# The Minimal Effect of Opinion Language on Review in a Judicial Hierarchy

Matthew P. Hitt* Daniel Lempert†

June 11, 2019

Abstract Word Count: 151
Manuscript Word Count: 3,996

*Assistant Professor of Political Science, Colorado State University. email: matthew.hitt@colostate.edu.
†Assistant Professor of Politics, SUNY Potsdam. email: lemperds@potsdam.edu.

**Abstract**

Does the Supreme Court seek out shoddily-written opinions for correction? Does it see well-written opinions as better vehicles? We construct an original dataset that contains information about textual features of a representative sample of published U.S. Courts of Appeals opinions from 1949-2002, as well as all published opinions for which cert was sought in OT 1982. We investigate whether textual properties are associated with discretionary review by the U.S. Supreme Court. We advance a multifaceted understanding of judicial writing quality, encompassing readability, cognitive complexity, affective language, and new to the judicial politics literature, informality. We find that all of these properties fail to exhibit a robust statistically significant association with review. These effects, furthermore, are uniformly negligible in magnitude. We conclude that, for the period we study, the Supreme Court does not use its discretionary docket to correct of poorly-written legal doctrine, nor does it seek better-written opinions as vehicles.

Keywords: Text Analysis; Certiorari; Circuit Courts of Appeals; Informality

# Introduction

We study the effect of judicial opinion quality, broadly construed, on the probability that a higher court elects to review a decision. The judicial politics literature offers little theory regarding the effect of textual properties on the U.S. Supreme Court's certiorari process. As a preliminary foray, we offer a simple resource and information based argument for why higher quality judicial opinions may be more likely to be chosen for review.

Along with the party briefs, the lower court opinion is the major source of information for the Supreme Court about the policy announced in a lower court's decision. Our review of the literature, as well as consultation with experts and a former Supreme Court clerk, did not yield a dispositive answer to the question of how often, or in what kind of cases, clerks (or justices) review lower court opinions prior to the cert vote. Though lower court opinions are not *always* consulted, they seem to have been regularly reviewed throughout the modern Court's history (Peppers 2006, 85, 94, 110, 121-122, 127, 128, 132, 138, 143, 153-154, 157-158; Perry 1991, 127, 287). Lower court judges also appear to *believe* that their writing matters for the probability of review. Perry (1991, 287) recounts that: "I was also told of attempts to 'certproof' a case. This was done by writing long, complicated opinions, resting the holding on various grounds. Such a case, of course, becomes a 'bad vehicle.' Justices want clean cases." Finally, we note that even if lower court opinions are not read directly by the Court, their writing quality could still influence the probability of review: for example, it is likely easier for a petition to poke holes in the arguments of a poorly-written opinion.

Ideological disagreements between the lower courts and the Supreme Court typically trigger the Supreme Court's monitoring role (Bueno De Mesquita and Stephenson 2002). Yet lower court opinions vary theoretically not only in their ideological location, but in their relative quality, broadly defined (Clark and Carrubba 2012). Perhaps the Supreme Court does not expend its valuable resources reviewing lower court opinions of low quality.[1] That is, when a lower court opinion is poorly-written, the clerks of the Supreme Court will likely view such an opinion as a bad vehicle for resolving important issues. The informational value of a lower court opinion to the Supreme Court may be influenced by its written quality; the clerks and justices may discount the doctrine promulgated in a low-quality opinion and therefore choose a different case to review. For these reasons, the Court may be less likely to review a decision as lower court opinion quality declines. We now turn to conceptualizing and measuring opinion quality.

## Properties of High-Quality Judicial Opinions

We draw on two literatures to operationalize judicial writing quality. First, we consider advice from academics and practitioners on what constitutes good legal writing. Second, we incorporate recent literature in writing style in judicial politics that speaks, sometimes indirectly, to elements of style that are associated with high quality writing.

Aldisert et al. (2009, 39) claim: "Good prose, in opinion writing as in any genre, must be clear." Lebovits, Curtin, and Solomon (2008, 7) echo that sentiment: "Judges must write precisely, simply, and concisely." Legal scholarship on writing style suggests opinions written in a clearer rhetorical style are of higher quality. For example, the Federal Judicial Center advises that "precision and clarity are the main concerns of good writing" (21) . In addition, practitioners are consistently instructed to use grammar, structure, and language consistent with more easily-readable text. Commentators urge writers to use the active voice, and avoid contractions, personal pronouns, intensifying adverbs and slang/idioms (Lebovits and Hildago 2009, 35). The FJC similarly suggests "[writers] should use active voice and...weed

[1]Some might argue instead that this line of reasoning is backwards. The justices claim that at least part of their work does revolve around resolving legal conflicts and legal errors (Baude 2015). If so, then poorly-written opinions are *more* likely to be reviewed.

out gratuitous adjectives and eliminate unnecessary adverbs" (23). Shapiro et al. (2013) also suggest writers be mindful of their structure: "Writers should strive to keep sentences short, to avoid excessive use of the passive voice, to use active verbs frequently, and to break up long paragraph" (733).

Legal academics and practitioners also advise writers to maintain a formal tone and to avoid emotional language. Scalia and Garner (2008, 118) reminds writers that "formality bespeaks dignity." Lebovits, Curtin and Solomon (2008, 22) cautions against affective language: "Although judges should write persuasively, they must avoid writing polemics or writing emotionally [... o]pinions are meant to be reasoned and solemn."

The literature in judicial politics has implemented empirical measures for each of these elements of style. Clarity has been conceptualized in two different ways: as the clarity of rhetorical language (which we refer to as "readability") and the clarity of the ideas discussed (which we refer to below as "cognitive complexity"; see Owens and Wedeking 2011, 1038 for discussion). The former has been operationalized as a function of word and sentence length, with longer words and sentences indicating less readable text (Black, Owens, Wedeking and Wohlfarth 2016). A prominent metric, which we use below, is the Flesch-Kincaid Grade Level (Kincaid, Fishburne, Rogers and Chissom 1975), calculated as $FKGL = 0.39(\frac{\text{Total Words}}{\text{Total Sentences}}) + 11.8(\frac{\text{Total Syllables}}{\text{Total Words}})$. This measure approximates the grade levels of education required to understand a text.

A second conceptualization of clarity is "cognitive complexity." Unlike readability, which captures the clarity of the language itself, cognitive complexity seeks to capture the clarity of ideas communicated by assessing the language choices employed. Scholars assessing cognitive complexity have primarily examined two elements: differentiation and integration. Differentiation taps whether individuals acknowledge multiple dimensions within an issue, while integration taps the connections an individual makes among these dimensions (Tetlock, Bernzweig and Gallant 1985). These elements collapse into a single dimension referred to as cognitive complexity. Texts that are more cognitively complex integrate concepts and ideas moreso than do simpler texts. This has led scholars to conclude that "as opinions become more cognitively complex, they become less clear (Owens and Wedeking 2011, 1038)."

A measure of informality presented in Budziak, Hitt, and Lempert (2019) combines several elements of informal language discussed above and in the computational linguistics literature: the presence of idiomatic language, phrasal verbs, personal pronouns, intensifying adverbs, contractions, polarized language, the absence of passive voice, and the extent to which the author selects the less formal of synonyms-pairs. A commonly used measure of emotional ("affective") language is based on the extent to which words indicating a high level of affect feature prominently in a text, (Black, Hall, Owens and Ringsmuth 2016).

Generally, we hypothesize that *as opinion quality increases, the probability that the Supreme Court grants certiorari increases.* We consider a number of plausible elements of writing quality in judicial opinions. Each of these elements is associated with a corollary hypotheses.

The hypothesis that the Court will seek to review better-written vehicles implies the following corollary hypotheses: (1) *as an opinion's informality increases, the probability that the Supreme Court grants certiorari decreases*; (2) *as an opinion's cognitive complexity increases, the probability that the Supreme Court grants certiorari decreases.*; (3) *as an opinion's use of affective language increases, the probability that the Supreme Court grants certiorari decreases*; (4) *as an opinion's readability decreases, the probability that the Supreme Court grants certiorari decreases.*

# Data

We seek to determine whether the quality of a judicial opinion, broadly defined, exhibits an association with its treatment by higher courts in a judicial hierarchy. Songer's Appeals Courts Database (`http://artsandsciences.sc.edu/poli/juri/appct.htm`) contains a stratified random sample of intermediate (Circuit Court) decisions 1925-2002. Inter alia, this database includes information indicating whether litigants at the Circuit level sought review by the Supreme Court and whether review was granted.[2]

Its benefits aside, the Songer database does not catalogue whether circuit court opinions result in actual ("square") conflict with opinions of another circuit. The presence of conflict massively increases the probability of a cert grant (Caldeira and Wright 1988), so controlling for actual conflict is crucial. New York University Law School undertook the measurement of actual conflict for all paid petitions filed during the OT 1982 term of the Court, an undertaking not since repeated (New York University Supreme Court Project 1984). We present a supplementary analysis of the published opinions for which cert was sought in OT 1982, controlling for actual conflict.

Readability is measured using the Flesch-Kincaid Grade Level score (Kincaid et al. 1975), as implemented in a custom Python script that uses tools from Bird, Loper and Klein (2009). Higher values indicate opinions that are less readable.

To measure cognitive complexity, we run the pre-processed opinions through LIWC 2015 version 1.1. Following (Owens and Wedeking 2011, 1040), we measure cognitive complexity as the factor score of 10 LIWC indicators: causation, insight, discrepancy, inhibition, tentativeness, certainty, inclusiveness, exclusiveness, negations, and percentage of words containing six or more letters. We utilize the 2007 LIWC dictionaries for direct comparability with prior published work (Owens and Wedeking 2011).

Judges and law clerks are encouraged to write in an emotionally neutral tone; as such, we are interested in affective language, whatever its polarity (positive or negative). Thus, we measure affective language via the "affect" category of LIWC (2007 dictionary), calculating the relative prevalence of some 900 words indicating high levels of affect in each opinion (Black, Hall, Owens and Ringsmuth 2016).

Our measure of informality is as described in Budziak, Hitt and Lempert (2019), which incorporates several components of informality: informal word choice, idiomatic language, phrasal verbs, personal pronouns, polarized language, contractions, intensifying adverbs, and

[2] We provide additional details about our data collection process in the Appendix.

(lack of) passive voice. First, to rank words by formality, we follow generally Brooke and Hirst (2014); we refer the reader to the Appendix and these articles for specifics and important details. The approach in Brooke and Hirst (2014) is to use the Ranking Support Vector Machine algorithm (SVM$^{rank}$) of Joachims (2002) to rank all words in a corpus according to formality, based only on a partial formality ranking of a subset of those words.

To measure idiomatic language, we searched for instances of idioms/slang phrases in opinions based on the idioms catalogued in Spears's (2008) *Essential American Idioms Dictionary.* To measure the use of phrasal verbs, we searched for instances of the 100 most common phrasal verbs (Gardner and Davies 2007, 352) in opinions. To measure usage of the passive voice, we searched for instances of a "to be" verb followed by a past-tense verb (e.g., "was attempted"). Instances of personal pronouns, contractions, and intensifying adverbs could all be counted in opinions using regular expressions in the `R` Programming Language (R Core Team 2014). For all of these frequency-based textual properties, we first scaled each count by the total number of words in the document. Linguistic polarization is measured as the absolute value of a document's polarity score. That is, the tone of a text from positive to negative can be calculated using the procedure described in Hu and Liu (2004); sentiment polarity is calculated in `R` using the `qdap` package (Rinker 2013). We measure linguistic polarization as the absolute value of an opinion's polarity score; higher values indicate more extreme sentiment. Finally, we standardized and aggregated these lower order properties of informal tone using factor analysis.

# Analysis

The various textual metrics are not highly correlated. Thus, we investigate whether any of the individual textual properties exhibit associations with the grant of review by the Supreme Court. The dependent variable in the analyses below is a dummy equal to 1 if the Supreme Court grants review. Of almost 15,000 circuit court decisions in our data, just over 3,000 of these decisions saw a certiorari petition filed. That is, for the majority of decisions in our data, the losing litigant did not seek review at the Supreme Court. Below, we estimate both a model that includes all decisions, and another that considers only decisions for which the losing litigant sought Supreme Court review.

We also control for a battery of case-level covariates suggested by prior literature on the certiorari process. These case-level covariates all derive from the Songer Database. First, we control for the presence of a brief amicus curiae from an outside interest at the lower court stage (Caldeira and Wright 1988).

Second, we also control for the presence of a dissenting opinion at the circuit court (Caldeira, Wright and Zorn 1999). We also include a dummy variable equal to one for cases for which the circuit court reverses a district court for similar reasons. Third, a long line of research in judicial politics indicates that some litigants (most notably the U.S. Government) fare better than others in the courts (e.g., Black and Boyd 2012). We include a dummy variable in the analysis below for the presence of the U.S. Government as a losing litigant, as the federal government may be particularly likely to successfully petition for cert.

Fourth, ideological considerations play a role in the votes of the justices over the granting of certiorari (Caldeira, Wright and Zorn 1999). As such, we control for the ideological compatibility the Supreme Court with the lower court decision by including a variable that is equal to the Martin-Quinn ideology score (Martin and Quinn 2002) of the Supreme Court's median justice if the lower court's decision is conservative, and $-1$ times the median justice's Martin-Quinn score if the lower court's decision is liberal (for an example of the same basic approach, see Bailey, Kamoie and Maltzman 2005, 78).

Fifth, we include fixed effects for the circuit of origin of each decision. We also include fixed effects for each decision's substantive issue area. Sixth, we control for opinion length by including the natural log of each opinion's word count in our models (as word counts are highly skewed in our data).

Seventh, as discussed above, we seek to control for the presence of actual and alleged legal conflict. In a separate model, we therefore analyze the population published opinions for which the losing litigants sought cert in the Supreme Court's 1982 term (Caldeira and Wright 1988). Thus, we can control for actual and alleged conflict in this set of cases.

Finally, as our data span multiple decades, we include a four-knot cubic spline in Models 1 and 2 to account for the age of decisions in our analyses below. This approach allows time to have a non-linear, non-monotonic effect on the dependent variable. We do not include this spline in Model 3, as the data in that analysis span only one term.

Table 1 presents the results of two logistic regression models, and one rare events logistic regression model (Model 3, due to the sparsity of review in a single term).[3] Model 3 is estimated via penalized maximum likelihood in order to reduce potential bias arising from rare event data (Coveney 2015). The dependent variable in each model is equal to 1 if the U.S. Supreme Court grants certiorari. Among the cases we analyze from our larger sample, 410 (about 11% of those that sought review) were granted cert. The models differ in that Model 1 considers all decisions in the data, while Model 2 considers only decisions for which the losing litigant petitioned for review. Model 3 considers only cases from OT 1982, including the actual and alleged conflict variables. These data are not a subset of the Songer data. Since these data differ slightly from our larger sample (some state supreme court decisions are included in the 1982 data, about 24% of the data), and due to the smaller N, we do not include circuit fixed effects in this specification.

Table 1 shows that of the four textual properties we analyze, none exhibit a statistically significant association with grants of certiorari across specifications. Informality is negatively and significantly associated with the probability of cert in models 1 and 2, but this association is no longer significant in model 3, when we include a control for actual circuit conflict. Similarly, increases in Flesch-Kincaid grade level (i.e., decreases in readability) are positively and significantly associated with the probability of cert in Model 1. However, once we analyze only those decisions for which a litigant sought review, the significance of this association vanishes, and this lack of significance persists in model 3. Across specifications, cognitive

[3] Results of Model 3 are robust to estimation by conventional logistic regression.

**Table 1.** Estimates of the Effect of Linguistic Characteristics on Cert Grant

| | (1) All Decisions $\beta$ (SE) | (2) Only Appellants $\beta$ (SE) | (3) OT82 $\beta$ (SE) |
|---|---|---|---|
| Textual Variables | | | |
| Informality | -0.52* | -0.41* | -0.30 |
| | (0.12) | (0.13) | (0.27) |
| Cognitive Complexity | 0.07 | 0.11 | 0.12 |
| | (0.07) | (0.07) | (0.17) |
| Reading Difficulty (Grade Level) | 0.02* | 0.01 | -0.16 |
| | (0.01) | (0.01) | (0.09) |
| Affective Language | 0.03 | 0.04 | -0.12 |
| | (0.06) | (0.07) | (0.15) |
| Words (Log) | 0.88* | 0.44* | 0.49* |
| | (0.10) | (0.09) | (0.19) |
| Control Variables | | | |
| Amicus | 1.12* | 0.91* | 1.88* |
| | (0.20) | (0.22) | (0.34) |
| Dissent | 0.69* | 0.42* | 0.75* |
| | (0.16) | (0.17) | (0.34) |
| Intermediate Reversal | 0.10 | 0.37 | 0.27 |
| | (0.15) | (0.16) | (0.30) |
| U.S. Loses | 0.42* | 1.07* | 3.31* |
| | (0.17) | (0.19) | (0.51) |
| Ideological Compatibility | -0.38* | -0.34* | -1.61* |
| | (0.14) | (0.15) | (0.38) |
| Actual Conflict | - | - | 3.94* |
| | | | (0.38) |
| Alleged Conflict | - | - | 0.36 |
| | | | (0.398) |
| Circuit FE | Yes | Yes | No |
| Issue FE | Yes | Yes | Yes |
| N | 14611 | 3693 | 924 |

$*p < 0.05$. Robust standard errors in parentheses. Dependent variable is equal to 1 if certiorari is granted, 0 otherwise. Models 1 and 2 include a four knot cubic spline to account for the age of decisions. Constant, circuit of origin, and legal issue area fixed effects not reported. Dependent variable is equal to one if the Supreme Court granted certiorari. Model 1 includes all decisions. Model 2 includes only decisions for which the losing party filed a cert petition (appealed the verdict). Model 3 includes all published decisions for which the losing party filed a cert petition in OT 1982, due to the availability of conflict data.

complexity and affective language do not exhibit statistically significant associations with the probability of cert.

The results generally show that the textual properties we consider do not exhibit consistent, statistically significant associations with grants of cert. Having not found statistically significant effects, can we *rule out* meaningful effects? In the Appendix, we follow the procedure described by Rainey (2014), and conclude that these metrics exert only negligible effects.

Briefly, we compare the marginal effect of each text-based metric to that of a control variable (the presence of an amicus brief) known to modestly impact the outcome. For no text metric does the 90% confidence interval of the marginal effect include this threshold.

In sum, there is no evidence that the Court seeks out high quality opinions, i.e., good vehicles, for review. Nor is there evidence that the Court seeks out low quality opinions for review and error correction. We fail to reject each of our null hypotheses.

Several explanations for these results seem plausible. First, perhaps our definition of opinion quality was too expansive. Alternatively, it may be that all the factors we consider do matter, but federal judges (and their clerks) write so uniformly well that it is difficult to detect the impact of style. Going forward, it is clear that further development of a theory of opinion writing by lower federal judgis needed. Both judges and their clerks write opinions, in service of various goals; when and why do these actors produce opinions of higher and lower quality, and to what end?

# Conclusion

We investigated the response of the U.S. Supreme Court to intermediate appellate opinions to determine whether any of a battery of textual properties are associated with grants of certiorari by the Court. To that end, we analyzed the textual properties of over 15,000 published opinions by intermediate federal appellate judges and state supreme court justices in the United States.

We find that several of these properties exhibit no statistically significant associations with grants of certiorari. Further, the substantive impact of all of the textual properties we analyzed is negligible. These results may suggest that textual properties make little difference with the justices and clerks of the Supreme Court. Alternatively, perhaps circuit judges uniformly write well enough so as not damage their interests. Distinguishing between these explanations demands a more fully developed theory of how judicial opinion quality should matter, and why some textual properties, but not others, should influence outcomes like cert grants.

# References


Aldisert, Ruggero J., Meehan Rasch and Matthew P. Bartlett. 1999. "Opinion Writing and Opinion Readers." *Cardozo Law Review* 31(1): 1–43.

Bailey, Michael A., Brian Kamoie and Forrest Maltzman. 2005. "Signals from the Tenth Justice: The Political Role of the Solicitor General in Supreme Court Decision Making." *American Journal of Poltical Science* 49(1): 72–85.

Baude, William. 2015. "Foreword: The Supreme Court's Shadow Docket." *New York University Journal of Law & Liberty* 9: 1–68.

Bird, Steven, Edward Loper and Ewan Klein. 2009. *Natural Language Processing with Python.* Sebastopol, CA: O'Reilly Media Inc.

Black, Ryan C. and Christina L. Boyd. 2012. "US Supreme Court agenda setting and the role of litigant status." *Journal of Law, Economics, and Organization* 28(2): 286–312.

Black, Ryan C., Matthew E.K. Hall, Ryan J. Owens and Eve Ringsmuth. 2016. "The Role of Emotional Language in Briefs Before the U.S. Supreme Court." *Journal of Law and Courts* 4(2): 377–407.

Black, Ryan C., Ryan J. Owens, Justin Wedeking and Patrick C. Wohlfarth. 2016. *US Supreme Court Opinions and their Audiences.* New York, NY: Cambridge University Press.

Brooke, Julian and Graeme Hirst. 2013. Hybrid Models for Lexical Acquisition of Correlated Styles. In *Proceedings of the 6th International Joint Conference on Natural Language Processing (IJCNLP '13).*

Brooke, Julian and Graeme Hirst. 2014. Supervised Ranking of Co-occurrence Profiles for Acquisition of Continuous Lexical Attributes. In *Proceedings of The 25th International Conference on Computational Linguistics (COLING 2014).*

Budziak, Jeffrey, Matthew P. Hitt and Daniel Lempert. 2019. “Determinants of Writing Style on the United States Circuit Courts of Appeal.” *Journal of Law and Courts* 7(1): 1–28.

Bueno De Mesquita, Ethan and Matthew Stephenson. 2002. “Informative Precedent and Intrajudicial Communication.” *American Political Science Review* 96(4): 755–766.

Caldeira, Gregory A. and John R. Wright. 1988. “Organized Interests and Agenda Setting in the U.S. Supreme Court.” *American Political Science Review* 82(4): 1109–1127.

Caldeira, Gregory, John Wright and Christopher Zorn. 1999. “Sophisticated Voting and Gatekeeping in the Supreme Court.” *Journal of Law, Economics, and Organization* 15(3): 549–572.

Clark, Tom S. and Clifford J. Carrubba. 2012. “A Theory of Opinion Writing in a Political Hierarchy.” *Journal of Politics* 74(2): 584–603.

Coveney, Joseph. 2015. “FIRTHLOGIT: Stata module to calculate bias reduction in logistic regression.”.

**URL:** *https://EconPapers.repec.org/RePEc:boc:bocode:s456948*

Gardner, Dee and Mark Davies. 2007. “Pointing Out Frequent Phrasal Verbs: A Corpus-Based Analysis.” *TESOL Quarterly* 41(2): 339–359.

Hayakawa, S.I. 1994. *Choose the Right Word*. 2nd ed. HarperCollins. Revised by Eugene Ehrlich.

Hu, Minqing and Bing Liu. 2004. Mining opinion features in customer reviews. In *American Association for Artificial Intelligence*. Vol. 4 pp. 755–760.

Joachims, Thorsten. 2002. Optimizing Search Engines using Clickthrough Data. In *9th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD02)*.

*Judicial Writing Manual: A Pocket Guide for Judges*. N.d.

Kincaid, J. Peter, Robert P. Fishburne, Richard L. Rogers and Brad S. Chissom. 1975. "Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel." Defense Technical Information Center Document Research Branch Report 8-75.

Lebovits, Gerald, Alifya V. Curtin and Lisa Solomon. 2008. "Ethical Judicial Opinion Writing." *Georgetown Journal of Legal Ethics* 21: 237–296.

Lebovits, Gerlad and Lucero Ramirez Hidalgo. 2009. "Advice to Law Clerks: How to Draft Your First Judicial Opinion." *Westchester Bar Journal* 36(1): 29–37.

Martin, Andrew D. and Kevin M. Quinn. 2002. "Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the US Supreme Court, 1953–1999." *Political Analysis* 10(2): 134–153.

New York University Supreme Court Project. 1984. "Summaries of cases granted certiorari during the 1982 term." *New York University Law Review* 59: 823–1003.

Owens, Ryan and Justin Wedeking. 2011. "Justices and Legal Clarity: Analyzing the Complexity of U.S. Supreme Court Opinions." *Law and Society Review* 45(4): 1027–1061.

Pedregosa, Fabian et al. 2011. "Scikit-learn: Machine Learning in Python." *Journal of Machine Learning Research* 12(1): 2825–2830.

Peppers, Todd. 2006. *Courtiers of the Marble Palace*. Standford, CA: Stanford Law and Politics.

Perry, H.W. 1991. *Deciding to Decide: Agenda Setting in the United States Supreme Court*. Harvard University Press.

R Core Team. 2014. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rainey, Carlisle. 2014. “Arguing for a Negligible Effect.” *American Journal of Political Science* 58(4): 1083–1091.

Rinker, Tyler W. 2013. *qdap: Quantitative Discourse Analysis Package*. Buffalo, New York: University at Buffalo/SUNY. Version 2.2.0.

Scalia, Antonin and Bryan A. Garner. 2008. *Making Your Case: The Art of Persuading Judges*. St. Paul, MN: Thomson West.

Shapiro, Stephen M., Kenneth S. Geller, Timothy S. Bishop, Edward A. Hartnett and Dan Himmelfarb. 2013. *Supreme Court Practice*. 10th ed. Arlington, VA: Bloomberg BNA.

Spears, Richard. 2008. *McGraw-Hill's Essential American Idioms*. McGraw Hill Professional.

Tetlock, Philip E., Jane Bernzweig and Jack L. Gallant. 1985. “Supreme Court Decision Making: Cognitive Style as a Predictor of Ideological Consistency of Voting.” *Journal of Personality and Social Psychology* 48(5): 1227–1239.

# Supplemental Appendix

In this Appendix we provide greater details about our data collection procedures, measures, and our analysis of negligible effects.

## Data Collection

In the paper, we analyze opinions from 1949-2002.We downloaded every circuit court opinion, 1949-2002, in HTML format that was available from `law.resource.org`. After converting the files to `.txt` format, we used a custom Python script to match the unanimous opinions to the citation information in the Songer dataset, and to trim extraneous material (i.e., text from the webpage that was not part of the actual opinion text). A research assistant manually downloaded nonunanimous opinions from Lexis; we then used a script to trim extraneous material, split each opinion in a case, and match each opinion to the citation in the Songer dataset.[4]

Two types of opinions are not included in the present analyses. First, we exclude all opinions smaller than 1.5 kB from the analysis. Opinions smaller than 1.5 kB are virtually all one sentence orders and as such are not particularly insightful when analyzed for textual properties. Second, due to occasional mistakes or irregularities in the file names at `law.resource.org` and in the opinion citations in the Songer dataset, for about three percent of the citations in the Songer dataset, our script did not find a match in the set of opinions we downloaded from `law.resource.org`. Some preliminary investigation reveals that this is due to occasional mistakes or irregularities in the file names from `law.resource.org`. When possible, we manually augmented our dataset with these "unmatched" opinions. After listwise deletion of observations with missing covariates, we are thus left with 14,961 majority opinions.

[4]We did not use `law.resource.org` for the nonunanimous opinions, since the format of the opinions from that site did not lend itself to reliable splitting at the opinion level (i.e., within a case).

## Measuring Informal Word Usage

The approach we adopt for measuring informality, as originally described by Brooke and Hirst (2014), is to use the Ranking Support Vector Machine algorithm (SVM$^{rank}$) of Joachims (2002) to rank all words in a corpus according to formality, based only on a partial formality ranking of a subset of those words. Using scikit-learn's (Pedregosa et al. 2011) CountVectorizer, we create a binary term-document matrix based on all opinions over 1.5 kb in size. We exclude a set of stop words, and words that appear in more than 99 percent or fewer than 1 percent of all opinions; we also use NLTK's (Bird, Loper and Klein 2009) WordNetLemmatizer to stem the words. Our set of profile words are words that appear in at least 10 percent, but no more than 50 percent of documents. Unlike Brooke and Hirst (2014), in constructing the co-occurrence profiles, we do not sample—that is, we take *every* document in which a word to be scored appears, and then calculate the proportion of those documents that each profile word appears in.[5] To train SVM$^{rank}$, we construct pair-(and triple-)wise rankings of formality, relying on Hayakawa (1994) and Brooke and Hirst (2013). Finally, we use the weights SVM$^{rank}$ outputs to rank each word; each document's informality score is the mean of the ranks of the scored words in the document—the higher this score, the more informal the document.

## Demonstrating the Negligible Effect of Textual Features on Grants of Certiorari

Here, we test the proposition that our linguistic features exert no meaningful influence on the cert process, by following the procedure described in Rainey (2014).

To understand the utility of this procedure, it is important to remember that statistical insignificance alone cannot demonstrate that some effect is substantively negligible. Rather, a negligible effect is one that is estimated to be near zero with precision (i.e., with small standard errors). Thus, following Rainey (2014), we set up the following test. First, we

[5]This is feasible because of our smaller corpus.

specify a threshold $m$ that represents a substantively meaningful effect. Then, we estimate an effect size $\Delta_t$ and an associated 90% confidence interval, for each textual feature. If $\Delta_t$ has a 90% confidence interval that includes only values smaller in magnitude than $m$, then we can conclude, with at least 95% confidence, that the effect of that property is substantively negligible (Rainey 2014).

The choice of $m$ is subjective, and open to debate in any given application (Rainey 2014, 1085). In our case, we set $m$ to be equivalent to the presence of one amicus brief in favor of certiorari, a factor known to be meaningfully associated with, but not one of the larger influences on, cert (Caldeira and Wright 1988). Thus, if a textual property has less of an influence on the cert decision than a single amicus brief (a change in predicted probability of about 0.1), it has little practical effect on the process.

Table A1 shows the 90% confidence intervals for $\Delta_t$ for all of the textual properties we analyze, across all three models. Since our textual properties are continuous, the effects shown in Table A1 represent a change from the 25th to 75th percentile of that variable.[6] Table A1 shows that in every specification, for every textual property, these properties exert at best a negligible effect on the probability of a cert grant. Indeed, no property, in any specification, exhibits even half of the effect size we stated a priori was necessary for a conclusion of substantive significance. While not a focus of this analysis, we also found that opinion length exerts only a negligible effect on cert across specifications as well.

[6] Predicted probabilities estimated via the margins command in Stata 13.1. Confidence intervals estimated via the delta method.

**Table A1.** Effect estimate for a change in probability of cert, by textual property

| Specification | Property | Effect Size ($\Delta_t$) | 90% CI | Conclusion |
|---|---|---|---|---|
| Baseline | Reading Difficulty | 0.001 | [0.0004, 0.002] | Negligible Effect |
| Appellants Only | Reading Difficulty | 0.002 | [-0.002, 0.006] | Negligible Effect |
| OT82 Only | Reading Difficulty | $-0.021$ | [-0.041, -0.0002] | Negligible Effect |
| Baseline | Informality | $-0.010$ | [-0.013, -0.006] | Negligible Effect |
| Appellants Only | Informality | $-0.027$ | [-0.040, -0.013] | Negligible Effect |
| OT82 Only | Informality | $-0.014$ | [-0.034, 0.007] | Negligible Effect |
| Baseline | Affect | 0.001 | [-0.002, 0.005] | Negligible Effect |
| Appellants Only | Affect | 0.005 | [-0.007, 0.017] | Negligible Effect |
| OT82 Only | Affect | $-0.009$ | [-0.029, 0.011] | Negligible Effect |
| Baseline | Complexity | 0.002 | [-0.001, 0.005] | Negligible Effect |
| Appellants Only | Complexity | 0.010 | [0.001, 0.019] | Negligible Effect |
| OT82 Only | Complexity | 0.008 | [-0.010, 0.026] | Negligible Effect |

Effect estimates represent a change from the 25th to 75th percentile for a given textual property.