Research Article

ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Psychological Science
1–13

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

Eye Movements Predict Large-Scale Voting Decisions

TC

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

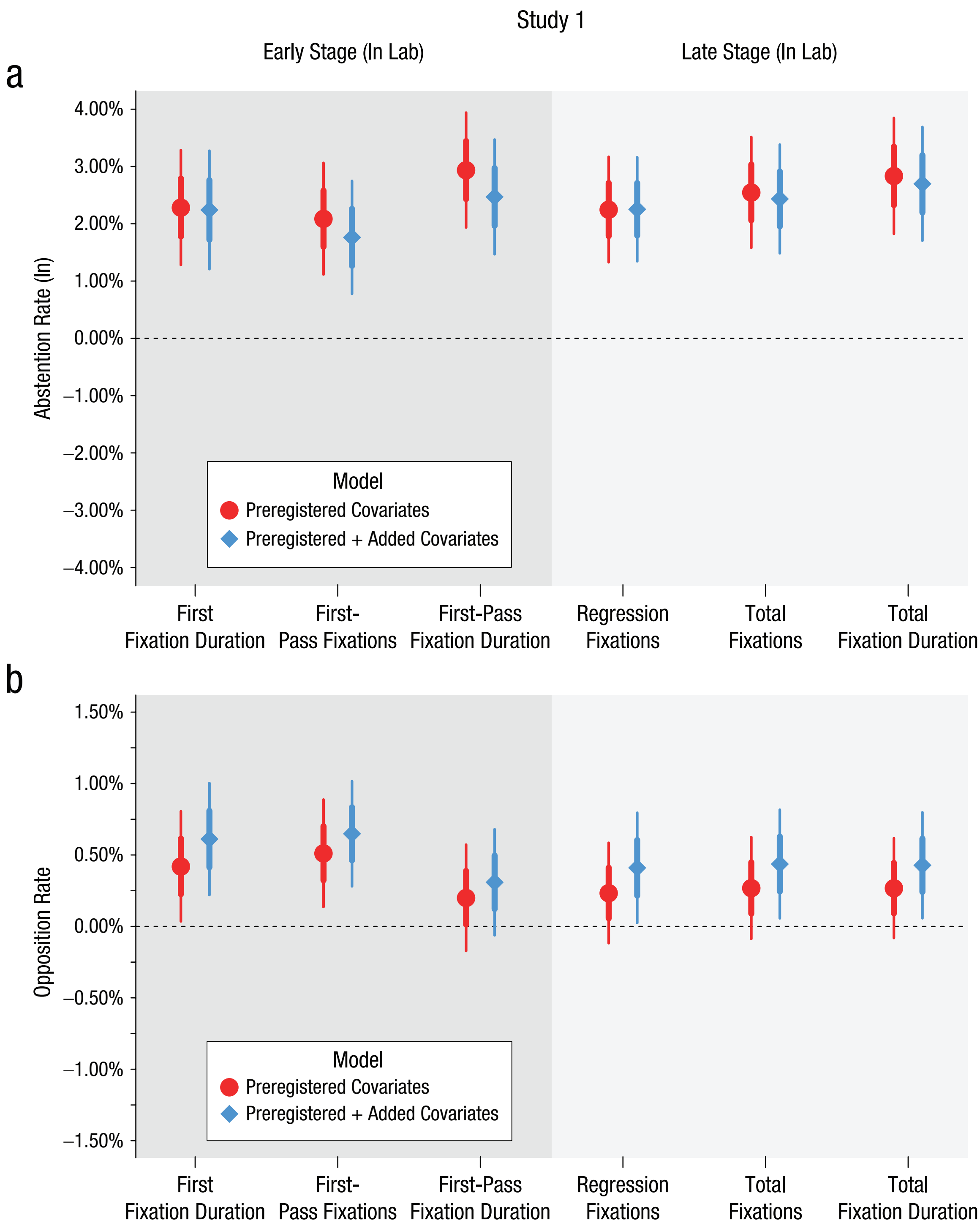

Fig. 1. Effect of a 1-*SD* increase on aggregate voting decisions in actual elections (Study 1; lab data were collected after actual voting occurred). The graphs show (a) the average abstention rate and (b) the average opposition rate for both the preregistered and alternative models, separately for 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 $ps > .10$).

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/).

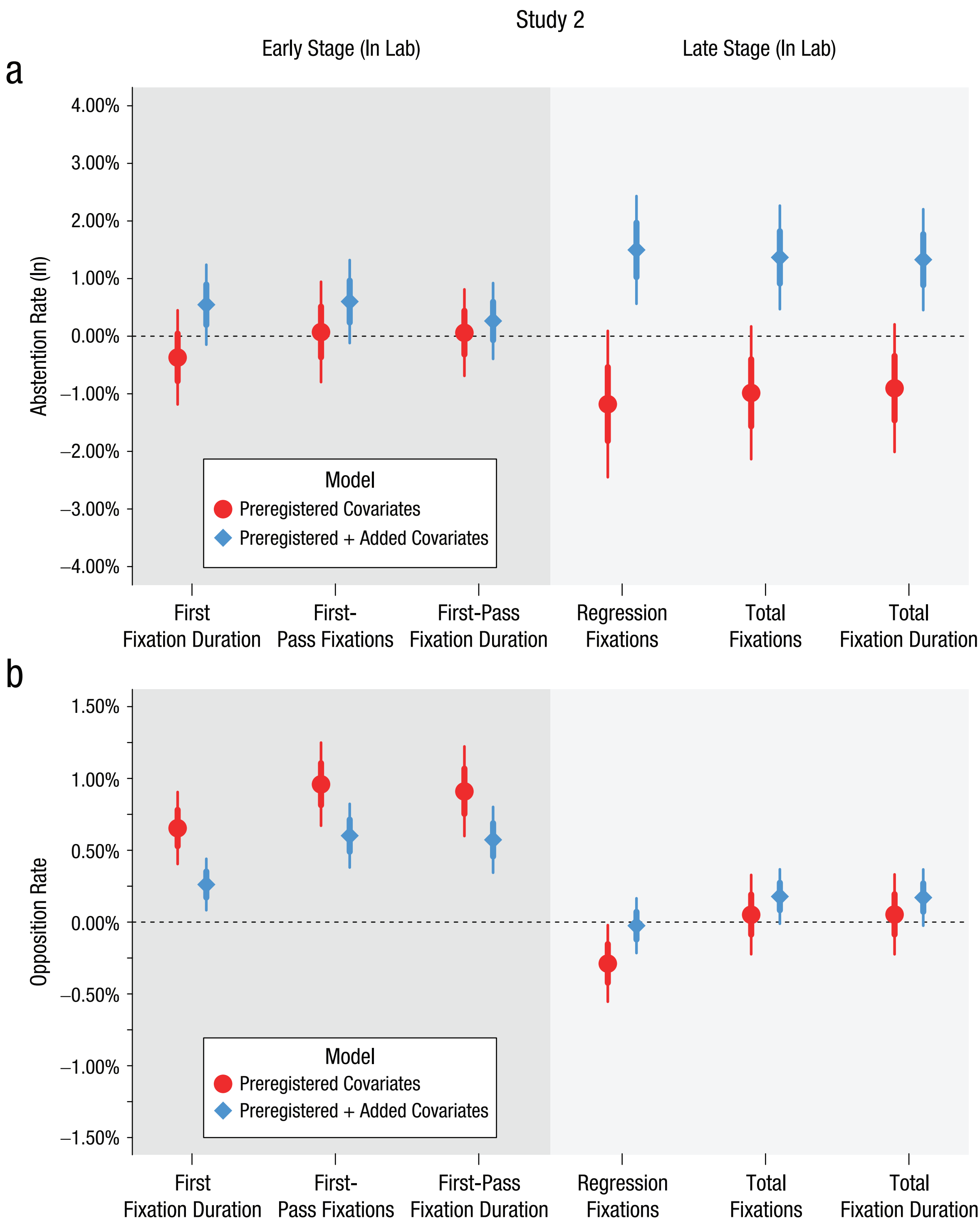

Fig. 2. Effect of a 1-*SD* increase on aggregate voting decisions in actual elections (Study 2; lab data were collected before actual voting occurred). The graphs show (a) the average abstention rate and (b) the average opposition rate for both the preregistered and alternative models, separately for 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.

TC

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.