Are Ballot Titles Biased?
Partisanship in California's Supervision of Direct Democracy

Christopher S. Elmendorf and Douglas M. Spencer*

This study investigates whether and if so under what conditions the California attorney general (AG), who authors the ballot title and summary (label) for statewide ballot initiatives, writes ballot language that is biased rather than impartial. State law demands an impartial label, but commentators frequently complain that the AG chooses misleading language to bolster (undermine) measures that the AG or the AG's party supports (opposes). In this Article, using a convenience sample of students from several universities, we measure ordinary observers' perceptions of bias in ballot labels for initiatives dating back to 1974. Separately, we calculate an objective measure of bias using a readability algorithm. We then test hypotheses about AG strategy, examining whether the extent of bias in ballot labels varies with the closeness of the election and the degree to which the measure elicits partisan division. We also examine the correlation between bias perceptions and observer characteristics such as support for the ballot measure, trust in government, and social trust.

INTRODUCTION

The brainchild of early twentieth-century Progressives, the ballot initiative was supposed to enable ordinary citizens to wrest control of lawmaking from elected officials whom the Progressives saw as marionettes of powerful interest groups and corrupt party bosses. Whether the existence of the ballot initiative actually results in greater congruence between state law and median voter preferences is a subject of ongoing study and debate.[1] It is clear, however, that the political insiders whom the ballot initiative was meant to check are not passive bystanders. Government officials have sometimes parried initiative proponents by putting topically related referendum measures on the same ballot (potentially confusing voters), by refusing to appropriate funds for the implementation of enacted ballot measures, and by narrowing enacted measures through the courts.[2]

Government officials may also be able to bolster or undermine a proposed initiative by controlling how the measure is characterized on the ballot and in the state-issued ballot pamphlet. In California, the attorney general (AG)—a partisan, elected official—authors a brief ballot label describing each measure that has qualified for the ballot, and a somewhat longer summary for the ballot pamphlet, which the state mails free of charge to all registered voters. The nonpartisan Legislative Analyst's Office (LAO) prepares a much more detailed description of the measure and its likely effects, which appear in the ballot pamphlet following the AG's summary.

State law directs the AG to write evenhanded, nonargumentative ballot labels and ballot-pamphlet summaries, but critics say AGs often flaunt this duty, crafting biased summaries to improve or diminish a measure's odds of passage. The LAO has faced much less criticism. If the critics are right and if the LAO is as nonpartisan as its name and reputation suggest, then responsibility for writing ballot labels and ballot-pamphlet summaries arguably should be transferred from the AG to the LAO or a similarly nonpartisan official.[3] A bill to this effect was introduced in the California Assembly in 2009.[4]

---

    1.   *See, e.g.*, JOHN G. MATSUSAKA, FOR THE MANY OR THE FEW: THE INITIATIVE, PUBLIC POLICY, AND AMERICAN DEMOCRACY (Benjamin I. Page ed., 2004); Jeffrey R. Lax & Justin H. Phillips, *The Democratic Deficit in the States*, 56 AM. J. POL. SCI. 148 (2012).

    2.   *See* ELISABETH GERBER ET AL., STEALING THE INITIATIVE: HOW STATE GOVERNMENT RESPONDS TO DIRECT DEMOCRACY 19 (Paul S. Herrnson ed., 2001).

    3.   State law presently provides that if the AG is the proponent (sponsor) of an initiative, then the AG's ordinary responsibilities with respect to that measure shall be performed by the Legislative

The present study provides the first empirical evidence regarding AG partisanship (or its absence) in the writing of ballot labels. We create and analyze objective and subjective measures of ballot-label bias, looking for the patterns we would expect to find if the AG strategically manipulated ballot labels.

Our objective measure is based on the label's readability. We argue that, with respect to propositions that are competitive and likely to polarize voters by party affiliation and reading level, Democratic AGs have incentives to write exceptionally easy-to-read labels and Republican AGs have incentives to write exceptionally difficult labels. But the data do not provide much support for these hypotheses. With respect to competitive measures, Republican AGs wrote significantly more difficult-to-read labels for very conservative measures than for centrist measures. But there is no evidence that Republican AGs manipulated readability on liberal measures, or that Democratic AGs manipulated readability at all. Labels authored by Democratic AGs were actually more difficult to read, on balance, than labels authored by Republicans.

Our subjective measure of ballot-label bias is based on university students' perceptions. The students were placed behind a veil of ignorance regarding the author of the ballot label, whether the measure passed or failed, etc. We gave students a one- to three-page description of the measure, which we said had been prepared by a "neutral, disinterested expert." In point of fact, subjects received the analysis prepared for the ballot pamphlet by the nonpartisan legislative analyst. We then asked students to read a proposed "ballot title and summary" (the actual ballot label), and to evaluate whether the title and summary was biased in several different ways. Subjects who reported bias were asked whether the bias favored a yes or no vote. Approximately ten students evaluated each label.

We model the existence and direction of perceived bias as a function of both the individual characteristics of the survey respondents and various political factors that we hypothesize may affect the AG's incentive to write a biased label. We find that bias perceptions are significantly correlated with the survey respondents' support for or opposition to the measure, but we do not obtain statistically significant results with respect to the political covariates. The sign of the coefficient on the key political interaction term is consistent with our strategic partisanship hypothesis, but we cannot reject the null hypothesis that this is due to chance.

Counsel instead. CAL. ELEC. CODE § 9003 (West 2003). However, the courts have allowed the AG to write ballot labels for measures that the AG takes a very active role in promoting or opposing, so long as the AG is not technically the "proponent," i.e., the person who initially files the proposed measure with the state, seeking a title and summary for the circulating petition. *See, e.g.*, Lungren v. Superior Court, 48 Cal. App. 4th 435, 440 n.1 (1996) (rejecting challenge to AG-titled measure for which the AG authored the "official Rebuttal to Arguments Against [the Measure]" in the ballot pamphlet, and rejecting plaintiffs' argument that at the very least the AG was not owed deference on the ballot label given his active role in promoting the measure).

    4.   Assemb. B. 319, 2009–2010 Reg. Sess. (Cal. 2009).

I. BACKGROUND

This Part motivates our study by briefly explaining the AG's legal responsibilities and duties in the ballot-initiative process, as well as the political science literature regarding the influence of ballot labels on voter decision making in initiative and referendum elections.

A. The Law and Politics of Ballot Labels

California delegates responsibility for producing the brief descriptions of ballot measures that citizens encounter when performing their direct democracy agenda setting and enactment functions to the attorney general. Proponents who wish to put a measure on the ballot first submit the text of the measure to the AG and pay a filing fee. Based on the submitted text, the AG writes a title and summary for the "circulating petition," the document that proponents must use to gather the signatures necessary to qualify the measure for the ballot.[5] For each measure that qualifies, the AG prepares a "title and summary" and a shorter "ballot label" of up to seventy-five words.[6] The title and summary introduce the measure in the state-issued ballot pamphlet, which is mailed to all registered voters shortly before voting begins.[7] Voters who dig deeper into the pamphlet also find a detailed analysis of the measure prepared by the nonpartisan Legislative Analyst's Office, as well as arguments for and against the measure submitted by activists on either side.[8]

On the ballot, initiatives are assigned a number and described by the AG-authored ballot label.[9] The label must include the measure's expected fiscal impact, as projected by the legislative analyst,[10] but everything else falls to the AG's discretion. The AG's discretion is not unfettered, however. The AG is bound by law to "give a true and impartial statement of the purpose of the measure," and not to use language that constitutes an "argument" or that is "likely to create prejudice, for or against the proposed measure."[11] The AG's job is to

---

5. CAL. ELEC. CODE §§ 9004–06 (West 2003).

6. *Id.* § 9051.

7. *Id.* § 9094.

8. *Id.* § 9086(c).

9. *Id.* §§ 9050–53.

10. *Id.* § 9051(2)(b).

11. *Id.* § 9051(2)(c). While this command explicitly applies only to the "ballot title and summary" (i.e., the description in the ballot pamphlet), it is clear by implication and from other provisions in the California Election Code that it also applies to the circulating title and summary, and to the ballot label. *See id.* Section 9051(2)(b) states that the ballot label is to be a "condensed version of the ballot title and summary . . . ." *Id.* § 9051(2)(b). As for the circulating petition title and summary, it "shall be prepared in the manner provided for the preparation of ballot titles and summaries in Article 5 (commencing with Section 9050)[.]" *Id.* § 9004(a).

Moreover, in an unpublished opinion, the California Court of Appeal treated the duty of impartiality in the labeling of ballot measures as derived from state constitutional provisions concerning free elections, and, as such, binding on any state actor to whom the legislature might

"reasonably inform the voters of the character and purpose of the proposed measure," not to be an advocate for or against it.[12]

Often, however, the AG has come under attack for crossing the line between information and advocacy. For example, a *Sacramento Bee* columnist wrote that Democratic AG Kamala Harris's circulating titles and summaries for a pair of pension reform measures were "nonsense" that "got many on the right screaming and even partisans on the left privately squirming."[13] Even the liberal *Fresno Bee* complained that the "official description of the two measures read like talking points taken straight from a public employee union boss' campaign handbook"— and that these talking points were "not true."[14] Initiative proponents blamed Harris's "ugly, partisan, and manipulative"[15] language when they withdrew their petition, claiming that "the Attorney General's false and misleading title and summary makes [sic] it nearly impossible to pass."[16]

Attacks have sounded in the courtroom too, but the courts have not been especially receptive. While California's courts review ballot labels and voter-pamphlet language for compliance with the statutory standard, and while they have been willing to order specific changes in the event of a successful challenge,[17] the standard of review is deferential. If "'reasonable minds may differ as to the sufficiency'" of AG-prepared ballot materials, the legal challenge will be rejected.[18] "'[O]nly in a clear case should . . . [the AG's materials] be held insufficient.'"[19] The normative touchstone is whether the label risks "misleading the public with inaccurate information."[20]

Beyond restating the standard of review, it is not easy to characterize what the courts generally look for and do when adjudicating these challenges. The

---

delegate this responsibility (including the legislature itself). Clark v. Superior Court, No. C064430, 2010 WL 928384, at 5–6 (Cal. Ct. App. Mar. 16, 2010).

12.   Yes on 25, Citizens for an On-Time Budget v. Superior Court, 189 Cal. App. 4th 1445, 1452 (2010).

13.   George Skelton, *Bias, Low Fee Make Monster of Initiative System*, SACRAMENTO BEE (Feb. 20, 2012, 4:55 PM), http://www.sacbee.com/2012/02/20/4H7896/skelton-bias-low-fee-make-monster .html.

14.   Editorial, *It's Up to Brown to Get Pension Reform Results*, FRESNO BEE, Feb. 15, 2012, at A15.

15.   Editorial, *Ballot Measures Summaries Should Be Written by Neutral Party*, CONTRA COSTA TIMES (Feb. 25, 2012), http://www.contracostatimes.com/ci_200040274.

16.   Patrick McGreevy, *GOP-led Drive for California Pension Initiative Dead for This Year*, L.A. TIMES (Feb. 8, 2012, 3:44 PM), http://latimesblogs.latimes.com/california-politics/2012/02/gop-led -drive-for-california-pension-initiative-dead.html.

17.   *See, e.g.*, *Clark*, 2010 WL 928384, at 9 (Cal Ct. App. Mar. 16, 2010) (finding the verb "reforms" to be argumentative in the context of a ballot label for an open-primary measure, and ordering its replacement with "changes").

18.   Amador Valley Joint Union High Sch. Dist. v. State Bd. of Equalization, 583 P.2d 1281, 1298 (1978) (quoting Epperson v. Jordan, 82 P.2d 445, 448 (1938)); *Yes on 25*, 189 Cal. App. 4th at 1453.

19.   *Yes on 25,* 189 Cal. App. 4th at 1453 (quoting *Epperson*, 82 P.2d at 448).

20.   Lungren v. Superior Court, 48 Cal. App. 4th 435, 440 (1996) (quoting *Amador Valley*, 583 P.2d at 1298).

litigation occurs preelection and in a brief window of time. Few cases make it to the court of appeal. Since the mid-1970s, the Superior Court of Sacramento County has had exclusive jurisdiction over preelection challenges to the AG's handiwork,[21] and this court's opinions and transcripts are not available through any electronic database. It may be helpful, however, to provide a few examples of recent disputes and their resolution by the courts.[22]

Proposition 32 (2012). Proposition 32 would have disallowed unions and corporations from donating money directly to candidates for elective office and from spending money raised via payroll deductions on any political purpose.[23] The original ballot label produced by Attorney General Kamala Harris described Proposition 32 as "restricting" union contributions to candidates.[24] Proponents sued, arguing that "'[v]oters deserve to be informed that Prop. 32 doesn't just reduce direct contributions from corporations and unions to politicians, it eliminates them entirely.'"[25] The superior court agreed, ordering the word "restricts" replaced with "prohibits."[26]

Proposition 25 (2010). For many years, the California Constitution required budgets to be adopted by a two-thirds vote.[27] Proposition 25 changed the two-thirds requirement to a simple majority requirement. The ballot title and label prepared by Attorney General Jerry Brown stated that Proposition 25 "retains two-thirds vote requirement for taxes." The Chamber of Commerce sued, arguing that this would mislead voters into believing that they had to pass the measure in order to retain the supermajority rule for taxes. The superior court concurred, but its judgment was vacated on appeal.[28] Pointing to dictionary definitions of the word "retain," the court of appeal concluded that most voters probably would not interpret the title as the Chamber of Commerce and the lower court had.[29] The Court also said the AG's decision to state in the label that the measure did not apply to taxes was reasonable "because debates regarding changes to California's budget process routinely include a discussion of whether the vote requirement for raising taxes should be lowered."[30]

---

21.   CAL. GOV'T CODE § 88006 (West 2005); CAL. ELEC. CODE § 9092 (West 2003).

22.   The following are drawn mostly from appellate opinions available on Lexis and Westlaw.

23.   *Proposition 32*, LEGISLATIVE ANALYST'S OFFICE 4 (Nov. 6, 2012), http://www.lao.ca.gov/ballot/2012/32_11_2012.pdf.

24.   *Judge Orders Changes to Prop. 32 Language*, L.A. TIMES (Aug. 13, 2012, 6:56 PM), http://latimesblogs.latimes.com/california-politics/2012/08/judge-orders-changes-prop-32-language.html.

25.   *Id.*

26.   *Id.*; *see* Titus v. Bowen, No. 34-2012-80001218 (Super. Ct. Sacramento Cnty. Aug. 13, 2012) (order granting in part and denying in part peremptory writ of mandate).

27.   For some background on the California budget process and potential reforms, see Ethan J. Leib & Christopher S. Elmendorf, *Why Party Democrats Need Popular Democracy and Popular Democrats Need Parties*, 100 CALIF. L. REV. 69, 100–07 (2012).

28.   Yes on 25, Citizens for an On-Time Budget v. Superior Court, 189 Cal. App. 4th 1445, 1452 (2010).

29.   *Id.* at 1454.

30.   *Id.*

Proposition 14 (2010). Proposition 14 established California's nonpartisan "top-two" primary election system by constitutional amendment. The amendment was put on the ballot by the legislature, rather than by a private citizen or interest group, and in passing Proposition 14 the legislature also specified that the following label was to be used to describe it on the ballot:

> ELECTIONS. PRIMARIES. GREATER PARTICIPATION IN ELEC-TIONS. Reforms the primary election process for congressional, statewide, and legislative races. Allows all voters to choose any candidate regardless of the candidate's or voter's political party preference. Ensures that the two candidates receiving the greatest number of votes will appear on the general election ballot regardless of party preference.[31]

After determining that the legislature is bound by the same obligations of impartiality as the AG, the superior court and the court of appeal held the label improper. The superior court replaced the title phrase "greater participation in elections" with "increases right to participate in primary elections," and added the fiscal impact statement required of ballot labels authored by the AG.[32] The court of appeal deemed this fix inadequate because it left the argumentative word "reform" in the label.[33] The Court quoted dictionaries to show that "reform" connotes melioration, and ordered the word replaced with "change."[34] But the Court rejected opponents' argument that the title phrase, "increases right to participate in primary elections," was argumentative or misleading.[35]

Proposition 8 (2008). Proposition 8, which qualified for the ballot shortly after the California Supreme Court recognized a constitutional right to same-sex marriage, amended the state constitution to limit marriage to opposite-sex couples. The AG's ballot label stated in relevant part:

> ELIMINATES RIGHT OF SAME-SEX COUPLES TO MARRY. INITIATIVE CONSTITUTIONAL AMENDMENT. Changes the California Constitution to eliminate the right of same-sex couples to marry. Provides that only marriage between a man and a woman is valid or recognized in California.[36]

Proponents had argued for the following:

---

31.   S.B. 19, 2009–2010 3d Extended Sess. (Cal. 2009).
32.   *Yes on 25*, 189 Cal. App. 4th at 1451–52.
33.   *Id.* at 1453–54.
34.   *Id.* It also pointed to an earlier case concerning a ballot measure that would increase city taxes on a utility plant. The title as originally written was, "Amendment of Utility Tax by Removing Electric Power Plant Exemption." The court determined that the word "exemption," "particularly in the tax context," "connote[s] unfair influence and special treatment," and it ordered the city to write "exclusion" in place of "exemption." *Id.* at 8 (discussing and quoting Huntington Beach City Council v. Superior Court, 94 Cal. App. 4th 1417, 1433–34 (2002)).
35.   *Id.* at 1454.
36.   Jansson v. Bowen, No. 34-2008-00017351, at *3 (Super. Ct. Sacramento Cnty. Aug. 7, 2008), *available at* http://ag.ca.gov/cms_attachments/press/pdfs/n1597_ruling_on_proposition_8.pdf.

> MARRIAGE. CONSTITUTIONAL AMENDMENT. Amends the California Constitution to provide that only marriage between a man and a woman is valid or recognized in California.[37]

Proponents contended that the AG's label was argumentative and prejudicial because it used "a strongly negative, active tense verb[—'eliminates'—]to characterize the effect of the measure."[38] The court disagreed.

 Proposition 209 (1996). The ballot label for Proposition 209 read:

> PROHIBITION AGAINST DISCRIMINATION OR PREFEREN-TIAL TREATMENT BY STATE AND OTHER PUBLIC ENTITIES. INITIATIVE CONSTITUTIONAL AMENDMENT. Generally prohibits discrimination or preferential treatment based on race, sex, color, ethnicity, or national origin in public employment, education, and contracting . . . .[39]

 Opponents maintained that the label was misleading because the true purpose of Proposition 209 was to end affirmative action by state and local governments, not to end discrimination.[40] In public statements, newspaper editorials, and arguments in the ballot pamphlet, proponents had characterized the measuring as "end[ing] affirmative action."[41] Relying on such extrinsic evidence of intent, the superior court agreed with the plaintiffs.[42] But the court of appeal reversed, stating that "the title, summary and label provided by the Attorney General are essentially verbatim recitations of the operative terms of the measure," using words that "are all subject to common understanding."[43]

B. Do Ballot Labels Matter?

 The fact that proponents and opponents of ballot measures litigate seemingly small differences in the language of the label suggest that they believe these details may affect whether a measure passes or fails. Their beliefs find considerable support in the political science literature.

 In the study most directly on point, Craig Burnett and Vladimir Kogan conducted survey experiments on a national sample of voting age adults to examine the effects of alternative ballot labels on vote intentions.[44] Depending on the treatment condition, subjects viewed either the official ballot label or a

---

 37. Letter from Andrew Pugno, Attorney for Proponents of Proposition 8, to Attorney General Initiative Coordinator (Oct. 15, 2007) (on file with authors).

 38. *Jansson*, No. 34-2008-00017351, at *4.

 39. All ballot labels discussed in this Article are on file with the authors.

 40. Lungren v. Superior Court, 48 Cal. App. 4th 435, 441–42 (1996).

 41. *Id.*

 42. *Id.*

 43. *Id.*

 44. Craig M. Burnett & Vladimir Kogan, The Case of the Stolen Initiative: Were the Voters Framed? (June 15, 2012) (unpublished manuscript), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1643448.

plausible alternate label, and subjects either received or did not receive prominent interest group endorsements. The measures under study included California's same-sex marriage ban, a Colorado anti-abortion measure, and a San Diego schools bond.[45] In the no-endorsements condition, Burnett and Kogan found that the alternate label resulted in an eight percentage-point swing in support for the abortion and same-sex marriage measures, but had no effect on the school bond measure.[46] Providing respondents with interest group endorsements reduced the magnitude of the label effect by about fifty percent,[47] but even a four percentage-point swing in reported vote intentions is quite striking.

One must be cautious about extrapolating from Burnett and Kogan's study to the real world. Their subjects included nonvoters as well as voters, and subjects were asked how they would vote in either an informational vacuum or an informational environment containing a single endorsement that the researchers (rather than the subjects) deemed relevant. In the real world of ballot initiative elections, voters may acquire information from political advertising, the ballot pamphlet, or friends and neighbors—information that makes them less sensitive to variations in ballot-label wording. Then again, most voters do not pay much attention to politics,[48] and the political economy of getting a measure on the ballot tends to result in the selection of measures that provide concentrated benefits for proponents and diffuse costs for everyone else.[49] These measures do not engender much campaign spending on the no side.[50] As a consequence, voters are unlikely to learn many endorsements that counsel in favor of a no vote, making them quite dependent on the ballot label.[51]

Other findings generally corroborate the intuition that voters rely on ballot

---

45.   The alternative labels were realistic and fairly chosen. In the case of the same-sex marriage ban, the alternative label was the title and summary on the circulating petition. In the case of the abortion measure, the alternative label was the label used for a similar measure in another state. In the case of the school bond measure, the alternative stated that the increase in property taxes that would be necessary to service the debt. *Id.* at 15–18.

46.   *Id.* at 22–23.

47.   *Id.*

48.   *See generally* Christopher S. Elmendorf & David Schleicher, *Informing Consent: Voter Ignorance, Political Parties, and Election Law* (UC Davis Legal Studies Research Paper Series No. 285, 2012) (examining how law can influence generally uninformed voters to obtain meaningful representation).

49.   Thad Kousser & Mathew D. McCubbins, *Social Choice, Crypto-Initiatives, and Policymaking by Direct Democracy*, 78 S. CAL. L. REV. 949, 951–57 (2005).

50.   *Id.*

51.   Surveys of California voters done in the early 1990s found that fifty-four percent of California voters *say* that they rely on the ballot pamphlet and that, of the voters who read pamphlets, eighty to ninety percent reported that the pro/con arguments or names of endorsers (presented in pro/con arguments) were especially helpful. SEAN BOWLER & TODD DONOVAN, DEMANDING CHOICES: OPINION, VOTING, AND DIRECT DEMOCRACY 55–59 (1998). However, these numbers may be exaggerated (respondents who wish to present themselves as "good" citizens may overreport reading the ballot pamphlet, just as they overreport voting). And even read in the most favorable light, they suggest that about fifty percent of voters *do not* glean useful information about endorsements from the ballot pamphlet.

labels in initiative and referendum elections. Using opinion-poll data, Sean Bowler and Todd Donovan demonstrated that self-interest explained much more of the variation in vote intentions on a school voucher ballot measure when respondents were provided with the proposition's ballot description and name, rather than the name alone.[52] Without an accurate label, many voters were unable to derive a position from their interests.

In an exit-poll study of voting on a California renewable energy measure, Craig Burnett, Elizabeth Garrett, and Mathew McCubbins show that voters' factual knowledge about the measure and awareness of endorsements had essentially no impact on vote choice.[53] Regardless of knowledge or endorsements, there was strong support for the proposition among voters who said they supported renewable energy even if electricity rates may rise, and strong opposition among those who disagreed.[54] A possible explanation is that the ballot label—which everyone sees at the moment they vote—swamps other influences on vote choice.[55] Burnett has since replicated these findings with exit polls on seven other ballot initiatives.[56]

Two other recent studies speak to the potential importance of ballot labels for voting in direct democracy. Michael Binder examined twenty-two ballot measures and found that anywhere from eighteen percent to more than fifty percent of voters acknowledged being confused about the measure.[57] For seven of the measures, Binder asked voters about their policy preferences with respect to the subject of the measure, and inferred the "correct vote" for each respondent. The percentage of incorrect votes cast by respondents who acknowledged being confused was five to fifteen points higher than the rate among nonconfused voters. With respect to five of the seven measures, the erroneous votes cast by confused voters basically canceled one other out, but the confused erred systematically in favor of voting yes on the other two measures.[58]

---

52. *Id.* at 55–59.

53. Craig M. Burnett et al., *The Dilemma of Direct Democracy*, 9 ELECTION L.J. 305 (2010).

54. *Id.* at 314–17.

55. *Id.*

56. Craig M. Burnett, Informed Democracy? How Voter Knowledge of Initiatives Influences Consistent Voting (2009) (unpublished manuscript), *available at* http://www.olemiss.edu/depts/ political_science/state_politics/conferences/2009/papers/20.pdf. Some caveats are in order. First, Burnett's measure of "knowledge of endorsements" is based on one or two endorsements *he* deems most useful for voting on each measure. It is possible that voters relied on other endorsements, knowledge of which Burnett did not measure. Second, Michael Binder's study of some of the same ballot measures finds that endorsement knowledge is correlated with voting correctly (i.e., in accordance with the voter's policy preference). *See* Michael M. Binder, Getting it Right or Playing it Safe? Confusion, the Status Quo Bias and Correct Voting in Direct Democracy 126–27 (2010) (unpublished Ph.D. dissertation, UC San Diego).

57. Binder, *supra* note 56, at 46–76.

58. *Id.* at 128–30.

Binder shows that voter confusion is correlated with individual-level attributes such as interest in politics, relevant factual knowledge, and knowledge of endorsements,[59] and also strongly related to the "difficulty" of the issue.[60] He did not examine ballot-label effects, but it is plausible that complex ballot labels contribute to voter confusion. It also seems likely that confused voters are susceptible to being swayed by argumentative, prejudicial, or otherwise inaccurate ballot labels. Certainly if we were in the business of drafting ballot labels to skew the vote, we would be pleased to learn that anywhere from twenty to fifty percent of voters are likely to be confused about the measure labels we drafted, and that these confused voters are much more likely than others to vote against their policy preferences.

In addition to causing "mistaken" votes, a complicated ballot label may lead some voters to skip the measure or "roll-off"[61] the ballot altogether. Studying more than 1200 recent measures, Shauna Reilly and Sean Richey found a positive, statistically significant correlation between voter roll-off and language complexity in the ballot label.[62] Their study relies on aggregate data, and the apparent causal relationship between language complexity and voter abstention may or may not hold up in future studies with individual-level data.[63] But Reilly and Richey's results are at the very least suggestive and, along with Binder's work, they motivate one of the empirical strategies we pursue in this Article.

II. METHODS, HYPOTHESES, AND RESULTS

The principal barrier to studying bias in the labeling of ballot measures is the lack of any agreed-upon method for ascertaining the existence and extent of bias. If there is no room for reasonable disagreement about whether any given ballot label is argumentative, prejudicial, or otherwise likely to mislead voters, there would be no point in litigating the label because the case's outcome would be a foregone conclusion. Yet without a defensible, quantifiable measure of ballot label bias, it would seem impossible to empirically assess the allegation that attorneys general behave as strategic partisan actors when writing labels.

We propose two solutions to this problem. First, following Reilly and

---

59.   *Id.* at 71–76.

60.   *Id.* at 58–62. Binder measured issue difficulty using the classification system developed in Edward G. Carmines & James A. Stimson, *The Two Faces of Issue Voting*, 74 AM. POL. SCI. REV. 78 (1980).

61.   "Roll-off" refers to voters that complete the first part of a ballot but then leave down-ballot races and measures blank. Roll-off is the product of "voter exhaustion" or "voter fatigue" due to various issues such as the complexity, length, or design of a ballot. *See, e.g.,* Charles S. Bullock III & Richard E. Dunn, *Election Roll-Off: A Test of Three Explanations*, 32 URB. AFF. REV. 71 (1996).

62.   Shauna Reilly & Sean Richey, *Ballot Question Readability and Roll-Off: The Impact of Language Complexity*, 20 POL. RES. Q. 1 (2009).

63.   Binder, *supra* note 56, at 88–104 (showing that standard results linking voter confusion to roll-off, which were established with aggregate data, do not hold up with individual-level data).

Richey, we code ballot labels using an objective measure of their readability. We then investigate deviations from the readability norm, asking whether the pattern of deviations is consistent with what one would expect from strategic partisan actors. This empirical strategy does not establish that low-readability (or high-readability) labels are "biased" in the sense of using argumentative or prejudicial language, or failing to convey the central purpose of the measure. Nor does it establish that the average level of readability is optimal or fair. Rather, this strategy does allow us to say whether AGs behave like strategic actors who seek to enable or disable correct voting by citizens with limited reading ability.[64]

Our other strategy relies on ordinary observers' perceptions of bias. Though individual bias evaluations may be erratic (high variance) or distorted by the prejudices of the observer, the average judgment of many observers should approximate the truth if there's a truth to be found. This intuition is grounded in the Condorcet Jury Theorem.[65] So long as each observer's judgment of bias contains some information as well as noise, and the observers' collective errors are not too highly correlated, the average of the observers' bias opinions will converge on the truth as the number of observers grows larger.[66]

The balance of this Part describes each of our methods in some detail, explains the hypotheses we propose to test with each, and presents our results.

A. Readability and AG Strategy

1. Theory and Hypotheses

As noted above, Reilly and Richey found a strong correlation between ballot label readability and voter roll-off.[67] The more complex the ballot label, the greater the likelihood that citizens who voted in the top-of-the-ballot race abstained from voting on the proposition.[68] It seems likely that the roll-off associated with language complexity is more pronounced among voters whose reading ability is limited.[69] This problem represents something of an opportunity for strategic attorneys general: If you do not like the way that low-reading-level (LRL) populations are likely to vote on the initiative if they understand it, write a very complex ballot label. Conversely, if you agree with the majority view among LRL

---

64. Note that our readability scores only capture the readability of the English-language version of the ballot label. Translated versions could differ. We are grateful to Kevin Quinn for raising this issue.

65. Condorcet Jury Theorem is a political science theorem that "establishes that under certain conditions a majority of a group . . . is more likely to choose the 'better' alternative than any one member of the group." Krishna K. Ladha, *The Condorcet Jury Theorem, Free Speech, and Correlated Votes*, 36 AM. J. POL. SCI. 617, 617 (1992).

66. *Id.* at 618–19.

67. Reilly & Richey, *supra* note 62, at 64–65.

68. *Id.*

69. Reilly and Richey do not address this question.

populations, write an exceptionally simple label.[70] These strategies should also work if complicated labels increase incorrect voting rather than roll-off among LRL populations.

An informal model may help to clarify our hypotheses. Assume that the least-cost strategy for the AG is to write a label of ordinary complexity. Extremely simple or extremely complex labels are more costly to write: they require more effort to craft, they may subject the AG to accusations of mischief or impropriety, and they may induce litigation and be rewritten by the courts.[71] Whether a strategic AG is willing to bear these costs depends on the payoff. The payoff is likely to vary with (1) the strength of the AG's own opinion on the issue or that of her political backers, (2) the degree to which voter opinion is polarized on reading-level lines, and (3) the expected closeness of the election.

The logic here is straightforward. The payoff from manipulating language complexity is a function of the payoff from winning the election (factor 1) and the likelihood that the manipulation will be outcome-determinative (factors 2 and 3). The manipulation is more likely to affect the outcome insofar as the election is likely to be close, and insofar as LRL voters who understand the measure all support it or all oppose it, with high-reading-level (HRL) voters taking the contrary position. If, instead, LRL voters split fifty-fifty on the merits of the measure, then the effect of language manipulation inducing them to abstain or to vote randomly would have no effect on whether the measure passes or fails.

Homogeneity of opinion among HRL voters matters too. If voters who agree with the AG are concentrated among HRL populations, then the AG who writes a complex label to disable LRL voters does not have to worry very much about losing votes or inducing error among citizens with whom the AG agrees. Conversely, the AG who writes an exceptionally simple label to help LRL citizens probably forgoes some degree of detail or nuance, which may increase the incorrect voting rate among HRL voters relative to baseline levels associated with typical ballot labels. The greater the polarization of opinion by reading level, the more this marginal increase in incorrect voting by HRL citizens benefits an AG who shares the opinion of LRL citizens.

---

70. This strategy should work so long as more complex labels result in disproportionately greater confusion among LRL voters compared to the electorate as a whole. Even if, as Binder finds, confused voters do not abstain, voter confusion is likely to introduce a random element into choice. As confusion grows, the share of people voting yes among the LRL population is likely to converge on fifty percent. At the limit, this is functionally equivalent to abstention. In either case, the confused have no impact on whether the ballot measure passes or fails.

71. This assumes that there is a positive or U-shaped relationship between language complexity and the probability of judicial invalidation. That proposition is not certain, but it is plausible insofar as the courts in reviewing ballot labels focus on whether the measures are likely to mislead voters. *See supra* Part I.A.

Figure 1: Expected Relationship Between Ballot Label Complexity and Rates
of Voter Abstention and/or Error Among Strong and Weak Readers

Ballot Label Difficulty (higher scores = harder to read)
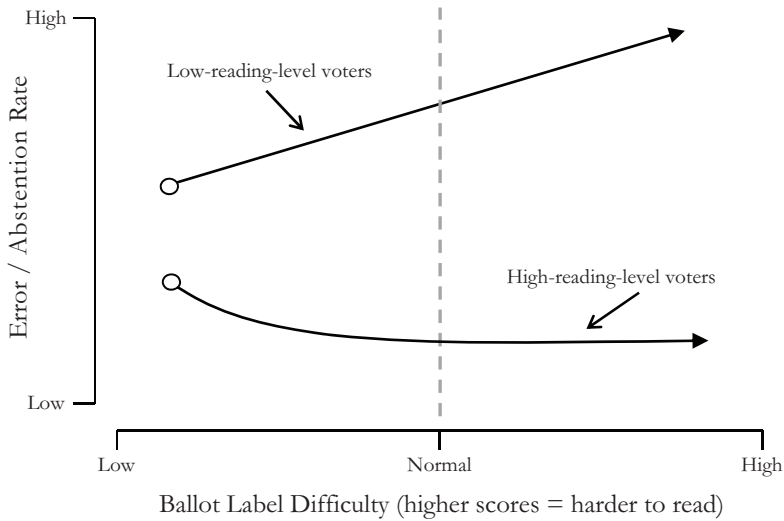
2. Operationalization

To test our hypotheses, we need measures of ballot label readability, the
strength of AG opinion about the proposition and the opinion of the AG's
supporters, the expected closeness of the election at the time the label was drafted,
and the expected polarization of opinion by reading levels at the time of drafting.

Following Reilly and Richey, we code readability using the Flesch-Kincaid
Grade Level formula, which linguists have used for more than half a century to
estimate the number of years of education necessary to read and comprehend a
passage.[72] The formula is based on the average sentence length and average
number of syllables per word in each passage.[73]

As for strength of AG opinion, we cannot observe how much the AG
personally cares about each measure.[74] But we can observe how thoroughly the

---

72.   James N. Farr et al., *Simplification of Flesch Reading Ease Formula*, 35 J. APPLIED PSYCHOL.
333, 334 (1951). We also tried several other measures of readability, none of which led to any
difference in our results (the measures of readability are highly correlated with one another).

73.   The formula is 0.39 × (average sentence length) + 11.8 × (average syllables per word)
− 15.59. *See id.* at 334. The resulting number estimates the number of years of education necessary to
read and comprehend the passage. *See* J. PETER KINCAID ET AL., DERIVATION OF NEW READABILI-
TY FORMULAS (AUTOMATED READABILITY INDEX, FOG COUNT AND FLESCH READING EASE
FORMULA) FOR NAVY ENLISTED PERSONNEL (1975).

74.   We considered trying to create same-scale measures of the ideal points of each AG and
the cutpoints of each ballot measure, using roll-call votes to place AGs (who had served in legislative
bodies), newspaper endorsements to place the measures, and newspaper endorsements on
referendums to bridge the issue spaces. But creating the database of newspaper endorsements would

measure divides voters along party lines. The AG, a partisan official, surely faces some pressure to toe the party line on issues where the parties have polarized. We therefore use partisan polarization of opinion as a rough proxy for how much the AG cares whether a proposition passes or fails.[75]

Ideally our metric of the measure's divisiveness (by party) would be contemporaneous with the drafting of the label. But contemporaneous polling data is not available for all measures, and even if it was, use of polling data would create other problems. For example, opinion polls include nonvoters. Also, partisan divisiveness at the time of drafting may be latent. Voters will not have paid attention to the measure yet, or received signals from party elites. Thus, polls taken at the time of drafting may grossly understate the eventual polarization of the electorate, and the AG, as a forward-looking partisan official, probably cares about elite opinion and anticipates how the electorate will react in the future.

To avoid these problems, we proxy partisan divisiveness using actual votes on the ballot measure. We ranked California's eighty assembly districts by the presidential vote share for each major party. We then identified the six most Republican-leaning and six most Democratic-leaning districts, and for each ballot measure we subtracted the average yes vote in the heavily Democratic districts from the average yes vote in the heavily Republican districts.[76] Finally, we normalized the result. This gives us a rough metric of the measure's "conservativeness."[77] Propositions with respect to which Republican and Democratic districts voted similarly have a conservativeness score of zero. Propositions that received more yes votes in Republican than in Democratic districts have a positive score, whereas propositions that did better in Democratic districts have a negative score. As a proxy for the measure's divisiveness (by party), we use the absolute value of conservativeness.

---

have required an enormous investment of resources, and, in any event, the undertaking would not have yielded ideal points for all of the AGs in our dataset because some neither served in the legislature (or as governor), nor ran against an opponent who had served in these capacities.

75.   It is an imperfect proxy because it does not capture the AG's personal strength of feeling, or the depth of concern on the part of interest groups that have supported the AG and have money to burn.

76.   The partisan divisiveness ratings are robust to small changes in the number of districts included; six is admittedly arbitrary. The scores are nearly identical whether we use four or eight districts.

Table 1: Pairwise Correlation Matrix of Polarization Scores

	Using 4 Districts	Using 6 Districts	Using 8 Districts
Using 4 Districts	1.000		
Using 6 Districts	0.997	1.000	
Using 8 Districts	0.994	0.998	1.000

77.   This can also be understood as a measure of the ballot initiative's "Republicanness," but that term is even more awkward than "conservativeness" so we will use the latter throughout.

This measure of divisiveness is imperfect. It is posttreatment in the sense that we are measuring opinion given the ballot label the AG drafted, rather than the degree of partisan division that would have occurred had the AG drafted a "normal" ballot label. If competitive ballot measures on which the electorate is latently divided by party tend to induce crafty ballot labels, and if these labels tend to dampen the expression of partisan disagreement (for example, by increasing the rate of incorrect voting), then a posttreatment measure of divisiveness by party will understate the latent partisan divide at the time of drafting in some cases.

There is also a potential ecological inference problem with our measures of conservativeness and divisiveness. Because we rely on aggregate data, we cannot be sure that Republicans and Democrats were as divided on the measure as our results suggest. For example, it is possible that some of the apparent divisiveness (as we measure it) is due to relatively conservative Democrats who live in Republican-dominated districts voting for conservative measures and against liberal ones, or relatively liberal Republicans who live in Democratic districts voting more like Democrats. That said, we are fairly confident that our rough measure of divisiveness by party helps to identify measures that divide liberals and conservatives,[78] and other things equal, we expect the AG to care more about and to face more political pressure on such measures.[79] We also take some comfort in the fact that the divisiveness scores remain nearly identical if we substitute precinct-level for district-level election returns.[80]

We proxied the expected competitiveness of ballot measures by calculating the absolute value of the difference in statewide vote totals (yes and no) for each ballot measure. We then inverted the result so that small values represent relatively noncompetitive measures and large values represent highly competitive measures.[81] Finally, we normalized the result. Like our measure of polarization, our measure of competitiveness is posttreatment. The ideal measure would be pretreatment and contemporaneous with the drafting of the label, since strategic manipulation of ballot labels is hypothesized to vary with the expected closeness rather than the actual closeness of the election. But the same obstacles that

---

78.   Many researchers before us have used presidential vote share to proxy the liberalism or conservatism of voters in a legislative district. *See, e.g.*, NOLAN MCCARTY ET AL., POLARIZED AMERICA: THE DANCE OF IDEOLOGY OF UNEQUAL RICHES 63–66 (2006). More importantly, it has recently been shown that presidential vote share is highly correlated with the ideological position of the median voter in California's assembly districts. *See* Seth E. Masket & Hans Noel, *Serving Two Masters: Using Referenda to Assess Partisan Versus Dyadic Representation*, 65 POL. RES. Q. 104 (2012).

79.   Of course, other things are not always equal. The AG may well face pressure on measures that are not ideologically divisive, but have strong support or opposition from high-spending interest groups or individuals. (Indian casino measures are perhaps an example.) In future work, it might be worthwhile to code measures by the degree to which campaign spending is concentrated (just a few donors) or diffuse (lots of donors).

80.   We compared our Assembly-based polarization scores from 2002–2010 to precinct-level data and the polarization scores were correlated at 94.4%.

81.   Competitiveness = abs(yea − nay) × (−1).

prevent us from obtaining a good contemporaneous measure of partisan polarization also prevent us from obtaining a good contemporaneous measure of competitiveness.

As a very rough proxy for the expected degree of polarization by reading level, we used a dummy variable set to one if a measure is redistributive, and zero otherwise. We hypothesize that strategic AGs will generally expect polarization by reading level on redistributive measures, given the well-known correlation between educational attainment and income, and given previous work showing that support for redistributive ballot measures is inversely related to the proportion of college-educated voters in a district.[82]

Each author independently classified each ballot measure as redistributive or not, based on the LAO's description of the measure. Measures were deemed redistributive if, in the coder's opinion, an ordinary voter's attitude toward redistribution from the rich to the poor would have a substantial effect on whether she supports or opposes the measure. Measures were therefore coded as redistributive if they would extract resources disproportionately from higher income/wealth populations and/or confer benefits disproportionately on lower income/wealth populations, *or* if they would do the opposite or erect political barriers to redistribution (e.g., supermajority requirements for tax increases). Thus, our redistribution dummy captures both liberal and conservative measures.[83] (In the results reported below, measures are coded as redistributive only if both authors agreed that the measure was redistributive.[84])

3. Results

As Table 2 and Figure 2 show, our results provide only partial support for our hypotheses. Figure 2 plots the readability and conservativeness scores of each ballot initiative. Higher readability scores mean the ballot label was more difficult to read. The curves show the relationship between conservativeness and readability for Democratic- and Republican-authored labels. Assuming that LRL populations tend to be liberal, we expected a U-shaped relationship between the

---

82.   James M. Synder, Jr., *Constituency Preferences: California Ballot Propositions, 1974–1990*, 21 LEG. STUD. Q. 463 (1996). Snyder shows that, during the period of his study (roughly half the period of ours), California voter preferences on ballot measures had three dimensions: a redistribution dimension, an environment/civil rights/public goods dimension, and a North-South dimension (mostly about water conflicts). Snyder finds that support for measures that load heavily onto the redistribution dimension is greatest in assembly districts with more Democratic *and fewer college educated* voters. *Id.* at 465, 471–75.

83.   We also included measures in the redistributive category if they provided equal benefits to everyone but would disproportionately benefit low-income voters. (Imagine a proposition that would establish free, state-provided health care and make it available to everyone.)

84.   Our individual assessments of redistributive measures were highly correlated (Krippendorff's alpha = 0.911). One author identified fifty-eight measures as redistributive and one identified fifty-five. Our sample is limited to the fifty-three measures where there was inter-coder agreement.

conservativeness of a ballot measure and the complexity of Republican-authored labels. If manipulating ballot labels is costly, AGs are strategic, and the payoff that AGs receive for the passage or defeat of a ballot measure increases with the measure's ideological divisiveness, we expected to obtain this result. Republican AGs will put in more effort to confuse LRL voters on very liberal and very conservative propositions. The U-shape should be most pronounced on competitive and redistributive propositions. Conversely, we expect to see an inverted-U relationship between ballot measures' conservativeness score and the readability of Democrat-authored labels, as Democratic AGs should be especially keen to avoid confusion among low-level readers on very liberal and very conservative measures.

The expected U-shaped and inverted-U-shaped relationships between AG party, ballot-measure conservativeness, and readability did not generally materialize. However, Republican AGs did write increasingly hard-to-read ballot labels on *competitive* measures that were more conservative than average. This effect is statistically significant. Graphically, the shaded gray region in the top-right panel of Figure 2 is a ninety-five percent confidence interval on readability; the upper bound of the interval when conservativeness equals zero (which describes a measure that is neither conservative nor liberal) is lower than the lower bound of the interval for the most conservative measures.

The top left panel in Figure 2 shows that there is no relationship between readability and conservativeness for Republican-authored labels on less competitive measures. The fact that Republican AGs have written hard-to-read labels only on competitive conservative measures tends to corroborate our strategic partisanship hypothesis. Note that Republican AGs did not write especially hard-to-read labels on redistributive measures as we hypothesized. However, this nonresult may be an artifact of an extremely small sample size. Our dataset contains only two conservative, competitive, redistributive measures whose labels were authored by a Republican AG.

Though Republican AGs' writing of increasingly hard to read labels on very conservative competitive measures is consistent with our hypotheses, the fact that Republican AGs did not also write more difficult labels for very liberal measures is puzzling. So too is the absence of the predicted relationship between readability and ballot-measure conservativeness for labels authored by Democrats. Indeed, as Table 2 shows, labels produced by Democratic AGs tend to be slightly harder to read than labels generated by Republican AGs. This is so whether one looks at the data in the aggregate or subsets it by competitiveness, divisiveness, or redistributiveness.

Table 2: Average Flesch-Kincaid Grade-Level Readability Scores for Ballot Labels Conditioned on the Competitiveness and Divisiveness of the Measure, and on the Party of the AG that Authored the Ballot Label

	50% Most Competitive			50% Least Competitive		
	All	Redistributive	Other	All	Redistributive	Other
Democratic AG	13.1	12.4	13.4	12.9	12.9	12.9
Republican AG	12.2	11.7	12.4	11.5	11.2	11.7

|	50% Most Divisive (by Party)			50% Least Divisive (by Party)		
	All	Redistributive	Other	All	Redistributive	Other
Democratic AG	13.1	12.4	13.5	12.9	13.1	12.9
Republican AG	12.3	10.7	13.4	11.4	12.6	11.1

What accounts for our limited results? It may be the case that the utility of passing conservative ballot measures is higher for Republican AGs than the utility of defeating liberal measures, although why this might be so is unclear. It may be the case that Democratic AGs have failed to see the potential payoff from manipulating language complexity. But this strikes us as unlikely; scholarly and elite attention to the readability of state-provided voter information materials is not new.[85] A third possibility is that the Flesch-Kinkaid algorithm does a poor job capturing the readability of very short texts (recall that the ballot labels are no longer than seventy-five words). But if measurement error was the story, one would not expect a strong correlation between readability and roll-off, as Reilly and Richey found, nor the relationship between conservativeness and readability for Republican-authored competitive measures that we found.

Perhaps norms and rules internal to the AG's office have professionalized the writing of ballot labels in ways that limit systematic manipulation of language complexity. Consistent with this hypothesis, Reilly and Richey report that the variance in the readability of California ballot labels is quite low compared to variance in the readability of ballot labels in most other states.[86] Our limited but suggestive results should certainly provide impetus for future investigations of strategic partisanship in the labeling of ballot measures in other states.

---

85. For an early and influential contribution, see DAVID B. MAGLEBY, DIRECT LEGISLATION: VOTING ON PROPOSITIONS IN THE UNITED STATES (1984). *See also* PHILIP L. DUBOIS & FLOYD F. FEENEY, IMPROVING THE CALIFORNIA INITIATIVE PROCESS: OPTIONS FOR CHANGE 138–40 (1992).

86. California has the fourth lowest variance in readability scores for labels among the forty-six states with statewide ballot measures between 1997 and 2007. Standard deviations ranged from 1.1 to 26.4 (in grade-level units). The standard deviation of California's readability scores was 1.8. *See* Reilly & Richey, *supra* note 62, at 63.

Figure 2: Flesch-Kincaid Grade-Level Readability Scores
for Every California Ballot Measure 1976–2010[87]
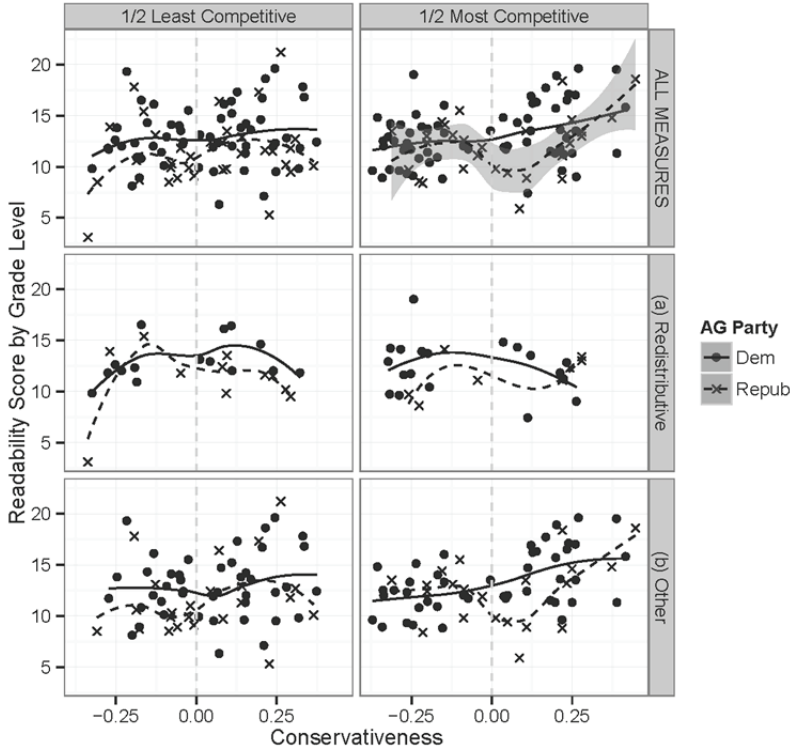
B. Bias as Ordinary Observers See It

1. *Theory and Hypotheses*

If the AG tries to write ballot labels that induce wavering or confused citizens to vote in accordance with the AG's preferences, these efforts should be detectable by dispassionate observers who closely study the text of the ballot measure and the AG's label. But observers may often err in evaluating bias. Some errors may be random noise. Others may be predictable, as each observer brings to the task of evaluating bias her own preconceptions and prejudices, as well as varying levels of care or capacity.

---

87. The top panels are all measures (*N*=186). The middle panels represent measures that the authors identified as redistributive (*n*=53). The bottom panels represent all nonredistributive measures (*n*=133). The *x*-axis is ballot-measure conservativeness as measured by the difference between presidential vote share in highly Republican assembly districts and highly Democratic assembly districts. Curves are locally weighted least squares (LOWESS) smoothers. The shaded region in the top right quadrant represents a ninety-five percent confidence interval.

If individual perceptions of bias contain an element of truth but also vary with the observer's capacities, prejudices, and preconceptions, then one can test the hypothesis that AGs behave strategically in writing ballot labels by modeling observers' perceptions of bias as a function of (1) a vector of individual-level covariates (traits likely to be correlated with perceived bias), and (2) a vector of political covariates (circumstances that strengthen or soften the AG's incentive to write biased labels). Of course, if perceptions of bias contain only a little information about actual bias, the researcher must obtain a large number of bias observations in order to detect the influence of AG partisanship.

We hypothesize that, at the individual level, bias evaluations will vary depending on whether the observer supports or opposes the measure, and whether the observer is a skilled reader. We also expect bias perceptions to depend on whether the observer thinks the ballot initiative process benefits the public generally or mainly benefits special interests, whether the observer is generally trustful of government and other people, and whether the observer identifies with the minority party in her state. As we explain below, we tried to minimize these influences by placing our observers behind a veil of ignorance regarding the author of the labels under study.[88]

Support/Opposition. Researchers have found that opinions about public policy color observers' perceptions of all sorts of matters touching on the policy question, ranging from the lessons of scientific studies to the existence of ambiguity in a statute.[89] It is also well established that losers in the political process—citizens who voted for the candidate or party that did not prevail—are much more likely than winners to believe the process to be unfair.[90] Though our subjects were not told whether the measure under evaluation passed or failed, we think subjects may rationalize the possibility of defeat, as it were, by imputing bias to the label ex ante. That is, voters who support a measure will be more likely to see bias in favor of a no vote, and voters who oppose the measure will be more likely to see bias in favor of a yes vote.

Regarding the *probability* of perceiving bias (as opposed to the *direction* of perceived bias), we draw two mutually inconsistent hypotheses from the literature.

---

88.   *See infra* Part II.B.2, specifically text accompanying notes 98–99.

89.   *See, e.g.*, Ward Farnsworth et al., *Ambiguity About Ambiguity: An Empirical Inquiry into Legal Interpretation*, 2 J. LEGAL ANALYSIS 1 (2010); Robert J. MacCoun & Susannah Paletz, *Citizens' Perceptions of Ideological Bias in Research on Public Policy Controversies*, 30 POL. PSYCHOL. 43 (2009).

90.   *See generally* CHRISTOPHER J. ANDERSON ET AL., LOSERS' CONSENT: ELECTIONS AND DEMOCRATIC LEGITIMACY (2005) (examining how election losers and their supporters respond to their loss and how institutions shape losing); Michael W. Sances & Charles Stewart III, *Partisanship and Voter Confidence, 2000–2010*, (MIT Political Science Dep't Research Paper No. 2012-12, 2012), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2035513 (exploring to what degree voter confidence in the fairness and trustworthiness of election procedures is driven by satisfaction in the outcome of an election).

The first hypothesis posits that supporters of a ballot measure will be less likely to perceive bias than opponents. This hypothesis is grounded in the large body of work showing that voters who supported the winning candidate in an election are more likely to perceive the electoral process as fair or legitimate than voters who supported the loser.[91] We did not tell subjects whether the ballot measure passed or failed, but the simple fact that a proposition has appeared on the ballot may count as a partial victory—a step toward victory—for supporters, and as a partial loss for opponents. If winning or losing at the agenda-setting stage has a similar (albeit weaker) effect on perceived fairness as winning or losing the election itself, then supporters of a ballot measure should be less likely to perceive bias than opponents.

The other hypothesis holds that the probability of perceiving bias will vary with the strength rather than the direction of support or opposition. On this view, people who care a lot about whether a measure passes or fails are more likely to perceive bias than people who only mildly support or oppose it. People whose support or opposition is strong will have firmer views about how the label should be written, and they will be alert to possible mischief by the "other side." It will also be harder for them to accept that a label that does not perfectly capture their understanding of the measure's merits or demerits may nonetheless be reasonable. The strength-of-opinion hypothesis finds some support in a recent study of perceptions of statutory ambiguity.[92] Law students who had firm, outcome-based preferences between two competing interpretations of a statute were much less likely to see the statute as ambiguous than students who had weak preferences.[93]

Reading Skills. We hypothesize that observers with stronger reading skills will be more likely to find that a label selectively includes or excludes information about the measure, in a manner calculated to induce a yes or no vote. We assume that better readers will learn more about the measure from the background materials we provide (more on this below), and that the more the reader knows about the measure, the more likely she is to find flaws in editorial decisions about what to include or exclude from the seventy-five-word label. We do not expect strong reading skills to increase the probability of perceiving other forms of bias.

Trust in Government, Society, and Direct Democracy; Minority-Party Identification. Outside the laboratory, ordinary citizens' perceptions of ballot-label bias probably have a lot to do with how the observer feels about government and society. Most obviously, observers who think the ballot-initiative process tends to benefit a few special interests rather than the public generally will be more likely to perceive bias than observers who think the ballot-initiative process serves the public good. Among citizens who know that a partisan, elected official authors ballot labels,

---

91.   *See supra* note 89.

92.   Farnsworth et al., *supra* note 89, at 257 (investigating statutory ambiguity in the law through a survey study administered to nearly 1000 law students).

93.   *Id.* at 271.

identification with the state's minority party probably increases the likelihood of perceiving bias. The likely effects of trust in government and social trust on perceived bias probably depend on what the observer knows or assumes about the identity of the ballot label's author. Observers who believe the label was written by a government official are probably more likely to perceive bias if they lack trust in government. Conversely, observers who think the label was written by the measure's proponents will probably perceive more bias if they lack social trust, that is, confidence in the fairness and integrity of their fellow citizens.

Because our primary goal in eliciting ordinary observers' bias perceptions is to obtain information about actual bias, we tried to minimize the influence of trust in government, minority-party identification, and the like by not revealing the label's author. Subjects were told that the label "may or may not" be the actual label that appeared on the ballot, and "may or may not have been prepared by a neutral and disinterested person."

AG Strategy and the Political Covariates of Bias. We hypothesize that labels for measures that the AG expects to be competitive are more likely to be seen as biased than labels for uncompetitive measures. As explained in Part II.A, strategic AGs have greater incentives to write biased labels if the election is expected to be close. We also predicted more perceived bias for measures that the AG or the AG's supporters strongly support or strongly oppose. A strategic AG will be more willing to bear the costs of writing a biased label insofar as either the AG or the AG's most important constituencies cares deeply about the outcome. We expected the direction of perceived bias to track the political incentives of the author of the label. That is, we expected to find perceived bias in favor of a yes vote on conservative measures labeled by a Republican AG and on liberal measures labeled by a Democrat, and bias in favor of a no vote on liberal measures labeled by a Republican AG and conservative measures labeled by a Democrat.

2. Operationalization

Implementing the ordinary observer strategy for estimating ballot-label bias presents a number of challenges. For starters, who should one use as "ordinary observers"? The ideal subjects for a study of California ballot propositions would be drawn at random from voting-age Californians—the people on whose behalf the attorney general acts when labeling ballot measures. Use of a true probability sample of voting-age Californians would also make it likely that measures of ballot-label bias incorporate much of the dispersed information about how voters are likely to see and understand the ballot labels under study.

We decided, however, to use convenience samples of students rather than a broader sample of Californians. We did this for reasons of cost and convenience, and because we wanted to ensure that our subjects were able to read and take notes on a hard copy of a long description of the ballot measure (more on this below). Had we tried to survey a more representative group of Californians, we

would have had to provide all materials to our subjects in electronic form, which would have made the materials harder to read and assimilate. We also would have had much less confidence that our subjects read the materials with any degree of care. Our subjects either completed the survey as a class assignment or took it under our supervision in an experimental lab.

We were concerned that a sample comprised of California students would contain too few conservatives for us to gauge whether agreement with conservative propositions or disagreement with liberal ones affects perceptions of ballot-label bias. We therefore recruited participation by students from Utah as well as California universities.[94]

The next implementation challenge was to provide students with a point of reference for evaluating bias. According to the courts, an unbiased label is one that reasonably informs voters about the gist of the ballot measure, without taking sides.[95] Evaluating bias therefore requires the observer to understand both how the ballot proposition would change state law and the likely effects of those changes. It would be a heroic undertaking for ordinary observers to wade through the text of a ballot measure, figure out exactly how it would change state law, gauge the likely effects of these changes, and finally judge whether the label fairly portrays the gist of the measure. California law recognizes as much, in that it directs the nonpartisan LAO to prepare a careful summary of each measure's principal provisions and likely effects for the ballot pamphlet.[96]

To simplify the observers' task, we asked them to evaluate bias in the ballot label *relative to the description of the measure in the ballot pamphlet*, rather than having them read and try to grasp the text of the measure itself. One might object to this strategy on the ground that the LAO's characterization of the proposition and its effects may itself be biased. If so, a ballot label that diverges from the legislative analyst's description of the measure could be *less biased* than a label that's faithful to the description.

We think our approach is nonetheless defensible because leading politicians and commentators who see the AG as biased favor delegating the authorship of ballot labels to the LAO,[97] which is widely regarded as impartial. If our results show that AGs are faithful to the LAO's characterization of a measure, then giving the task to the LAO is not likely to reduce bias. If, instead, our results show that AGs write biased measures (relative to the LAO's description) in

---

94.   We are extremely grateful to Quin Monson and Laura Marostica at Brigham Young University and to Sean Farhang at UC Berkeley who permitted us to administer our survey to their students. In addition, 242 students took the survey in UC Berkeley's Experimental Social Science Laboratory (XLab).

95.   *See supra* Part I.A.

96.   *See supra* Part I.A.

97.   *See supra* text accompanying notes 3–4.

circumstances where the AG has political incentives to do so, then our results would provide some support for the proposed reform.[98]

We did not tell our subjects that the long description for each measure had been prepared by the LAO, because we thought this might result in judgments of ballot-label bias becoming colored by opinions about the legislature or (possibly) the LAO. Instead, subjects were told that the long descriptions had been prepared by a neutral, disinterested expert.

Subjects received a packet containing the legislative analyst's descriptions of three ballot propositions, and were instructed as follows:

"Ballot initiatives" are citizen-proposed reforms that become law if approved by a majority of voters. They are described on the ballot with a title and a short written summary of their principal provisions and effects. In this study, we are collecting information about whether such titles and descriptions are perceived to be biased.

You will be asked to evaluate 3 ballot initiatives. For each ballot initiative, you will first carefully read a neutral, 1–2 page description of the measure and its likely effects, prepared by a disinterested expert (and included in this pamphlet). You will then read a short title and summary of the measure on the online survey. The short title and summary may or may not be the actual title and summary that appeared on the ballot, and it may or may not have been prepared by a neutral and disinterested person. We want to know whether you think the title/summary accurately and fairly describe the measure, or whether you think the title/summary is biased in some way.[99]

After reading through the rest of the instruction sheet, subjects logged onto a Qualtrics webpage, which prompted them to read carefully the first description

---

98. To be sure, it is formally possible that the legislative analyst could be biased in the opposite direction of the AG in precisely those circumstances where the AG has a political incentive to write biased labels, but that seems quite unlikely. It would mean that the nonpartisan and widely respected legislative analyst behaves like a Democrat when the AG is a Republican, and like a Republican when the Democrat is an AG. If nothing else, a finding (based on ordinary observer judgments of ballot-label bias relative to the legislative analyst's description) that ballot labels are biased in the manner expected of a strategic, partisan AG would serve up the question of whether it's the AG or the legislative analyst who is biased.

99. Christopher S. Elmendorf & Douglas M. Spencer, *Are Ballot Titles Biased? Partisanship and Ideology in California's Supervision of Direct Democracy*, 3 U.C. IRVINE L. REV. 511 app. B (2013), *available at* http://www.law.uci.edu/lawreview/vol3/no3/elmendorf_spencer_survey.pdf. Although California law uses the phrase "title and summary" to refer to the AG-written passage introducing a proposition in the ballot pamphlet, and the phrase "ballot label" to refer to the AG-written description on the ballot itself, CAL. ELEC. CODE § 9051 (2003), we described the label to our subjects as a "title and summary" for that seemed to us much more natural than the descriptor "ballot label."

in the pamphlet.[100] The next Qualtrics screen displayed the corresponding ballot label, followed by a series of questions about perceived bias.

Unlike people who have been asked to estimate the number of jellybeans in a jar, or the weight of a cow, the observers asked to gauge bias in a ballot label may have different ideas about what constitutes bias. We therefore asked our subjects not to assess bias generically, but rather to look for particular forms of bias as we defined them:

Prejudicial language. Does the title/summary use colorful or symbolic language, creating a risk that some voters who know little about the measure will end up voting for or against it because of their emotional response to the colorful or symbolic language?

Argumentative. Does the title/summary make an argument for or against the proposed measure?

Selective information. Does the title and summary selectively exclude or include information about the measure in a way that may cause some voters who know little about the measure to vote differently than they would have voted if they had carefully read the long description prepared by the neutral expert?

Other (please describe). Is the title/summary otherwise biased for or against adoption of the proposed measure?

These definitions of bias track those used by the courts when reviewing ballot labels, and ballot titles and summaries.[101]

Subjects who reported perceiving bias received a follow-up question about the direction of bias, with four answer options: (1) strongly biased toward no vote; (2) mildly biased toward no vote; (3) mildly biased toward yes vote; and (4) strongly biased toward yes vote.

After answering the suite of bias questions for each measure, subjects were asked, "Based on the long description of the measure you read in the pamphlet, how would you vote on this measure?" The corresponding answer options were: (1) definitely vote no; (2) probably vote no; (3) probably vote yes; (4) definitely vote yes.

The survey concluded with several questions designed to tap other individual-level attributes hypothesized to vary with perceived bias.[102] We measured trust in government, using the standard four-question battery developed by the American National Election Survey;[103] social trust, using both the

---

100. Qualtrics is a leading online survey platform that we used to create the survey. We customized the structure and design of the survey using Qualtrics' web-based software and then distributed a URL to all respondents that directed them to the site.

101. *See supra* Part I.A.

102. These are reported in full in Appendix B. Elmendorf & Spencer, *supra* note 99, at app. B.

103. *See Trust in Government Index 1958–2008*, AM. NAT'L ELECTION STUD. (2010), http://www.electionstudies.org/nesguide/toptable/tab5a_5.htm.

conventional General Social Survey (GSS) questions and alternatives developed by the economist Edward Glaeser;[104] and trust in the ballot-initiative process, for which we adapted one of the ANES trust-in-government questions.[105] To measure reading ability, we asked subjects about their score on the "Critical Reading" (Scholastic Aptitude Test, or SAT) or "Reading" (ACT) section of their college entrance exam.[106]

Because we expected observers' judgments of bias to be noisy and colored by individual-level traits, we figured we would need a substantial number of observations to detect strategic bias in ballot labels. Between 1974 and 2010 the California attorney general authored 185 ballot labels. We randomly selected ninety ballot labels for observation, stratified by attorney general: approximately sixteen labels per each of the six AGs during this time period.[107] We obtained a total of 996 ballot-label evaluations from 255 California and 77 Utah students,[108] roughly ten observations per measure.

All nondichotomous independent variables have been normalized to facilitate interpretation of the results.

3. Results

No doubt primed by our instructions, which cautioned that the labels "may or may not have been prepared by a neutral and disinterested person," the students found bias at strikingly high rates.[109] Seventy-four percent of the ballot-label observations include a finding of bias. By far the most common form of perceived bias is selective inclusion or exclusion of information. Students found the label to be faulty in this way in fifty-seven percent of the cases; by contrast, the prejudicial language, argumentative language, and "other" forms of bias were detected in thirty-five, thirty-three, and thirty percent of the cases, respectively.

Tables 3 and 4 report the results of a linear probability model (LPM) in

---

104.   Edward Glaeser et al., *Measuring Trust*, 115 Q. J. ECON. 811 (2000).

105.   Here the question we posed: "Would you say that laws enacted *through the ballot-initiative process* pretty much benefit a few big interests looking out for themselves or that they benefit all people?"

Because many of our observers are from Utah, a state that does not have the ballot initiative, we included the following definition as part of this question: "In states with the ballot initiative, measures proposed by individual citizens or groups are put on the ballot if the proponents gather signatures from a certain number of registered voters, and these measures become law if approved by a majority vote at the next election."

106.   *Cf.* Cheryl Boudreau, *Closing the Gap: When Do Cues Eliminate Differences Between Sophisticated and Unsophisticated Citizens?*, 71 J. POL. 964 (2009) (using SAT math scores as a measure of experimental subjects' ability).

107.   Our sample represents the following number and percentage of overall ballot labels authored by each AG: Younger (*n*=7, 100%); Deukmejian (*n*=13, 100%), Van De Kamp (*n*=16, 33%); Lungren (*n*=16, 38%); Lockyer (*n*=16, 34%); Brown (*n*=16, 62%).

108.   *See supra* note 83.

109.   Summary statistics are presented in Appendix A.

which the dependent variable is whether the observer perceived a particular form of bias and the independent variables include SAT score, trust in government, trust in direct democracy, social trust, minority-party identification, the measure's divisiveness (by party), and the competitiveness of the election. The divisiveness and competitiveness variables are the same ones we used in the readability study above. In Table 3, we include a dummy for whether the observer had "definite" vote intentions on the measure. In Table 4, we replace this variable with a dummy capturing whether the observer would definitely or probably vote yes.

The first column in these tables ("at least one type of bias") reports results from models that use an aggregated measure of perceived bias as the dependent variable. The dependent variable equals one if the observer perceived one or more types of bias in the label, and zero otherwise. We consider the results using this dependent variable to be most informative because, save for the predicted effect of reading ability, our hypotheses are not specific to particular types of perceived bias.

Our decision to utilize linear regression is predicated on Occam's razor. The mechanics of LPM are easy to understand and the results are simple to interpret: the coefficients represent the effect (shift in probability) on the dependent variable of a one unit change in the independent variable—for dichotomous variables a categorical change, and for normalized continuous variables a shift from the mean value to one standard deviation above the mean. Coefficients in traditional nonlinear models (e.g., logit and probit) are more difficult to interpret, especially coefficients on interaction terms.[110]

Looking at the results, we see no relationship between the strength of vote intention and the probability of perceiving bias. People who are really confident about how they would vote on a measure are no more likely to perceive bias than people who are wavering. This tends to undercut the hypothesis that strongly opinionated observers will see bias at a higher rate than observers who care less.[111]

---

110.   Chunrong Ai & Edward C. Norton, *Interaction Terms in Logit and Probit Models*, 80 ECON. LETTERS 123 (2003).

111.   This null result may reflect a problem in our survey instrument. We asked subjects whether they would "definitely" or "probably" vote yes (or no) on the measure as described in the pamphlet. The definiteness of vote intentions may reflect subjects' political knowledge as much as it does their level of care or concern about the issue. In later survey administrations, we asked a subset of respondents (n=46) how much they would care if each particular measure were passed—a lot, moderately, a little, or not at all—and note that this measure is uncorrelated to respondents' voting preferences (Pearson's r=0.06). However, this measure of concern about the issue was also not predictive of bias, nor did it alter any of the coefficients in Table 3 in models that included it (not presented).

Table 3: Linear Probability Model of Factors Affecting
Whether Bias Was Perceived (Not Direction of Bias)[112]

	At least one type of bias	Prejudice	Argument	Selective	Other
Probability of perceiving bias	0.739	0.353	0.326	0.576	0.291
Strong yes/no preference	0.019	0.018	0.027	−0.007	0.030
	(0.031)	(0.044)	(0.036)	(0.038)	(0.034)
SAT score	−0.006	0.014	0.053*	−0.041†	0.023
	(0.022)	(0.020)	(0.023)	(0.023)	(0.021)
Minority party ID	0.021	−0.023	0.030	−0.001	0.033
	(0.021)	(0.025)	(0.030)	(0.029)	(0.030)
ANES trust in gov't	0.001	−0.000	0.000	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Direct democracy trust	−0.022	−0.064†	0.042	−0.067	−0.001
	(0.036)	(0.038)	(0.043)	(0.042)	(0.040)
GSS social trust	−0.010	−0.012	−0.012	−0.010	−0.014
	(0.007)	(0.008)	(0.010)	(0.008)	(0.010)
Measure competitiveness	0.002	−0.026	0.012	−0.006	−0.018
	(0.018)	(0.021)	(0.017)	(0.022)	(0.018)
Measure divisiveness (by party)	0.008	0.020	−0.017	0.001	0.037
	(0.022)	(0.020)	(0.016)	(0.027)	(0.025)
N	771	771	771	771	771
R²	0.007	0.013	0.020	0.016	0.015

Note: Standard errors (in parentheses) are clustered on measure and respondent.
† Significant at $p < 0.10$; *$p < 0.05$

We do find support for the agenda-setters-as-partial-winners hypothesis. Observers who would vote yes are about seven percent less likely to perceive at least one type of bias than observers who would vote no (see Table 4). This effect (p=0.068) in the "at least one type of bias" model is driven almost entirely by perceptions of selective inclusion/exclusion and "other" bias. There is little relationship between whether one supports a measure based on the legislative analyst's description and whether one thinks the label uses prejudicial language or makes an argument for or against the measure.

---

112.   The outcome variable is a dummy set to one if respondents perceived bias. Models are run on the subset of respondents who reported a voting preference and their SAT score (219 ballot-label observations excluded).

Table 4: Linear Probability Model of Factors
Affecting General Perceptions of Bias[113]

	At least one type of bias	Prejudice	Argument	Selective	Other
Probability of perceiving bias	0.739	0.353	0.326	0.576	0.291
Vote intention (1=yes)	−0.065†	0.019	−0.036	−0.107*	−0.030**
	(0.035)	(0.036)	(0.033)	(0.042)	(0.039)
SAT score	−0.005	0.014	0.053*	−0.039†	0.025
	(0.022)	(0.020)	(0.024)	(0.023)	(0.021)
Minority party ID	0.023	−0.023	0.031	0.002	0.036
	(0.025)	(0.025)	(0.029)	(0.028)	(0.030)
ANES trust in gov't	0.001	−0.000	0.000	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Direct democracy trust	−0.015	−0.066†	0.045	−0.056	0.011
	(0.036)	(0.038)	(0.043)	(0.042)	(0.039)
GSS social trust	−0.011	−0.012	−0.013	−0.012	−0.016
	(0.007)	(0.008)	(0.011)	(0.008)	(0.010)
Measure competitiveness	0.004	−0.028	0.013	−0.001	−0.014
	(0.018)	(0.021)	(0.017)	(0.022)	(0.018)
Measure divisiveness (by party)	0.005	0.021	−0.018	−0.004	0.031
	(0.021)	(0.019)	(0.016)	(0.026)	(0.021)
N	771	771	771	771	771
R²	0.012	0.013	0.021	0.027	0.030

Note: Standard errors (in parentheses) are clustered on measure and respondent.
† Significant at $p < 0.10$; *$p < 0.05$

We find a significant effect of SAT reading comprehension scores on perceptions of bias—but not the effect we predicted. We thought good readers would see more selective inclusion or exclusion bias, for the simple reason that they would learn more than other readers about the measure from the legislative analyst's description. However, as Tables 3 and 4 show, strong readers are actually *less* likely than others to perceive this form of bias. They are, however, more likely than others to think the label makes an argument for or against the measure.

The rest of our individual-attribute results suggest that the veil of ignorance worked.[114] The various trust and minority-party variables have little if any effect

---

113. The outcome variable is a dummy set to one if respondents perceived bias. Models are run on the subset of respondents who reported a voting preference and their SAT score (219 ballot-label observations excluded).

114. We asked a small subset of our survey respondents (*n*=46) to identify whom they perceived most likely to have written the measures they evaluated. Just four of the respondents said

on perceived bias. Trust in direct democracy shows the expected negative relationship to perceptions of selective inclusion/exclusion bias, but not to other forms of perceived bias. The GSS indicator of social trust has a slight negative relationship with overall perceived bias, but the effect is trivial. Results using the Glaeser measure of social trust (not reported) are essentially the same.

What about the political covariates of perceived bias? We do not find much support for our strategic partisanship hypotheses. Ballot measure competitiveness is, on balance, positively related to the probability of perceiving bias, though the effect is not statistically significant. Partisan divisiveness has a slightly negative effect on the probability of perceiving some forms of bias, and a stronger, positive effect on the probability of perceiving other forms of bias. In the aggregated, "at least one type of bias" model—which in our view is the most relevant model—the sign of the coefficient on divisiveness (by party) is positive but the magnitude is very small. Because our measure of polarization is normalized (meaning that each value is recoded as its distance from the group average), the interpretation of the "divisiveness (by party)" coefficient in Tables 3 and 4 is the effect on perceived bias between measures with an average divisiveness score (when the variable equals zero) and measures with a divisiveness score that is one standard deviation above the average (when the variable equals one). In other words, the coefficient of 0.008 in Table 3 and 0.005 in Table 4 represents the increased probability (less than one percent) that a survey respondent perceived bias on measures that were one standard deviation above the mean, relative to the mean.

In a normal distribution, about sixty-eight percent of the data fall within one standard deviation of the mean. Thus, in a linear probability model with normalized independent variables, the effect of shifting an independent variable from one standard deviation below the mean from one standard deviation above—equivalent to a shift from the sixteenth percentile to the eighty-fourth percentile—is twice the coefficient. We therefore estimate that the effect of shifting partisan divisiveness from the sixteenth percentile to the eighty-fourth percentile on the probability of perceiving bias is about 1.0% (Table 4) or 1.6% (Table 3).[115] Taking account of the standard errors, we can with ninety-five percent confidence *rule out* an effect of this sixteenth-to-eighty-fourth-percentile shift larger than 5.8% (Table 3) or 5.2% (Table 4).[116] Given that our subjects perceived some form of bias in nearly seventy-four percent of the labels, we can be quite confident that a measure's divisiveness has very little substantive effect

---

they thought an elected partisan official wrote the ballot language, compared to seventeen who thought the language was written by the initiative's proponents, and to thirteen who thought the authors were trained or nonpartisan professionals.

115.   All effects described in percentage terms in the text are actually *percentage point* shifts. To illustrate, an effect of "five percent" is an increase in the probability of the dependent variable being equal to one of five percentage points.

116.   These effects are calculated as: 2 × (coefficient) + 2 × (standard error).

on perceived bias in the label. Bias is perceived in the vast majority of cases and at most is only a couple of percentage points more likely to be perceived with respect to measures that are much more divisive of partisans than average.

Even if observers actually perceive bias at meaningfully higher rates on highly divisive measures—which is unlikely, given our results—there are at least two mechanisms that could explain the effect. First, observers may be detecting actual bias resulting from AG strategy. Alternatively, observers may have a tendency to project bias when evaluating labels for ballot measures they consider controversial or highly partisan.

Table 5: Linear Probability Model of Factors Affecting
Perceptions of Bias Toward a Yes Vote[117]

	Aggregate	Prejudice	Argument	Selective	Other
Probability of bias toward "yes" vote	0.703	0.676	0.795	0.702	0.748
Vote intention (1=yes)	−0.044	0.047	−0.011	−0.102*	−0.032
	(0.049)	(0.058)	(0.057)	(0.043)	(0.071)
Measure conservativeness	−0.033	−0.064	−0.064	0.005	−0.044
	(0.029)	(0.044)	(0.040)	(0.026)	(0.047)
AG party	0.049	0.084	0.035	0.033	0.060
	(0.047)	(0.062)	(0.051)	(0.050)	(0.069)
AG party * measure conservativeness	0.029	0.082	0.045	0.013	0.057
	(0.044)	(0.060)	(0.048)	(0.048)	(0.075)
N	666	369	333	582	309
R^2	0.007	0.021	0.015	0.015	0.012

Note: Standard errors (in parentheses) are clustered on measure and respondent.
*$p < 0.05$

We could adjudicate between these hypotheses by looking at the *direction* of perceived bias. Though observers may have a tendency to project bias when they encounter controversial measures, they have no reason (controlling for their vote intention) to project bias *in favor of a yes vote* or *in favor of a no vote* unless they know the party or ideology of the label's author. Thus, if observers behind the veil discern a directional bias that accords with the political preferences of the AG who authored the label, this would be pretty good evidence of actual, strategic bias.

---

117. The outcome variable is a dummy where, conditional on perceiving a particular type of bias, respondents perceived bias in a yes direction. The first model ("aggregate") excludes eighty ballot-label observations where the respondent perceived bias in a yes direction for one type of bias (e.g., prejudice) and in a no direction for another type of bias (e.g., argumentative).

In the interest of completeness, we report directional bias findings in Tables 5 and 6. (We recognize that it would be quite odd to find directional bias consistent with our strategic partisanship hypothesis, given the lack of a meaningful positive correlation between measures' divisiveness by party and the probability of labels being perceived as biased.)

Table 6: Linear Probability Model of Factors Affecting
Perceptions of Bias Toward a Yes Vote[118]

	Aggregate	Prejudice	Argument	Selective	Other
Probability of bias toward "yes" vote	0.703	0.676	0.795	0.702	0.748
Vote intention (1=yes)	−0.034	0.049	−0.017	−0.089†	−0.020
	(0.052)	(0.063)	(0.056)	(0.046)	(0.073)
Utah Dem.	0.028	−0.169	−0.099	0.136	0.207
	(0.133)	(0.157)	(0.139)	(0.165)	(0.130)
CA Repub.	0.015	0.009	−0.013	0.075†	0.032
	(0.040)	(0.064)	(0.055)	(0.044)	(0.083)
Measure conservativeness	−0.104**	−0.074	−0.035	−0.078	−0.116
	(0.038)	(0.066)	(0.060)	(0.050)	(0.073)
Utah Dem. * measure conservativeness	0.169†	−0.073	0.114	0.115	0.190
	(0.980)	(0.168)	(0.135)	(0.138)	(0.118)
CA Repub. * measure conservativeness	0.103**	0.060	−0.018	0.105*	0.118†
	(0.039)	(0.070)	(0.063)	(0.050)	(0.070)
N	666	369	333	582	309
R^2	0.011	0.021	0.017	0.024	0.017

Note: Standard errors (in parentheses) are clustered on measure and respondent.
† Significant at $p < 0.10$; *$p < 0.05$; **$p < 0.01$

The directional models are limited to cases in which the respondent found one or more forms of bias in the label. The dependent variable is whether, in the respondent's judgment, the bias she discerned favors a yes vote. In a surprising number of cases (eight percent total), respondents found that a given label was biased toward a yes vote in some ways (e.g., making an argument) and toward a no vote in other ways (e.g., selectively including or excluding information). We exclude these observations from the model in which the dependent variable

118.   The outcome variable is a dummy where, conditional on perceiving a particular type of bias, respondents perceived bias in a Yes direction. The first model ("aggregate") excludes eighty ballot-label observations where the respondent perceived bias in a yes direction for one type of bias (e.g., prejudice) and in a no direction for another type of bias (e.g., argumentative).

aggregates all four types of bias into a single dummy,[119] but we include them otherwise.[120]

The independent variables in Table 5 include a vote-intention dummy, the measure's conservativeness, the attorney general's party, and an interaction of AG party and ballot-measure conservativeness. The AG variable is coded *–1* for Democrats and *+1* for Republicans. Thus, AG interacted with "measure conservativeness" will be positive when the AG's party supports the measure and negative when the AG's party opposes it.

As Table 5 shows, the coefficient on the interaction term is positively signed in all models—consistent with our strategic partisanship hypothesis—but small in magnitude and not even close to statistically significant. To get a sense of the substantive import of the point estimate, it is useful to compare the effect of shifting "measure conservativeness" from one standard deviation below the mean (a very liberal measure) to one standard deviation above (a very conservative measure) under a Democratic AG, with the effect of such a shift under a Republican AG.[121] The following discussion reports effects in the "aggregate" model, which we consider most informative.[122]

Measure conservativeness affects the probability of perceiving bias toward a yes vote both directly and through the interaction with AG party. The direct effect of increasing measure conservativeness on perceived bias toward a yes vote is negative and by construction the same regardless of AG party. Under a Republican AG, the direct and interaction effects basically cancel each other out; the net effect of shifting conservativeness from one standard deviation below the mean to one standard deviation above is a reduction in the probability of perceived bias toward a yes vote of about 0.7%.[123] Under a Democratic AG, the direct and interaction effects reinforce each other; the net effect is a reduction in perceived bias toward a yes vote of about 12.4%. The *difference* between the effect of the *−1* to *+1* standard deviation shift in measure conservativeness, under

---

119.   In the "aggregate" model, the dependent variable is equal to one if the observer classified the measure as biased in one or more category(ies) and if that bias was perceived to favor a yes vote.

120.   We do not have strong views on whether the "bias in both directions" cases should be included in the models in which the dependent variable is the direction of a particular kind of bias, rather than the direction of bias generally. We ran these models both ways (i.e., including or excluding the bias-in-both-directions cases) and the results are virtually identical.

121.   Because the independent variables have been normalized, this is roughly equivalent to shifting measure conservativeness from the sixteenth percentile to the eighty-fourth percentile. *See supra* text accompanying notes 105–06 for further explanation.

122.   We have no a priori reason to expect strategic AGs to use one form of bias rather than another. The aggregate model accounts for all forms of bias, and excludes arguably suspect observations where bias of one type was seen to favor a yes vote and bias of another type pointed toward a no vote.

123.   As explained earlier, the effect of shifting a normalized, nondichotomous independent variable in a linear probability model from its mean to one standard deviation above the mean is equal to the coefficient. The effect of shifting the variable from one standard deviation below to one standard deviation above is therefore 2 × (coefficient).

Republican and Democratic AGs, on the probability of perceiving bias toward a yes vote, is about 11.6%. However, this point estimate is not precise. We can only rule out, with ninety-five percent confidence, differences of more than forty-six percent.[124] An effect of forty-six percent would seem to us quite worrisome, especially given our initial expectation that individual observers' bias judgments would be quite noisy. Thus, while the fairly small and statistically insignificant coefficient on the interaction term between AG party and measure conservativeness means that we cannot confirm our strategic partisanship hypothesis,[125] neither can we rule it out.

---

124. A straightforward way to calculate the largest plausible *difference* (with ninety-five percent confidence) between a shift in ballot measure conservativeness from *−1* to *+1* (i.e., one standard deviation below the mean to one standard deviation above) under Republican and Democratic AGs on the probability of perceived bias toward a yes vote, is to set the coefficient on the interaction term to the upper bound of the ninety-five perecent confidence interval and then to multiply this "upper bound" estimate by four. This yields an estimate of 46.8%. (One must multiply by four rather than two, because the effect of increasing a measure's conservativeness changes signs between Democratic and Republican AGs, and we're interested in the *difference* in the direction of perceived bias on Democrat and Republican-authored labels. Note that the direct effect of conservativeness on the direction of perceived bias cancels out when one is looking at the *difference* between the effect of a shift in measure conservativeness under Democratic and Republican AGs, so the direct effect, and uncertainty in the estimate of the direct effect, can be ignored when one is asking about the differential effect of a shift in ballot measure conservativeness on the direction of perceived bias under Democratic and Republican AGs.)

Another way to calculate the confidence interval on the difference is by bootstrapping. We ran the "Aggregate" model in Table 5 on 10,000 bootstrap replications of the data, and calculated the 2.5 and 97.5 percentile values of the estimated difference between the effect of shifting ballot-measure conservativeness from *−1* to *+1* under Democratic and Republican AGs. This yielded an upper bound estimate 39.6%, a bit smaller than the difference calculated directly from the estimated coefficients. (The difference may be due to the model's clustering of standard errors.)

125. With a larger sample size we could measure the effect size with more precision, but even if our findings were statistically significant it is not clear that our point estimate of the interaction term would be substantively worrisome from a public policy perspective.

Our sample size prevents us from precisely measuring the effect sizes we observe. Using power analysis we can estimate the sample size necessary to either reject the null hypothesis that Attorneys General do not engage in strategic behavior, or confirm the null hypothesis, or both. Given a sample size of 666 (see Table 5), we should be able to estimate a statistically significant difference between Democratic and Republican AGs if respondents perceived bias at least eleven percent more often when reading labels written by one or the other. There are 358 observations of Democratic-authored ballot labels and 308 observations for Republican-authored ballot labels. The standard error of the difference is $b = $ sqrt$(p_1(1\text{-}p_1)/(0.54n) + p_2(1\text{-}p_2)/(0.46n))$ with an upper bound of $0.5 \times b$. *See* ANDREW GELMAN & JENNIFER HILL, DATA ANALYSIS USING REGRESSION AND MULTILEVEL/HIERARCHICAL MODELS 439–47 (2007). Using the conventional power level of 80% (meaning 80% of the 95% confidence intervals will not overlap 0.5), we can estimate the effect size x by plugging in 666 to n in the equation above, which simplifies to $666 = (2.8 / x) \times 2$ or $x = 0.11$.

The observed effect on the AG party dummy is just 4.5%. In order for us to statistically significantly estimate an effect of that size we would need a sample size of 3900. We are also interested in the interaction of AG and polarization. In order for the observed interaction effect size of 13.3% to be statistically significant, we would need a sample size of 5128. Because standard errors are proportional to $1/$sqrt(n) and for 80% power the estimated effect must be 2.8 standard deviations from 0, then the standard error must be at least $0.133 / 2.8 = 0.048$. Thus we would need a sample size $(0.133 / 0.048) \times 2$ times as large as the current sample size. $666 \times 7.7 = 5128$.

As for the individual-level covariates, we see some limited evidence for our hypothesis that subjects project bias against their personal position. The coefficients on the "vote yes" dummy are negative in nearly all of the models, but they reach statistical significance only when the dependent variable concerns selective inclusion or exclusion bias (the magnitude of the effect is about ten percentage points). This is further evidence that personal policy preferences particularly distort selective inclusion or exclusion bias.[126]

We find strangely mixed evidence for our minority-party identification hypothesis. Positive coefficients on the interaction term between Utah Democrat and conservativeness and negative coefficients on the interaction terms between California Republican and conservativeness would corroborate our hypothesis that minority-party identifiers see bias in the direction of the majority party's ideology. For Utah Democrats, the sign on the coefficient is positive in most models and becomes marginally significant in the aggregate model ($p = 0.09$). For California Republicans, however, the sign is also positive and highly significant ($p < 0.01$). The magnitude of this effect is captured by the sum of the interaction and the coefficient on "measure conservativeness." In the "Aggregate" model of Table 6, the effect for Utah Democrats is 6.5% with a 95% confidence interval of −14% to 29%. The effect for California Republicans is 0.2% with a 95% confidence interval of −4% to 4%. In other words, we observe some evidence that Utah Democrats were more likely to perceive bias in favor of a yes vote on conservative measures, but no evidence that California Republicans perceived a bias towards yes on liberal measures.

III. DISCUSSION AND CONCLUSION

As readers of the daily paper, we began this project with a dim view of California AGs' performance in writing ballot labels. If asked to vote on a proposition to transfer authority over ballot-measure communications from the AG to the LAO, we would have certainly voted yes.[127] But the results reported here give us pause.[128]

---

126. *See supra* Tables 3, 4.

127. In previous work, one of us has argued for new independent checks on the administration of elections. *See* Christopher S. Elmendorf, *Election Commission and Electoral Reform: An Overview*, 5 ELECTION L.J. 425 (2006); Christopher S. Elmendorf, *Representation Reinforcement Through Advisory Commissions: The Case of Election Law*, 80 N.Y.U. L. REV. 1366 (2005).

128. It is certainly possible that the nonpartisan legislative analyst would do a better, more impartial job of labeling ballot measures than the partisan attorney general. But this is far from obvious; our results provide little if any evidence that the AG has behaved as a strategic partisan in labeling ballot measures. It is also possible that transferring ballot-label responsibilities to the legislative analyst would make it harder for him or her to maintain the trust and confidence of legislative leaders on both sides of the aisle, given the litigation and casting of aspersions that seems almost inevitably to attend the labeling of ballot measures. An analysis of this trade-off is outside the scope of this Article.

On balance, we found little evidence to support our hypotheses about strategic manipulation of ballot-label complexity. To be sure, our results do suggest that Republican AGs may have tried to confuse or mislead LRL voters by writing difficult labels on conservative, competitive measures. But this result should be treated very cautiously, since the same pattern does not occur on very liberal measures, since there are only two Republican AGs in our sample, and since there is no evidence that Democratic AGs have tried to manipulate ballot-label complexity. Indeed, contrary to our expectations, Democratic AGs wrote harder-to-read labels than Republican AGs. On balance, we think the results of our ballot measure readability models are strong enough to motivate further investigation of readability manipulation in other states—particularly in states whose ballot labels exhibit lots of variation in readability—but not strong enough to impugn Republican AGs in California.

We found essentially no association between the competitiveness of a ballot proposition and veiled observers' perceptions of bias. If AGs were behaving as strategic actors, it is on competitive measures where one would expect to find the most bias.

Similarly, the effects of ballot-measure divisiveness (by party) on the probability of perceiving bias were small and statistically insignificant. We can be quite certain that the labels of very divisive measures (one standard deviation above the mean) are no more than 5.8% more likely to be perceived as biased than the labels of very nondivisive measures (one standard deviation below the mean). And even if the effect were this large—which is very unlikely—it would not necessarily imply that AGs have been writing biased labels on measures that divide the electorate on party lines. It is also possible that observers are predisposed to "see" (i.e., project) ballot-label bias whenever they recognize the subject of a measure as one over which Democrats and Republicans clash.

This projection effect should not, however, cause veiled observers to perceive a *directional* bias that favors the party of the AG. As such, the strongest indicator of strategic partisanship would be positive, statistically significant coefficients on the interaction terms between AG party and ballot-measure conservativeness in the directional models. We obtained small and statistically insignificant estimates for these coefficients. But because the standard errors are fairly large we cannot rule out a substantively important effect.

This illustrates a more general point: our Article does not prove that AGs are "not biased." Our measures of partisan divisiveness, competiveness, and redistributiveness are imperfect. Our observers may not have taken their task very seriously. Our sample size does not leave us with much statistical power to detect small effects. It may be that the AG is biased but only on a very small number of extremely controversial measures, which are not well captured by our divisiveness scores. Or it may be that the AG is biased on measures that would have big financial consequences for well-organized interests but do not excite and divide

partisans. Or it may be that for any given ballot measure there is a range of plausible labels that nonetheless have different consequences for how people will vote, and that AGs use their discretion to pick from these defensible labels the alternative that best serves their party's positions.

It is an open question whether the responses we elicited from our subjects are just noise or whether in the aggregate they contain meaningful information about which labels are more or less biased. The best evidence of meaningful information would be positive, statistically significant coefficients on the political variables that, we think, strengthen or weaken the incentive for the AG to write a biased label. But we've obtained no such result. In results not reported here, we confirmed that the distribution of bias scores across all measures in our study is extremely unlikely to have occurred by chance if all subjects had the same probability of perceiving bias and the probability of perceiving bias was unrelated to the measure or label at issue. But even if bias judgments are no better than the flip of a weighted coin, there is no reason to expect the weighting of the coin to be the same across observers.[129]

One final point is worth reiterating. We *do* find that individuals' support for or opposition to a measure based on the LAO's description affects their probability of perceiving bias. This effect is fairly large and statistically significant in the case of selective inclusion or exclusion bias. It is inconsequential in the case of argument and prejudicial-language bias. Interestingly, when California courts review ballot labels for bias, they generally have focused on whether the label makes an argument or uses prejudicial language rather than on whether the label is misleadingly selective.[130] The courts' focus, which seems to us fairly arbitrary, is perhaps explained by an intuition that judges' policy preferences are more likely to contaminate their assessments of the latter than the former.[131]

---

129. In future work, we may create a rank ordering of California's post-1974 ballot propositions by perceived bias. We would purge bias observations of the effects of individual-level covariates (vote intention, SAT score), and then aggregate the purged scores into a directional, measure-level estimate of bias. We could then randomly assign experimental subjects to read either the ballot label or the legislative analyst's long description of the measure, and afterwards ask subjects about their vote intention. If our aggregated measure of perceived bias is meaningful, respondents who are assigned to read biased labels should vote for or against the measure (depending on the direction of bias) at higher rates than respondents who read only the legislative analyst's description, whereas respondents who receive neutral labels should vote similarly to respondents in the legislative analyst group.

130. We base this claim on our reading of the small number of published opinions. *See, e.g.*, *supra* notes 17–20 and accompanying text.

131. That said, the courts in reviewing the AG's "title and summary" for the ballot pamphlet do evaluate whether the AG fairly conveyed the "chief purposes and points" of the ballot measure. *See* Yes on 25, Citizens for an On-Time Budget v. Superior Court, 189 Cal. App. 4th 1445, 1452 (2010); Lungren v. Superior Court, 48 Cal. App. 4th 435, 439–440 (1996). This is tantamount to asking whether the AG in writing the title and summary selectively included or excluded information in a manner calculated to induce a yes or no vote.

Appendix A:
Summary Statistics of Variables Used in LPMs

INDIVIDUAL-LEVEL COVARIATES

Variable	Values	*N*	%
University	UC Berkeley	255	77
	BYU	77	23
Self-reported party ID	Strong Dem.	57	17
	Dem.	56	17
	Leans Dem.	73	22
	Independent	57	17
	Leans Repub.	48	15
	Repub.	24	7
	Strong Repub.	16	5
SAT (Critical Reading) / ACT (Reading) score	700–800 / 31–36	148	45
	650–699 / 28–30	66	20
	600–649 / 25–27	36	11
	550–599 / 21–24	8	2
	Below 550 / Below 21	2	1
	Don't know	70	21

POLITICAL COVARIATES

Variable	Min	Max	Average (SD)
Attorney general	−1 (D)	1 (R)	−0.10 (1.00)
Trust in government (ANES)	11	83	42.48 (21.88)
Ballot trust	1	2	1.44 (0.50)
GSS social trust	−4.24	4.02	0.08 (2.20)
Glaeser social trust	−6.40	3.37	0.04 (2.06)
Competitiveness	−0.62	−0.02	−0.21 (0.13)
Measure divisiveness (by party)	0.01	0.39	0.20 (0.10)
Conservativeness	−0.31	0.39	0.03 (0.22)