical; ST = stereotypical. Log transformations to obtain appropriate skewness ranges were performed on attitudes toward older adults. Means not sharing subscripts across rows are significantly different ($p < .05$), except for the following where the difference was marginally significant ($p < .10$): the differences between the CST-Episodic and Control conditions for attitudes toward Social Security, and the Thematic and Control conditions for the individual attributions mediator. The overall *F* statistic is significant for attitudes toward Social Security and individual attributions; it is nonsignificant for the other two analyses.

Table 2 Correlation Matrix for Major Study Variables

Variable	(1)	(2)	(3)	(4)
1. Attitudes toward older adults	—			
2. Attitudes toward Social Security	.10	—		
3. Individual attribution	.07	−.18**	—	
4. Government attribution	−.003	.25**	−.27**	—

Note: $N = 197$.
**$p < .01$

in their posttest attitudes toward Social Security, $F(3, 190) = 4.43$, $p = .005$, $\eta_p^2 = .07$. Tukey's pairwise comparisons indicated that the stereotypical and counterstereotypical episodic groups had significantly lower support of Social Security than the thematic group, and did not differ from each other (Table 1). The stereotypical episodic group had significantly lower support of Social Security than the control group; the corresponding difference for the counterstereotypical episodic group was marginally significant. Results supported the hypothesis.

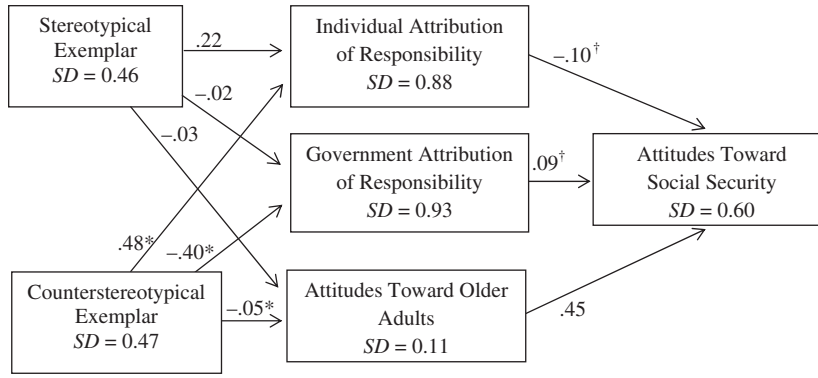


Figure 1 Results of regression analysis showing that the effect of counterstereotypical message frame on attitudes toward Social Security is mediated by individual attribution of responsibility and government attribution of responsibility. The numbers are unstandardized regression coefficients. Controlling for preexisting attitudes toward Social Security, race, and political affiliation did not affect the results. $^*p < .05$, $^\dagger p > .08$, $N = 151$.

Hypothesis 2a predicted that the effect in H1 was mediated by attributions of individual responsibility, and Hypothesis 2b predicted that the effect in H1 was mediated by attributions of government responsibility. Results showed two significant mediated pathways in the model: one from the counterstereotypical frame, through attributions of individual responsibility, to the outcome, CI [−0.138, −0.003], and another from the counterstereotypical frame, through attributions of government responsibility, to the outcome, CI [−0.1190, −0.0004].

As shown in Figure 1, participants in the counterstereotypical episodic condition attributed more responsibility to individuals to save for retirement than did those in the thematic condition, and individual attributions of responsibility resulted in reduced support for Social Security. Conversely, participants in the thematic condition attributed more responsibility to the government to provide for retirement than did those in the counterstereotypical episodic condition, and government attributions of responsibility resulted in increased support for Social Security. These mediated paths are also significant in simpler models incorporating a single mediator and a simple contrast of stereotypical and counterstereotypical conditions. The direct path from counterstereotypical frame to the outcome was not significant after inclusion of the mediators, CI [−0.40, 0.04], suggesting that attributions fully mediate the effect.

The ratio of indirect to total effect indicates that approximately 24% of the *total effect* of the counterstereotypical frame on attitudes toward Social Security is carried *through* individual attributions (the mediator), and the overall R^2 mediating effect size indicates that 2.64% of the total variance in attitudes toward Social Security is explained via the mediated pathway. Further, approximately 19% of the *total effect* of the counterstereotypical frame on attitudes toward Social Security is carried *through* government attributions, and the overall R^2 mediating effect size indicates that 2.05%

of the total variance in attitudes toward Social Security is explained via the mediated pathway (Preacher & Kelley, 2011). Although these effect sizes are small, they are not trivial: Effects of this size or smaller are common in media effects research, including work on framing (e.g., O'Keefe and Jensen's 2007 meta-analysis of 93 studies shows an effect of $r = .03$ ($r^2 = .0009$) for the effects of gain/loss frames on health behaviors).

No mediation effects were significant for the comparison of the stereotypical episodic versus thematic conditions: attributions of individual responsibility CI [−0.10, 0.01]; and, attributions of government responsibility CI [−0.06, 0.03]. Attitudes toward older adults were not a significant mediator of the effects of either condition: counterstereotypical CI [−0.09, 0.01]; stereotypical CI [−0.08, 0.01]. Results supported Hypotheses 2a and 2b only for the counterstereotypical episodic condition.

Hypothesis 3 predicted that the episodic frames would lead to more positive attitudes toward older adults than the thematic frame. The stereotypical episodic, counterstereotypical episodic, thematic, and control groups did not differ significantly $F(3, 190) = 1.65$, *ns*. A contrast of the counterstereotypical episodic and thematic conditions was significant, $F(1, 99) = 4.56$, $p = .04$, $\eta^2 = .04$, but there was no significant difference between the stereotypical episodic and thematic conditions. Those in both episodic conditions reported more negative attitudes about older people than those in the thematic condition (see Table 1 for means), which contradicts H3. However, the results of H4 suggest that the logic underlying the hypothesis was partially correct.

Hypothesis 4 predicted that the effect in H3 was mediated by attributions of individual responsibility. Analysis indicates there is one significant mediated pathway in the model from the counterstereotypical frame, through attributions of individual responsibility, to the outcome, CI [0.001, 0.030]. As shown in Figure 2, participants in the counterstereotypical episodic condition attributed more responsibility to individuals to save for retirement than those in the thematic condition, and individual attributions resulted in more positive attitudes toward older adults. The relationship between counterstereotypical message frame and attitudes toward older adults remained significant (and negative) after inclusion of the mediator, CI [−0.10, −0.01]. This mediated path is also significant in a simpler model incorporating a single mediator and a simple contrast of stereotypical and counterstereotypical conditions. The ratio of indirect to total effect indicates that 26% of the *total effect* of the counterstereotypical frame on attitudes toward older adults is carried *through* individual attributions (the mediator), and the overall R^2 mediating effect size indicates that 1.98% of the total variance in attitudes toward older adults is explained via the mediated pathway.

As suggested previously, the positive indirect effects support the logic of H3, even though the total effects did not support H3. Counterstereotypical episodic frames about Social Security make attitudes about older adults worse (total effect), but they also improve attitudes about older adults *through* attributions of individual responsibility for retirement planning. No other mediation effects were significant: stereotypical episodic versus thematic conditions through attributions of individual

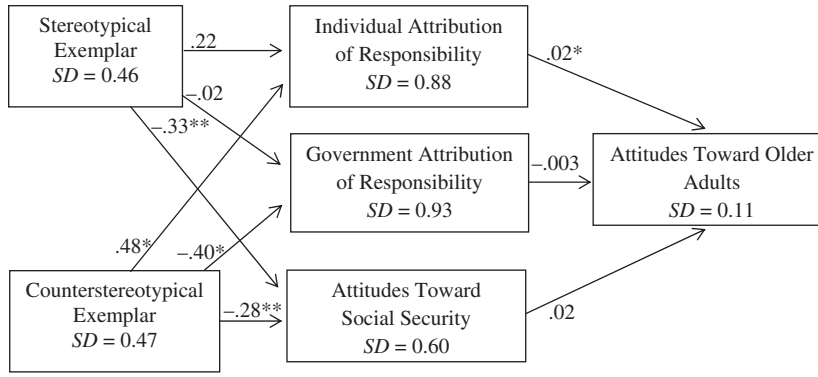


Figure 2 Results of regression analysis showing that the effect of counterstereotypical message frame on attitudes toward older adults is mediated by individual attribution of responsibility. The numbers are unstandardized regression coefficients. Controlling for preexisting attitudes toward Social Security, race, and political affiliation did not affect the results. $^{*}p < .05$, $^{**}p < .01$, $N = 151$.

responsibility, CI [−0.002, 0.019]; as well as both conditions through attributions of government responsibility: counterstereotypical CI [−0.007, 0.013], stereotypical CI [−0.004, 0.005]; and attitudes toward Social Security: counterstereotypical CI [−0.024, 0.002] or, stereotypical CI [−0.022, 0.003].

Discussion

We proposed that the framing of political messages has effects on policy attitudes, but also unintended effects on attitudes toward the beneficiaries of the policy; we explored attributions of responsibility as a theoretically derived mediator of these effects. The effects were explored in the context of an *anti*-Social Security message.

Effects on policy attitudes

Message frame directly affected respondents' attitudes toward Social Security—those in the episodic conditions were more negative about Social Security than those in the thematic or control conditions. For the comparison of the counterstereotypical episodic versus thematic conditions, this effect was mediated by both individual and government attribution of responsibility for retirement planning. This provides the first evidence for personal responsibility as a mediator of episodic framing effects, an effect first hypothesized by Iyengar (1987) who suggested that episodic frames encourage individualist perspectives on issues.

As predicted, young adults in the counterstereotypical episodic condition rated individual responsibility significantly higher than the thematic condition—a message-consistent ideology that supports the message. This translated into respondents following the opinion of the exemplar by rating attitudes toward Social Security lower, indicating the exemplar affected perceptions of the message (Brosius & Bathelt,

1994; Perry & Gonzenbach, 1997). We believe our work is the first to explicitly model this process and demonstrate that individualist orientation mediates the effects of episodic frames on attitudes (Iyengar, 1991).

In a complementary effect, the episodic stereotypical condition also reduced perceptions of government responsibility, and lowered perceptions of government responsibility were associated with more negative attitudes toward the policy. In other words, individual and collective responsibility attributions both independently mediated the effects of an episodic manipulation. However, this is not a complete confirmation of Iyengar (1991) as we discuss next.

The mediation pattern occurred for the *counter*stereotypical episodic condition, but not for the *stereotypical* episodic group. This may indicate the vividness of the counterstereotypical older adult exemplar makes the message more memorable, as exemplification theory suggests (Zillmann & Brosius, 2000). In particular, in the context of a message advocating less support for older adults, our counterstereotypical exemplar was a better embodiment of the message—he was presented as wealthy, successful, and highly competent, and hence exemplified precisely the type of older person for whom the message would make sense. In contrast, the stereotypical exemplar was experiencing age-related decline (memory problems) and hence although he was verbally making the case for not supporting older adults, he actually exemplified a reason why support for older adults might be necessary. Clearly our current data are only suggestive, but they indicate that a greater match between exemplar circumstances and argument content should lead to greater persuasiveness. This is somewhat similar to Perry and Gonzenbach's (1997) finding that local exemplars are more powerful in influencing opinion on local issues; they have better "fit" to those issues. More generally, some exemplars are clearly better than others (Lefevere et al., 2012).

Although our results show full mediation, globally they do not explain a large amount of variance. Therefore, we should consider other mediating factors that might be operating simultaneously. In particular, work on entertainment overcoming resistance (Moyer-Gusé, 2008) and transportation (Slater & Rouner, 2002) suggests that episodic messages reduce counterarguing or reactance and "entertain" the reader into submission. Depending on the specific message and outcome, resistance-reduction mediators and individual-societal attribution mediators might complement each other in influencing attitudes. In other cases, they might operate in opposition. For instance, Niederdeppe, Shapiro, and Porticella (2011) used messages that emphasized external causes of obesity. Their episodic messages were more persuasive (for a subset of their sample), an effect that was mediated by reduced counterarguing: Political liberals were more supportive of societal solutions to obesity problems in a narrative (episodic) condition.

Our results and Iyengar's theorizing would suggest that an individual-societal attribution-mediated path would (simultaneously) operate in the opposite direction in Niederdeppe et al. (2011) data, leading to *less* support for societal solutions in the episodic condition. In this context, we find it telling that the most effective message overall in Niederdeppe et al.'s (2011) study is one that combines episodic

and thematic features, perhaps short-circuiting the type of mediated pathway that we hypothesize. Future work should simultaneously examine contradictory mediators for messages with persuasive goals that involve both individual and societal solutions.

Our mediation results suggest that Iyengar's (1991) focus on the effect of attributions is not entirely warranted. Only a small portion of the effect of thematic versus episodic framing results from perceptions of attribution of responsibility. Furthermore, not all episodic frames are equivalent as evidenced by differing results for the two episodic exemplars. This clearly suggests that additional mechanisms such as specific characteristics of the exemplar are important in this process.

We suspect additional processes favoring the episodic frame derive from young people's use of personal news sources. This bias should lead people to pay more attention to episodic versus thematic news frames; as such they might read episodic frames more carefully and retain the arguments better. Episodic frames also allow for perspective taking with the characters in the stories — presumably a characteristic of personal news stories. Resistance to the "abolish" Social Security message is likely to be grounded in the perception that older adults need and want Social Security. Reading a story in which a *specific* older adult opposes that notion reduces this perception.

Effects on attitudes toward groups

Our results also show an unintended effect of framing on attitudes toward older adults. Counterstereotypical episodic framing of messages advocating abolishing Social Security worsened attitudes about older adults (compared to thematic frames) as a direct effect. However, the same frame also increased attributions of individual responsibility which in turn improved attitudes about older adults. This supports our general idea that policy messages have unintended effects on attitudes toward the policy's prime beneficiaries. However, the direction of the direct effect is the opposite of our predictions; we expected exposure to an independent and competent older adult to improve attitudes.

Perhaps our respondents did not *like* the particular older adults in either of our episodic conditions — contact with disliked outgroup individuals is not expected to yield positive attitudinal outcomes (Paolini, Harwood, & Rubin, 2010). Our data suggest a fairly tepid level of liking for the characters in the episodic conditions ($M = 2.55$, $SD = .85$, on a 1–5 scale, where high scores indicate more liking; no difference between the two conditions). The characters may have been downgraded for speaking out against their in-group's interests, or simply for failing to confirm respondents' "worldviews" concerning older adults (e.g., by being moderately competent and independent; Glick & Fiske, 2001). Another possibility is that when young people see older people apparently arguing against their group's interest that they view older adults as not unified on this issue, and perceptions of disagreement among the policy's beneficiaries result in negative attitudes about the group (e.g., "they're disorganized").

That said, the mediator effects here provided support for our underlying logic: Respondents in the counterstereotypical episodic condition attributed more responsibility to individuals to save for retirement, which led to more positive attitudes toward older adults. In other words, the indirect effect supported the direction that we had predicted for this effect and demonstrates that episodic policy messages can have implications for members of policy-relevant groups.

Limitations and future directions

A limitation in this study is the use of single-item measures to explore individual and government attributions of responsibility as mediators. No scales were found in a search for measures of individual and government responsibility, so single-item measures were created to explore the concept for this study. Creating multi-item scales for use in future studies would benefit the literature. Another limitation is the lack of a pro-Social Security condition in the experiment. Demonstrating that episodic frames perform *worse* for a pro-Social Security message, and that that effect is also mediated by individual attributions is the next building block in the argument we make here. A more diverse sample in terms of age and work experience would benefit future investigations into moderators of our effects: Older respondents with more work experience should be more engaged with the issue and have more complex perceptions of aging and older adults (Hummert, Garstka, Shaner, & Strahm, 1994).

Finally, we acknowledge the limitations of mediated models in examining causality throughout the model. Specifically, we are confident in the causal effects of our independent variable on the mediator (because it is an experimental manipulation), but note that the causal effects of the mediator on the dependent variable are inferred from what are effectively correlational data; it is possible. Although not consistent with our theoretical position, there might be effects in the reverse direction (DV to mediator). This speaks to the need for separate work experimentally manipulating individual and governmental attributions to examine their effects on policy attitudes.

Our study provides a valuable addition to framing research in the context of political message effects on attitudes. We show how specific political message frames related to public policies are associated with attitudes toward those policies and the people who are the target of the policies, partially supporting Iyengar's claims. However, there are clearly other factors also involved, as discussed in more recent work (Lefevere et al., 2012; Shah et al., 2004; Tran, 2012). This enhances the framing literature by demonstrating the role of attribution of responsibility as a potential mediator in the attitudinal effects of frames, while simultaneously drawing attention to the fact that additional mediators also operate in this process.

References

Allport, G. W. (1954). *The nature of prejudice*. Cambridge, MA: Addison-Wesley.

Brosius, H. B., & Bathelt, A. (1994). The utility of exemplars in persuasive communications. *Communication Research*, **21**, 48–78. doi:10.1177/009365094021001004

Cuddy, A. J., Norton, M. I., & Fiske, S. T. (2005). This old stereotype: The pervasiveness and persistence of the elderly stereotype. *Journal of Social Issues*, **61**, 267–285. doi:10.1111/j.1540-4560.2005.00405.x

Dovidio, J. F., Eller, A., & Hewstone, M. (2011). Improving intergroup relations through direct, extended and other forms of indirect contact. *Group Processes and Intergroup Relations*, **14**, 147–160. doi:10.1177/1368430210390555

Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, **42**, 51–58. doi:10.1111/j.1460-2466.1993.tb01304.x

Fazio, R. H. (1986). How do attitudes guide behavior?. In R. M. Sorrentino & E. T. Higgins (Eds.), *The handbook of motivation and cognition: Foundations of social behavior* (pp. 204–243). New York, NY: Guildford Press.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.

Gamson, W. A., & Modigliani, A. (1987). The changing culture of affirmative action. In R. G. Braungart (Ed.), *Research in political sociology* (Vol. **3**, pp. 137–177). Greenwich, CT: JAI.

Gitlin, T. (1978). Media sociology: The dominant paradigm. *Theory and Society*, **6**, 205–253. doi:10.1007/BF01681751

Glick, P., & Fiske, S. T. (2001). Ambivalent stereotypes as legitimizing ideologies: Differentiating paternalistic and envious prejudice. In J. T. Jost & B. Major (Eds.), *The psychology of legitimacy: Emerging perspectives on ideology, justice, and intergroup relations* (pp. 278–306). New York, NY: Cambridge University Press.

Gross, K. (2008). Framing persuasive appeals: Episodic and thematic framing, emotional response, and policy opinion. *Political Psychology*, **29**, 169–192. doi:10.1111/j.1467-9221.2008.00622.x

Harwood, J., Hewstone, M., Paolini, S., & Voci, A. (2005). Grandparent-grandchild contact and attitudes toward older adults: Moderator and mediator effects. *Personality and Social Psychology Bulletin*, **31**, 393–406. doi:10.1177/0146167204271577

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: Guilford.

Hummert, M. L., Garstka, T. A., Shaner, J. L., & Strahm, S. (1994). Stereotypes of the elderly held by young, middle-aged, and elderly adults. *Journals of Gerontology: Psychological Sciences*, **49**, 240–249. doi:10.1093/geronj/49.5.P240

Iyengar, S. (1987). Television news and citizens' explanations of national affairs. *The American Political Science Review*, **81**, 815–832.

Iyengar, S. (1991). *Is anyone responsible?: How television frames political issues*. Chicago, IL: University of Chicago Press.

Iyengar, S. (2011). *Media politics: A citizen's guide* (Vol. ,2nd ed.). New York, NY: W. W. Norton.

Kite, M. E., & Johnson, B. T. (1988). Attitudes toward older and younger adults: A meta-analysis. *Psychology and Aging*, **3**, 233–244. doi:10.1037/0882-7974.3.3.233

Kite, M. E., Stockdale, G. D., Whitley, B. E., & Johnson, B. T. (2005). Attitudes toward younger and older adults: An updated meta-analytic review. *Journal of Social Issues*, **61**, 241–266. doi:10.1111/j.1540-4560.2005.00404.x

Lefevere, J., Swert, K. D., & Walgrave, S. (2012). Effects of popular exemplars in television news. *Communication Research*, **39**, 103–119. doi:10.1177/0093650210387124

Levy, B. R. (2003). Mind matters: Cognitive and physical effects of aging self-stereotypes. *Journal of Gerontology*, **54B**, 203–211. doi:10.1093/geronb/58.4.P203

Marmor, T. R., Cook, F. L., & Scher, S. (1997). Social security politics and the conflict between generations: Are we asking the right questions?. In E. R. Kingson & J. H. Schulz (Eds.), *Social Security in the 21st century* (pp. 195–207). New York, NY: Oxford University Press.

Moyer-Gusé, E. (2008). Toward a theory of entertainment persuasion: Explaining the persuasive effects of entertainment-education messages. *Communication Theory*, **18**, 407–425. doi:10.1111/j.1468-2885.2008.00328.x

Nelson, T. E., & Kinder, D. R. (1996). Issue frames and group centrism in American public opinion. *Journal of Politics*, **58**, 1055–1078.

Niederdeppe, J., Shapiro, M. A., & Porticella, N. (2011). Attributions of responsibility for obesity: Narrative communication reduces reactive counterarguing among liberals. *Human Communication Research*, **37**, 295–323. doi:10.1111/j.1468-2958.2011.01409.x

O'Keefe, D. J., & Jensen, J. D. (2007). The relative persuasiveness of gain-framed loss-framed messages for encouraging disease prevention behaviors: A meta-analytic review. *Journal of Health Communication*, **12**, 623–644. doi:10.1080/10810730701615198

Ortiz, M., & Harwood, J. (2007). A social cognitive theory approach to the effects of mediated intergroup contact on intergroup attitudes. *Journal of Broadcasting and Electronic Media*, **51**, 615–631. doi:10.1080/08838150701626487

Paolini, S., Harwood, J., & Rubin, M. (2010). Negative intergroup contact makes group memberships salient: Explaining why intergroup conflict endures. *Personality and Social Psychology Bulletin*, **36**, 1723–1738. doi:10.1177/0146167210388667

Perry, S. D., & Gonzenbach, W. J. (1997). Effects of news exemplification extended: Considerations of controversiality and perceived future opinion. *Journal of Broadcasting & Electronic Media*, **41**, 229–244. doi:10.1080/08838159709364403

Preacher, K. J., & Kelley, K. (2011). Effect size measures for mediation models: Quantitative strategies for communicating indirect effects. *Psychological Methods*, **16**, 93–115. doi:10.1037/a0022658

Scheufele, D. A. (2000). Agenda-setting, priming, and framing revisited: Another look at cognitive effects of political communication. *Mass Communication & Society*, **3**, 297–316. doi:10.1207/S15327825MCS0323_07

Scheufele, D. A., & Tewksbury, D. (2007). Framing, agenda setting, and priming: The evolution of three media effects models. *Journal of Communication*, **57**, 9–20. doi:10.1111/j.0021-9916.2007.00326.x

Schiappa, E., Gregg, P. B., & Hewes, D. E. (2005). The parasocial contact hypothesis. *Communication Monographs*, **77**, 92–115. doi:10.1080/0363775052000342544

Semetko, H. A., & Valkenburg, P. M. (2000). Framing European politics: A content analysis of press and television news. *Journal of Communication*, **50**, 93–109. doi:10.1111/j.1460-2466.2000.tb02843

Shah, D. V., Kwak, N., Schmierbach, M., & Zubric, J. (2004). The interplay of news frames on cognitive complexity. *Human Communication Research*, **30**, 102–120. doi:10.1111/j.1468-2958.2004.tb00726.x

Signorielli, N. (2004). Aging on television: Messages relating to gender, race, and occupation in prime time. *Journal of Broadcasting and Electronic Media*, **48**, 279–301. doi:10.1207/s15506878jobem48027

Silverstein, M., & Parrott, T. M. (2001). Attitudes toward government policies that assist informal caregivers: The link between personal troubles and public issues. *Research on Aging*, **23**, 349–374. doi:10.1177/0164027501233004

Slater, M. D., & Rouner, D. (2002). Entertainment-education and elaboration likelihood: Understanding the processing of narrative persuasion. *Communication Theory*, **12**, 173–191. doi:10.1111/j.1468-2885.2002.tb00265.x

Social Security Administration (SSA), Board of Trustees. (2011). *The 2011 annual report of the board of trustees of the federal old-age and survivors insurance and federal disability insurance trust funds* (*71*). Retrieved from http://www.ssa.gov/OACT/TR/2011/tr2011.pdf

Tran, H. (2012). Exemplification effects of multimedia enhancements. *Media Psychology*, **15**, 396–419. doi:10.1080/15213269.2012.723119

Wright, S. C., Aron, A., McLaughlin-Volpe, T., & Ropp, S. A. (1997). The extended contact effect: Knowledge of cross-group friendships and prejudice. *Journal of Personality and Social Psychology*, **73**, 73–90. doi:10.1037/0022-3514.73.1.73

Zillmann, D. (1999). Exemplification theory: Judging the whole by some of its parts. *Media Psychology*, **1**, 69–94. doi:10.1207/s1532785xmep0101_5

Zillmann, D., & Brosius, H. B. (2000). *Exemplification in communication: The influence of case reports on the perception of issues*. Mahwah, NJ: Erlbaum.