On Separation of Powers and Obfuscation in U.S. Supreme Court Opinions*

Daniel Lempert†

September 10, 2021

Abstract

A longstanding debate in American judicial politics concerns whether the U.S. Supreme Court anticipates or responds to the possibility that Congress will override its decisions. A recent theory proposes that opinions that are relatively hard to read are more costly for Congress to review—and that as a result, the Court can decrease the likelihood of override from a hostile Congress by *obfuscating* its opinions: writing opinions that are less readable when congressional review is a threat. I derive a straightforward but novel empirical implication of this theory; I then show that the implication does not in fact hold. This casts serious doubt on the claim that justices strategically obfuscate opinion language to avoid congressional override. I also discuss sentence tokenization as a source of measurement error in readability statistics for judicial opinions.

*I thank Jeff Budziak for helpful discussion and William Minozzi for suggesting a data source. I appreciate the particularly constructive comments from the Editor and three reviewers.

†Associate Professor of Politics, SUNY Potsdam. email: dalempert@gmail.com.

A longstanding debate in American judicial politics concerns whether the U.S. Supreme Court anticipates or responds to the possibility that Congress will override its decisions. A widely-cited paper, Owens, Wedeking and Wohlfarth (2013, 38–40) theorizes that opinions that are relatively hard to read are more costly for Congress to review.[1] Thus, the paper argues, the Court can decrease the likelihood of override from a hostile Congress by *obfuscating* its opinions: writing opinions that are less readable when congressional review is a threat. Owens, Wedeking and Wohlfarth (2013, 48) provides evidence that the Court's majority opinions are relatively less readable when the Court is constrained by Congress (in a sense to be made precise below).

On the one hand, as Owens, Wedeking and Wohlfarth (2013, 52) acknowledges, this result is, broadly speaking, inconsistent with a fair bit of empirical research about congressional influence on the Court (see e.g., Owens 2010; Owens 2011; Sala and Spriggs 2004; Segal 1997; Segal, Westerland and Lindquist 2011).[2] On the other hand, the theory in Owens, Wedeking and Wohlfarth (2013)—which I will refer to as *strategic obfuscation theory*—has an important strength: it is *elaborate*, in the sense used by, for example, Rosenbaum (2010, Ch. 19). This is to say, in short, that strategic obfuscation theory's proposed causal mechanism has several different testable implications.[3] In this research note, I propose and test one straightforward implication of the theory's causal mechanism.

Theory

Strategic obfuscation theory draws loosely on literature formally modeling how courts or other agencies can raise the costs of review for supervisory bodies (e.g., Staton and Vanberg 2008). The initial premise of strategic obfuscation theory is intuitive and well-supported: Congress has limited resources and time (e.g., Cox and McCubbins 2005; Lee 2010). As such, the costs of taking a given action are always relevant for Congress. Owens, Wedeking and Wohlfarth

[1] As of 7/5/21, the paper has been cited 71 times according to Google Scholar, including by at least three textbooks.

[2] But see King (2007) and, somewhat more generally, Clark (2009).

[3] The terminology dates back to R.A. Fisher and William G. Cochran (Rosenbaum 2015). The perspective is of course consistent with mainstream philosophy of science, which prefers theories that make relatively more falsifiable predictions; see Rosenbaum (2017, Ch. 7) discussing, among others, Popper (2002) [1959].

(2013, 39) cites, for example, collective action problems, the need to regularly credit-claim and produce benefits for constituents, and the shrinking size of staffs as factors limiting congressional capacity to act on issues, particularly those that are complex.

The novel proposal in Owens, Wedeking and Wohlfarth (2013) is that the Court can raise the costs for Congress to review and potentially override its opinions by obfuscating the language therein. The argument is as follows:

> Obfuscated Court opinions can generate heightened review costs and thereby deter congressional responses. To understand complex and obscure Court decisions, Congress must expend additional—and scarce—resources. A member who wishes to alter the Court's policies or otherwise punish the Court must examine the central logic and tenets of the Court's opinions and may even need to examine how the opinion compares to others written in the past by the Court. In some cases, the Court's opinion may be clear. In those cases, members may easily internalize the degree to which they favor the political content of the majority opinion. Yet the Court also has the ability to obfuscate opinions by making them less readable. In those instances, the heightened legislative costs required to address the opinion may increase. By writing a less readable opinion, justices might craft a desired judicial policy while simultaneously deterring a legislative response by making it more difficult for Congress to address it (Owens, Wedeking and Wohlfarth 2013, 39).

Owens, Wedeking and Wohlfarth (2013, 39–40) recognizes that making opinions less readable will not absolutely bar review, and that obfuscation has costs; for example, it may cause lower courts or relevant agencies to implement Court policies inaccurately (see also Black, Owens, Wedeking and Wohlfarth 2016). Nonetheless, strategic obfuscation theory proposes that when the threat of congressional override is great—in particular, when the Court is constrained—the Court can reduce the chances of review by obfuscating the language in the majority opinion. To this end, under the proposed causal mechanism, the majority opinion author intentionally obfuscates when facing a hostile Congress. Owens, Wedeking and Wohlfarth (2013) presents results indicating that majority opinions are written less readably when the Court is constrained; the magnitude of the effect is as much as one full grade level.[4]

Owens, Wedeking and Wohlfarth (2013) does not discuss dissenting opinions or the theoretically-predicted behavior of dissenters. Nonetheless, as I argue here, strategic obfuscation theory has clear implications for how dissenting opinions should be written. Consider the incentives of the dissenters. When should dissenters seek to increase the probability of review? Assuming policy-motivated actors (as does strategic obfuscation theory), they should do so when they prefer the policy that would result from an override to the policy announced in the majority opinion. Let C be the policy announced in the Court's majority opinion, D the dissenters' most preferred policy, and R the policy that results from congressional review and override. Given a very basic spatial model, the dissenters prefer an override whenever $|D - R| < |D - C|$—whenever the dissenters are closer to Congress than to the Court majority.[5] Since, under strategic obfuscation theory, the probability of review increases as obfuscation decreases, dissenters should write particularly *readably* when the majority is constrained but the dissenters prefer an override on policy grounds. The benefit to dissenters from writing more clearly when the Court is constrained is in making an override—and thus, a policy outcome they prefer to that resulting from the majority opinion—more likely.[6]

To summarize, under strategic obfuscation theory, while the Court majority has incentive to obfuscate when the Court is constrained, the dissent does not. Rather, the dissenters have the opposite incentive insofar as they prefer the policy that would result from an override: they should write particularly readably.[7] By making it easier for Congress to understand what the

[4]This is as measured by the Coleman-Liau Index of readability. I describe the measure and detail other empirical specifics below.

[5]I discuss operationalization of these policy locations below.

[6]As stated just above, this is conditional on $|D - R| < |D - C|$; my empirical tests below account for this caveat.

[7]One might object that dissenters may wish to avoid overrides from Congress even if they favor Congress' policy preferences over those of the majority, perhaps for institutional reasons. There are two answers to this point. First, we know that, on a regular basis, justices do explicitly invite overrides from Congress (e.g., Hausegger and Baum 1999; Rice 2019); thus, if explicit requests are normative, surely so are implicit actions that increase the chances of review. Second, even granting for the moment that dissenting justices do not intentionally seek to increase the readability of their opinions when Congress is hostile to the majority, they—at the very least—have no reason to affirmatively obfuscate under strategic obfuscation theory. This is a slightly different, in a sense weaker, implication; but given the results I present below, the distinction turns out to be

majority opinion implies, and where the majority opinion has erred, the dissenters can reduce the costs of review for Congress, and make it more likely that an opinion they disagree with is overridden. I test this implication of strategic obfuscation theory below.

Measurement, Sample, Hypothesis

Measurement of Variables

Constructing a dependent variable requires a measure of obfuscation, that is, (lack of) readability.[8] Following Owens, Wedeking and Wohlfarth (2013, 42–44), I use the Coleman-Liau Readability Index (CLI). This index is a function of word and sentence length. A key advantage of the measure is ease of interpretation, since it is scaled to approximate the (U.S.) grade levels of education needed to understand a text. CLI is defined as:

$$\text{CLI} \equiv 5.88 \left(\frac{\text{Number of Letters}}{\text{Number of Words}} \right) - 29.6 \left(\frac{\text{Number of Sentences}}{\text{Number of Words}} \right) - 15.8. \tag{1}$$

Thus, as intuitive, a text becomes more readable (or less obfuscated) as average word length and average sentence length decrease. Later I discuss specifics regarding implementation, and discuss of the non-trivial challenges associated with measuring CLI's constituent terms.

The most straightforward way to construct an appropriate dependent variable is to use the average CLI for dissents in a given case; I call this variable *Dissent CLI*.[9] Under strategic obfuscation theory, the expectation is that *Dissent CLI* decreases when the Court (i.e., the majority) becomes constrained, since the dissenters then have incentive to make the opinions more readable.

The key independent variables are measures of congressional constraint. In the results I present in the main text, I follow Owens, Wedeking and Wohlfarth's (2013) operationalization

irrelevant.

[8]In the relevant literature, readability is also referred to as (rhetorical) clarity.

[9]In practice, this involves combining all dissents for a given case into a single text file, and calculating a CLI for the combined text. Thus, this is a weighted average of all dissents in a case, where the weights are a function of individual opinion length.

in all respects.[10] Conceptually, the measures are set to 0 when the Court is unconstrained—that is, when it is located ideologically between the most extreme congressional pivots (very generally, see Krehbiel 1998). When the Court is constrained—that is, when it is to the right of the rightmost pivot or left of the leftmost pivot—it takes on the value of the ideological distance between the Court and the pivot closest to it. Figure 1 illustrates. If the scenario shown on the top axis obtains, the measure of constraint equals 0. If the scenario on the bottom holds, the measure equals the euclidean distance between C and P_L.

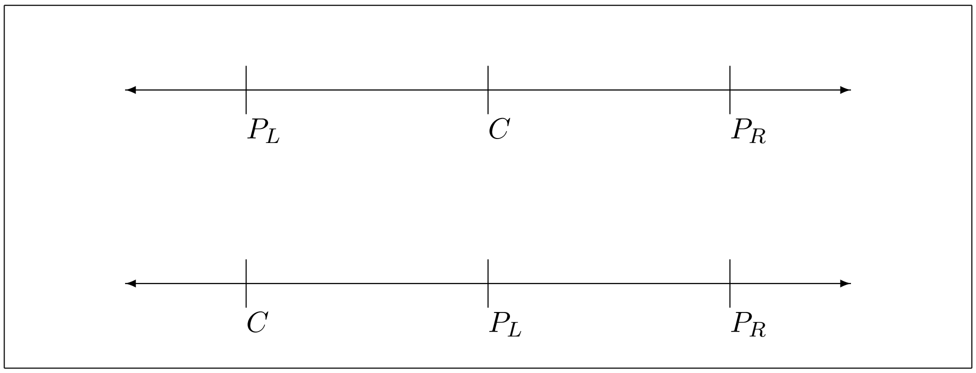

Figure 1. The Court (C) is unconstrained on the top axis, since it is between the leftmost congressional pivot (P_L) and the rightmost pivot (P_R). It is constrained on the bottom axis, because it is to the left of the leftmost pivot. Were the Court to the right of the rightmost pivot, it would also be constrained.

It remains to locate the Court and relevant pivots in ideological space. My approach is exactly that of Owens, Wedeking and Wohlfarth (2013). I locate actors using Judicial Common Space (JCS) scores (Epstein, Segal, Spaeth and Walker 2007); that is, 1st Dimension DW-NOMINATE scores for legislators (Lewis, Poole, Rosenthal, Boche, Rudkin and Sonnet 2021) and Martin-Quinn (2002) scores transformed into DW-NOMINATE space for justices. JCS scores thus range from -1 (most liberal) to 1 (most conservative). The Court's ideal point in a given case is identified as that of the median in the majority coalition. There are four different ways of locating relevant pivots, each motivated by a different model of congressional policymaking (Owens, Wedeking and Wohlfarth 2013, 45). For each of the four models, the relevant pivots are the leftmost and the rightmost of the following actors:

[10] I do this to limit researcher degrees of freedom (Simmons, Nelson and Simonsohn 2011). But there are any number of reasonable alternative variable specifications; I discuss some of these in the Appendix.

1. *Filibuster Pivot Model.* For and after the 94th Congress: the House median, the 40th most conservative Senator (i.e., the Senator with the 40th greatest JCS score), and the 60th most conservative Senator. Before the 94th Congress: the House median, the Senator at the 33rd percentile of conservatism, and the Senator at the 67th percentile of conservatism.[11]

2. *Chamber Median Model.* The median member of the Senate and the median member of the House.

3. *Committee Median Model.* The median member of the Senate Judiciary Committee and the median member of the House Judiciary Committee.[12]

4. *Majority Party Median Model.* The median member of the majority party in the Senate, and the median member of the majority party in the House.

There are thus four variants of the key independent variable measuring the degree to which the Court majority is constrained: *Distance To Filibuster Pivot*, *Distance To Chamber Median*, *Distance To Committee Median*, and *Distance To Majority Party Median.* Each is defined as the absolute difference between the ideal point of the Court (i.e., the Judicial Common Space score for the median of the majority coalition in a case) and the ideal point of the closest pivot, as defined just above, if the Court is constrained, and 0 otherwise. Thus, as the Court becomes "more constrained," these variables increase. Below, I refer to these variables collectively as the *constraint variables.*

I include the same control variables as Owens, Wedeking and Wohlfarth (2013): *Lower Court Conflict*, *Case Complexity*, *Precedent Alteration*, *Judicial Review*, and *Coalition Heterogeneity.* The definitions follow those given in Owens, Wedeking and Wohlfarth (2013, 46–47). Unless noted, the variables are based on information in Spaeth, Epstein, Segal, Ruger, Martin and Benesh (2017).

[11] This is of course because before the 94th Congress, two thirds of Senators were required to vote for cloture to end a filibuster, while starting with the 94th Congress, only 60 Senators were required. I refer to percentiles since in the earliest years of my sample there were only 96 senators.

[12] I identified committee members using two datasets: Swift, Brookshire, Canon, Fink, Hibbing, Humes, Malbin and Martis (2009) and Stewart III and Woon (2017).

Lower Court Conflict equals 1 if the Court notes that the sole reason it granted a case is to resolve a conflict in the lower federal or state courts, and 0 otherwise. Case Complexity is the number of amicus briefs in a case. These data are from Collins (2008) and Box-Steffensmeier and Christenson (2012). Precedent Alteration equals 1 if the majority opinion alters existing Court precedent, and 0 otherwise. Judicial Review equals 1 if the majority struck down a federal law, and 0 otherwise. Coalition Heterogeneity is the standard deviation of Martin-Quinn (2002) scores for justices voting in a given majority or minority coalition.[13]

Sample

Building on the Supreme Court Database (Spaeth et al. 2017), I construct a dataset of all signed Supreme Court majority opinions from October Terms 1947-2012. This is altogether 6699 majority opinions, 6690 once observations with data missing on covariates are dropped. I am chiefly concerned with those cases where at least one dissent was authored.[14] There are 3374 such cases, 3372 once observations with data missing on covariates are dropped.

I exclude one (small) subset of cases for reasons discussed above: those where the majority is constrained, but the dissenters prefer the policy in the majority opinion to that which would result from a congressional override. (I.e., where $|D - R| > |D - C|$.) To locate the policy from an override (R), I take the midpoint between the House and Senate chamber medians for the Filibuster Pivot, Chamber Median, and Committee Median models, and the midpoint between the House and Senate majority party medians for the Majority Party Median Model (for detailed discussion, see e.g. Harvey and Friedman 2006, 540–542). I locate the dissent (D) at the median of dissenting justices' ideal points. (And as stated above, I follow Owens, Wedeking and Wohlfarth (2013) in locating the Court majority opinion (C) at the median of the majority coalition.) This leaves between 2755 and 2964 cases in the estimation sample, depending on the theoretical model used to locate congressional pivots. Notably, this sample is significantly larger

[13] In other words, if a given observation is a majority opinion, Coalition Heterogeneity is the standard deviation of the Martin-Quinn scores for the majority justices, and if the observation is a dissenting opinion (or opinions), Coalition Heterogeneity is the standard deviation of the Martin-Quinn scores for the dissenting justices). I define dissenting opinion Coalition Heterogeneity for a single dissenting justice as 0.

[14] I exclude from the definition of *dissent* those dissents in part where the "dissenters" were coded as agreeing with the majority disposition by Spaeth et al. (2017).

(and slightly broader in temporal scope) than the sample in Owens, Wedeking and Wohlfarth (2013), which is a random sample of 529 majority opinions 1953–2008; thus, a lack relative lack of power is not a concern.

Hypothesis

As discussed above, strategic obfuscation theory implies that dissent authors will write more readably when the Court majority is constrained. Thus, the hypothesis is that if the Court majority's distance to a relevant pivot increases, readability of dissenting opinions increases. Precisely, as Distance to Filibuster Pivot, Distance to Chamber Median, Distance to Committee Median, and Distance to Majority Party Median, respectively, increase, Dissent CLI decreases.

Analysis

For each of the four constraint variables (Distance To Filibuster Pivot, Distance To Chamber Median, Distance To Committee Median, and Distance To Majority Party Median), I estimate three models predicting a dissenting opinion's CLI. The Baseline model includes no controls, the Add Controls model includes the controls mentioned above (Lower Court Conflict, Case Complexity, Precedent Alteration, Judicial Review, Coalition Heterogeneity), while the Add FEs model includes these controls and also fixed effects for dissenting opinion author and primary issue area. The models are OLS regressions, with standard errors clustered by term.[15]

Table 1 presents regression coefficients and standard errors for the twelve models. For none of the specifications is the coefficient on the constraint variable statistically significant and negative; this is clearly contrary to theoretical predictions. In fact, all of the coefficients are positive, and statistically significant in seven specifications, including all three specifications where Distance to Filibuster Pivot is the variable measuring constraint. Thus, there is no

[15]Again, this follows the procedure in Owens, Wedeking and Wohlfarth (2013), though there, of course, the fixed effects are for the majority opinion author, since the outcome is majority opinion CLI. I combine all justices writing fewer than 30 opinions into a single "Other Justice" category, to allow for valid estimation of the clustered standard errors.

indication that dissenters write more readably when the majority is constrained, contrary to the expectation derived from strategic obfuscation theory.

Key IV	Baseline	Add Controls	Add FEs
Distance to Filibuster Pivot	1.230*	1.221*	0.514*
	(.20)	(.18)	(.16)
Distance to Chamber Median	0.760*	0.779*	0.215
	(.18)	(.15)	(.12)
Distance to Commitee Median	0.465*	0.510*	0.119
	(.18)	(.15)	(.11)
Distance to Majority Party Median	0.208	0.082	0.069
	(.17)	(.16)	(.09)

Table 1. Dissenting opinion readability as a function of Court majority constraint. DV: Dissenting opinion CLI. OLS coefficients and standard errors (clustered by term), for twelve models: four variants of a distance to relevant pivot, and three model specifications. See text for details. (*: $p < 0.05$.)

Discussion

Can these results be made consistent with strategic obfuscation theory? One argument, based on a certain form of unobserved confounding, is as follows. Suppose that there is a case-level confounder that happens to be positively associated with an opinion's CLI, and also (by unfortunate chance) with the constraint variables. On the face of it, at least the first association is not unreasonable: the fixed effects for primary issue are relatively crude, consisting of 13 issue categories (Spaeth et al. 2017). If such a confounder exists, the results in Table 1 could still hold if this confounding overwhelms dissenters' efforts to write readably; in other words, one might propose that the coefficients in Table 1 would be *even greater* (due to the confounding) if dissenters did not make a particular effort to write readably when strategically warranted.[16]

One way to rule out this out is to examine how the putative effects of the constraint variables vary between majority opinions and dissents. Even under the proposed confounding, the positive association between the constraint variables and opinion CLI should be greater for

[16]But note that such confounding would imply that the results in Owens, Wedeking and Wohlfarth (2013) are also overestimates.

majority opinions, whose authors are trying to obfuscate, than for dissenting opinions, whose authors are trying to write readably.

I test for this possibility by analyzing both majority opinions and dissents in a single model; specifically, I add the majority opinion associated with each case whose dissent(s) were analyzed in Table 1. I estimate the same models shown in Table 1, except I include a binary variable indicating whether a given opinion is (=1) or is not (=0) a majority opinion, which I interact with the constraint variable in the model. If there is a stronger positive association between a constraint variable and CLI for majority opinions, the coefficient on the interaction term should be positive.

Key Interaction	Baseline	Add Controls	Add FEs
Distance to Filibuster Pivot	-0.829^*	-0.833^*	-0.526^*
× Majority Opinion	(.16)	(.17)	(.19)
Distance to Chamber Median	-0.525^*	-0.560^*	-0.325^*
× Majority Opinion	(.13)	(.14)	(.14)
Distance to Commitee Median	-0.487^*	-0.487^*	-0.266^*
× Majority Opinion	(.12)	(.12)	(.11)
Distance to Majority Party Median	-0.093	-0.103	0.023
× Majority Opinion	(.12)	(.12)	(.09)

Table 2. Opinion readability—conditional on opinion status (majority vs. dissent)—as a function of Court majority constraint. DV: Opinion CLI. OLS coefficients and standard errors (clustered by term), for twelve models: four variants of a distance to relevant pivot, and three model specifications. The table presents the coefficients and standard errors for the key interaction, which should be positive if majority opinions, more so than dissenting opinions, are written as majority opinions are predicted to be written by strategic obfuscation theory. See text for details. (*: $p < 0.05$.)

The results, given in Table 2, are contrary to the prediction derived from strategic obfuscation theory. For none of the 12 specifications is the coefficient attending the interaction term positive and statistically significant. In 11 of 12 specifications, the coefficient is negative; in nine of those 11—in all but those where Distance to Majority Party Median is the measure of constraint—the coefficient is statistically significant. Thus, the weight of the evidence indicates that the effect of majority constraint is greater for dissents than for majority opinions. That is, dissenters apparently increase their level of obfuscation *more* than the majority, as the threat of congressional override increases.[17]

These results are robust. A similar pattern of results obtains if I modify the sample used in Table 2 by adding unanimous majority opinions. The same is true for a slightly different analytical approach: setting the case as the unit of analysis, and the difference between the majority and dissent CLIs as the dependent variable. These results are in the Appendix, along with other specifications involving an alternative measure of ideology, definition of constraint, and location of Court majority opinion. In each of those specifications, as in all but one specification above, the dissenters' response to constraint is closer to the response theoretically predicted for the majority, *than the response of the majority itself.*

In short, not only is there no evidence that dissenters write more readably when the majority is constrained, but there is not even evidence that majority opinion authors obfuscate as a function of Court constraint to a greater degree than dissenters. This is incompatible with strategic obfuscation theory, since majority obfuscation is strategically warranted when the Court is constrained, and dissenting opinion obfuscation is strategically unwarranted.

In sum, I have shown that a straightforward empirical implication derived from the theory of strategic obfuscation receives no support. This casts serious doubt on the theory—specifically, its proposed causal mechanism: that justices strategically manipulate writing style to avoid review from Congress. To be explicit, strategic obfuscation theory has no straightforward theoretical explanation for why justices would seek to affect the probability of override when in the majority, but *those same justices* would not seek to do so when in dissent. (Even worse for the theory, justices in dissent tend to write in ways that are not just nonstrategic, but strategically counterproductive.) Turning now from the theoretical to the practical, I consider the potential role of measurement error in the original result supporting strategic obfuscation theory.

[17] The alert reader may note that the coefficients from the analysis in Owens, Wedeking and Wohlfarth (2013) are generally greater than those in Table 1 above. Below, I explain why, despite this apparent discrepancy, the results in Table 2 and the associated discussion should be credited.

On Calculating Accurate Sentence Counts

In this section, I sketch my approach to calculating the number of sentences in a legal text and show that it is a major improvement on the approach used by Owens, Wedeking and Wohlfarth (2013). Then I replicate the original analysis in Owens, Wedeking and Wohlfarth (2013) on my larger sample with an improved sentence counter, showing that the coefficients on the constraint variables are several times smaller than those originally reported; the discrepancy is likely due in large part to measurement error in the original study.

Recall the definition of that the definition of CLI (given in Eq. 1) requires counting the number of letters, words, and sentences in a text. Counting the number of letters and words in a legal opinion is relatively straightforward. The challenge is accurately counting the number of sentences in a text.

A naive approach is to count as a sentence any segment of text that ends in an end-of-sentence punctuation mark like `.`, `?`, or `!`. This is not satisfactory, however, since abbreviations within sentences can also contain periods. (See also initials and ellipses.) And legal opinions are full of abbreviations—most notably, but not only, as part of citations: `U.S.`, `L.Ed`, `F.2d`, and so on.

I implement my approach in the Python programming language, relying on tools in Bird, Loper and Klein (2009). The first step is to use the unsupervised sentence boundary detection method (or sentence *tokenizer*) in Kiss and Strunk (2006). Specifically, I train the tokenizer on a corpus of appellate opinions. Essentially, the tokenizer looks for collocations: pairs (more precisely, n-tuples) of text strings with a period between them that are likely to be abbreviations. I also add a set of abbreviations from the pre-trained English language sentence tokenizer from Kiss and Strunk (2006). In my application, this only gives a slight improvement over the naive approach.

Thus, I manually augment the abbreviations and collocations detected by the Kiss and Strunk (2006) tokenizer with various "legal" abbreviations and collocations (e.g., `civ. p.`, `id. at`, `u.s.c.`, and many more). This step gives the greatest improvement over the naive method. I then take sentences tokenized by the augmented tokenizer and disallow certain "sentences" that are unlikely to actually be sentences: for example, those that start with a lower-case letter, those that contain fewer than 3 words, and those that end with certain abbreviations. This gives a further slight improvement.

Owens, Wedeking and Wohlfarth (2013, 43) gives a sentence count for one of the opinions in its sample: *Washington* v. *Recueno* (548 U.S. 212), which is classified as having 400 sentences. (Indeed, the opinion includes 400 periods.) My method counts 99 sentences. The opinion is reproduced in the Appendix. An exact manual count of sentences is not entirely straightforward because readers might have the occasional disagreement on what constitutes a sentence, but in any case, it is clear that 99 is at worst a slight overestimate, and 400 overstates the number of sentences by at least a factor of 4.[18] This overestimate of course distorts the CLI: while Owens, Wedeking and Wohlfarth (2013) gives a CLI score of 6.1 (implying that the opinion is comprehensible to a sixth-grader), my estimate of the CLI is 13.3.

The face validity of some other scores cited in Owens, Wedeking and Wohlfarth (2013, 43, fn. 8) is also open to doubt; for example, *Crane* v. *Cedar Rapids and Iowa City Railway Co.* (395 U.S. 164) starts with these two (not atypical) sentences:

> The Federal Safety Appliance Act of 1833 requires interstate railroads to equip freight cars "with couplers coupling automatically by impact," but does not create a federal cause of action for employees or nonemployees seeking damages for injuries resulting from a railroad's violation of the Act. The Federal Employers' Liability Act of 1908 provides a cause of action for a railroad employee based on a violation of the Safety Appliance Act, in which he is required to prove only the statutory violation and the carrier is deprived of the defenses of contributory negligence and assumption of risk.

Owens, Wedeking and Wohlfarth (2013, 43) score this opinion's CLI as a 4.3, while my estimate

[18] Other off-the-shelf methods are only somewhat better. The `quanteda` package (Benoit, Watanabe, Wang, Nulty, Obeng, Muller, Matsuo and Lowe 2021) in `R` counts 268 sentences. NLTK's (Bird, Loper and Klein 2009) `sent_tokenize`—which is in fact an unmodified version of `punkt` trained on English language texts—gives 160 sentences; better than `quanteda` but still a significant overestimate.

is 12.6. In short, it is very likely that measurement error due to a naive sentence count affects the dependent variable in the original analysis.

Ex ante, it is not obvious how much this biases the central results in Owens, Wedeking and Wohlfarth (2013): the rank ordering of opinions' CLI could be more-or-less preserved if there is a global overestimation of the number of sentences. I thus replicate the main analyses, using the larger sample I describe above (and the improved sentence counter/CLI score). Table 3, analogous to Table 1 in Owens, Wedeking and Wohlfarth (2013), presents these results.

Key IV	(1)	(2)	(3)	(4)
Distance to Filibuster Pivot	0.216			
	(0.11)			
Distance to Chamber Median		0.114		
		(0.08)		
Distance to Committee Median			0.072	
			(0.08)	
Distance to Majority Median				0.010
				(0.05)
Lower Court Conflict	0.029	0.029	0.029	0.029
	(0.03)	(0.03)	(0.03)	(0.03)
Case Complexity	0.024*	0.023*	0.023*	0.023*
	(0.00)	(0.00)	(0.00)	(0.00)
Precedent Alteration	0.021	0.020	0.019	0.021
	(0.06)	(0.06)	(0.06)	(0.06)
Judicial Review	0.260*	0.261*	0.261*	0.258*
	(0.08)	(0.08)	(0.08)	(0.08)
Coalition Heterogeneity	0.005	−0.002	−0.006	−0.012
	(0.02)	(0.02)	(0.02)	(0.02)

Table 3. Majority opinion readability as a function of Court majority constraint. $N = 6690$. DV: Majority opinion CLI. OLS coefficients; standard errors clustered by term. Constant and fixed effects for primary issue area and opinion author not shown. See text for details. (*: $p < 0.05$.)

The coefficient on each of the key independent variables (Distance To Filibuster Pivot, Distance To Chamber Median, Distance To Committee Median, and Distance To Majority Party Median) remain positive, as strategic obfuscation theory predicts. But, whereas in the original analysis, all except one of the key independent variables (Distance to Majority Party Median) had a statistically significant attending coefficient, now none of them remain significant. Even more to the point, the coefficients are much smaller than those reported in the original analysis. A one-unit increase in Distance to Filibuster Pivot increases CLI by less than a quarter grade level; in fact, Distance to Filibuster Pivot ranges only from 0 to about .55, so the maximum in-sample effect is approximately an eighth of a grade level.[19]

This, then, explains why Table 1 and Table 2 are correct despite the results in Owens, Wedeking and Wohlfarth (2013). The results in the original article vastly overestimate, in all likelihood, the coefficients on the key independent variables. Once those results are corrected, by using a more accurate sentence tokenizer to count sentences, it is clear that the coefficients in Table 1 are larger than those in Table 3, which effectively implies the results in Table 2.[20]

Conclusion

This research note has presented two central results. First, I have shown that a straightforward empirical implication derived from the theory of strategic obfuscation receives no support. Specifically, dissenting justices who prefer a congressional override on policy grounds do not write more readably to an increase the probability of an override when majority is constrained (and thus subject to potential override). This casts serious doubt on the theory—specifically, its proposed causal mechanism: that justices strategically manipulate writing style to affect the probability of review from Congress.

Still, there are positive lessons to be learned from this result. Because strategic obfuscation theory was elaborate (in the sense I discussed above), it allowed for testing of multiple implications. This is must be acknowledged as a strength of the theory, even though it did not ultimately find empirical support. As Rosenbaum (2015) eloquently points out, scholars should

[19] This difference from the original result cannot be attributed to the different samples. I do not have access to the particular sample of 529 cases used in Owens, Wedeking and Wohlfarth (2013), but I investigate the role of sampling by re-running the analysis in Column 1 of Table 3 on 500 random samples of 529 cases. The median coefficient attending Distance to Filibuster Pivot across the 500 sample draws is .19 (standard deviation: .32); the largest across 500 draws is 1.20—less than the 1.81 in Owens, Wedeking and Wohlfarth (2013, 48). It is possible of course that *some* of the difference in our results is due to an (un)lucky sample draw in Owens, Wedeking and Wohlfarth (2013); it is very unlikely that this accounts for any meaningful part of the difference. In any case, since my sample—which is effectively the population of relevant cases from 1947–2012—subsumes the sample in Owens, Wedeking and Wohlfarth (2013), it is the results here that should be credited, at least on sampling grounds.

[20] I use the qualifier "effectively" because the results in Table 2 exclude unanimous opinions and (as discussed) a small set of non-unanimous opinions, whereas the results in Table 3 above include both unanimous and non-unanimous majority opinions (as do the results in Owens, Wedeking and Wohlfarth (2013)).

hesitate to accept theories that make only a single prediction or a few related predictions; an elaborate theory, which makes several independent predictions, is to be preferred.[21] True, the evidence from testing several implications of an elaborate theory may be ambiguous, or even disappointing. But "inconsistency and uncertainty [are] necessary stepping stones on a path to greater consistency and greater certainty (Rosenbaum 2015, 209)."

The second central result in the note is that the initial analysis supporting strategic obfuscation theory, in Owens, Wedeking and Wohlfarth (2013), was likely affected by measurement error in the outcome variable, CLI. Specifically, overestimation of the number of sentences in the opinions led to underestimation of CLI scores, which in turn appears to have inflated the effect estimates by a factor of at least four.

This has implications not just for strategic obfuscation theory, but more generally for researchers who seek to calculate readability metrics for legal texts. In particular, researchers should ensure that the sentence tokenizer (segmenter) used to calculate the number of sentences is adapted to the peculiarities of judicial opinions. Most importantly, the tokenizer should account for the abbreviations that are common in legal texts but not other English-language texts. The customized tokenizer I use here, which outperforms several off-the-shelf solutions, is included with the replication materials.[22]

[21]There is a statistical literature that makes these points more precise. For citations, see Rosenbaum (2015, 209) and Cook (2015, 145).

[22]The tokenizer is most appropriate for federal appellate opinions (Courts of Appeals and Supreme Court). Some caution should be used if applied to legal texts other than opinions, if applied to state court or federal trial court opinions, or opinions distant from the time period considered here. It should not be used for opinions from non-U.S. courts.

References

Benoit, Kenneth, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Muller, Akitaka Matsuo and William Lowe. 2021. "quanteda: Quantitative Analysis of Textual Data." Version 3.0.0. Available at: `https://cran.r-project.org/web/packages/quanteda/index.html`.

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

Black, Ryan C., Ryan J. Owens, Justin Wedeking and Patrick C. Wohlfarth. 2016. *Supreme Court Opinions and Their Audiences.* Cambridge, UK: Cambridge University Press.

Box-Steffensmeier, Janet and Dino P. Christenson. 2012. "Database on Supreme Court Amicus Curiae Briefs." Version 1.0 [Computer File]. Available at: `amicinetworks.com`.

Clark, Tom S. 2009. "The Separation of Powers, Court Curbing, and Judicial Legitimacy." *American Journal of Political Science* 53(4): 971–989.

Collins, Paul M. 2008. *Friends of the Supreme Court: Interest Groups and Judicial Decision Making.* Oxford, UK: Oxford University Press.

Cook, William D. 2015. "The Inheritance Bequeathed to William G. Cochran that he Willed Forward and Left for Others to Will Forward Again: The Limits of Observational Studies that Seek to Mimic Randomized Experiments." *Observational Studies* 1(1): 141–164.

Cox, Gary W. and Matthew D. McCubbins. 2005. *Setting the Agenda: Responsible Party Government in the US House of Representatives.* Cambridge, UK: Cambridge University Press.

Epstein, Lee, Jeffrey A. Segal, Harold J. Spaeth and Thomas G. Walker. 2007. *The Supreme Court Compendium.* 4th ed. Washington D.C.: CQ Press.

Harvey, Anna and Barry Friedman. 2006. “Pulling Punches: Congressional Constraints on the Supreme Court’s Constitutional Rulings, 1987-2000.” *Legislative Studies Quarterly* 31(4): 533–562.

Hausegger, Lori and Lawrence Baum. 1999. “Inviting Congressional Action: A Study of Supreme Court Motivations in Statutory Interpretation.” *American Journal of Political Science* 43(1): 162–185.

King, Chad M. 2007. “Strategic Selection of Legal Instruments on the U.S. Supreme Court.” *American Politics Research* 35(5): 621–642.

Kiss, Tibor and Jan Strunk. 2006. “Unsupervised Multilingual Sentence Boundary Detection.” *Computational Linguistics* 32(4): 485–525.

Krehbiel, Keith. 1998. *Pivotal Politics: A Theory of US Lawmaking.* Chicago, IL: University of Chicago Press.

Lee, Frances E. 2010. “Senate Deliberation and the Future of Congressional Power.” *PS: Political Science and Politics* 43(2): 227–229.

Lewis, Jeffrey B., Keith Poole, Howard Rosenthal, Adam Boche, Aaron Rudkin and Luke Sonnet. 2021. “Voteview: Congressional Roll Call Votes Database.”. Available at: voteview.com.

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.

Owens, Ryan J. 2010. “The Separation of Powers and Supreme Court Agenda Setting.” *American Journal of Political Science* 54(2): 412–427.

Owens, Ryan J. 2011. “An Alternative Perspective on Supreme Court Agenda Setting in a System of Shared Powers.” *Justice System Journal* 32(2): 183–205.

Owens, Ryan, Justin Wedeking and Patrick Wohlfarth. 2013. "How the Supreme Court Alters Opinion Language to Evade Congressional Review." *Journal of Law and Courts* 1(1): 35–59.

Popper, Karl R. 2002 [1959]. *The Logic of Scientific Discovery*. New York: Routledge.

Rice, Douglas. 2019. "Placing the Ball in Congress' Court." *American Politics Research* 47(4): 803–831.

Rosenbaum, Paul R. 2010. *Design of Observational Studies*. New York, NY: Springer.

Rosenbaum, Paul R. 2015. "Cochran's Causal Crossword." *Observational Studies* 1(1): 205–211.

Rosenbaum, Paul R. 2017. *Observation and Experiment*. Cambridge, MA: Harvard University Press.

Sala, Brian R. and James F. Spriggs. 2004. "Designing Tests of the Supreme Court and the Separation of Powers." *Political Research Quarterly* 57(2): 197–208.

Segal, Jeffrey A. 1997. "Separation-of-powers Games in the Positive Theory of Congress and Courts." *American Political Science Review* 91(1): 28–44.

Segal, Jeffrey A., Chad Westerland and Stephane A. Lindquist. 2011. "Congress, the Supreme Court, and Judicial Review: Testing a Constitutional Separation of Powers Model." *American Journal of Political Science* 55(1): 89–104.

Simmons, Joseph P., Leif D. Nelson and Uri Simonsohn. 2011. "False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant." *Psychological Science* 22(11): 1359–1366.

Spaeth, Harold J., Lee Epstein, Jeffrey A. Segal, Ted Ruger, Andrew D. Martin and Sarah Benesh. 2017. "The Supreme Court Database." Available at: `scdb.wustl.edu`.

Staton, Jeffrey K. and Georg Vanberg. 2008. "The Value of Vagueness: Delegation, Defiance, and Judicial Opinions." *American Journal of Political Science* 52(3): 504–519.

Stewart III, Charles and Jonathan Woon. 2017. “Congressional Committee Assignments, 103rd to 114th Congresses, 1993–2017: House and Senate.” Datset updated 11/17/2017. Downloaded from http://web.mit.edu/17.251/www/data_page.html#2.

Swift, Elaine K., Robert G. Brookshire, David T. Canon, Evelyn C. Fink, John R. Hibbing, Brian D. Humes, Michael J. Malbin and Kenneth C. Martis. 2009. “Database of Congressional Historical Statistics, 1789–1989.” Distributed by ICPSR. Version of 9/3/2009.

Appendix to "On Separation of Powers and Obfuscation in U.S. Supreme Court Opinions"

August 19, 2021

Abstract

This appendix presents various alternate specifications of the models in the main text. It also includes the text of the majority opinion in *Washington* v. *Recueno*.

Here I briefly summarize certain alternative specifications of the models presented in the main text. The complete set of regressions are too lengthy to present in full, but they are included with the replication code.

First, consider setting the case as the unit of analysis, and predicting an alternative dependent variable: *CLI Difference*. This is the CLI for a case's dissent(s), subtracted from the CLI of the majority opinion. This measure is positive if the dissents are more readable than the majority, and negative if the majority is more readable. Accordingly, the theoretical expectation is that CLI Difference increases when the Court becomes constrained, since that is when the majority has incentive to obfuscate and the dissent has incentive to write readable opinions.

These results, presented in Table A1 are contrary to the predictions of strategic obfuscation theory as well. (Of course, this is not really surprising, given the results in Table 2 of the main text.) For all 12 regressions analogous to those in Table 2 of the main text, the coefficient attending CLI Difference is negative. That is, as the Court becomes more constrained, dissenters write less clearly, relative to the majority opinion. For all but the three specifications where Distance To Majority Median measures constraint, the coefficient on the constraint variable is statistically distinguishable from zero.

Key IV	Baseline	Add Controls	Add Issue FEs
Distance to Filibuster Pivot	-0.829^*	-1.142^*	-1.133^*
	(.16)	(.14)	(.14)
Distance to Chamber Median	-0.525^*	-0.767^*	-0.764^*
	(.13)	(.12)	(.12)
Distance to Commitee Median	-0.487^*	-0.630^*	-0.644^*
	(.12)	(.11)	(.11)
Distance to Majority Party Median	-0.093	-0.066	-0.051
	(.12)	(.12)	(.12)

Table A1. CLI Difference (majority CLI − dissenting opnion CLI) as a function of Court majority constraint DV: CLI Difference. OLS coefficients and standard errors (clustered by term), for twelve models: four variants of a distance to relevant pivot, and three model specifications. See text for details. (*: $p < 0.05$.)

I next turn to alternative operationalizations of the constraint variables. As mentioned, in the main text I follow the operationalization in Owens, Wedeking and Wohlfarth (2013), to limit researcher degrees of freedom. However, there are other reasonable specifications of the constraint variables.

Specifically, there are good arguments that the XTI scores (Bailey 2021) better capture the relevant ideological dimension of Congress-Court relations, during the time period I consider, than do JCS scores (see Bailey 2007). As well, there are strong theoretical reasons to expect that the Court median's ideal point, not the Court majority coalition median's ideal point, gives the best estimate of Court policy location (Anderson and Tahk 2007).[1] Finally, many scholars (e.g., Segal 1997) have argued that the Court is effectively unconstrained in cases where it exercises judicial review, since the possibility of a constitutional amendment overriding such a decision is so remote.

Thus, I report here the estimates of the key interaction, for additional sets of models analogous to those in Table 2 of the main text. In Table A2 I use XTI scores (Bailey 2007; Bailey 2021) to locate justices and congresspersons. Otherwise, the relevant actors are located as in the main text. I define two versions of the constraint variables Distance To Filibuster Pivot and Distance To Chamber Median.[2] The first version is same as defined in the main text. The second version is modified so that the variable equals 0 in any case where the Court rested its decision on the U.S. constitution; i.e., the Court is by definition considered unconstrained in these cases. These differences aside, the models (and samples) are defined exactly as in Table 2.

As in the main text, the results are contrary to the prediction of strategic obfuscation theory. The consistently negative coefficients indicate, again, that dissenters write less clearly, relative to the majority, when the Court is constrained.

Finally, I repeat the analyses presented in Table A2, with the Court (majority's) ideal point identified as the ideal point of the Court median, rather than that of the majority coalition median (Anderson and Tahk 2007). Again, the results, presented in Table A3, are

[1] To be sure, there is no consensus on this point (Carrubba, Friedman, Martin and Vanberg 2012).

[2] I do not construct Distance To Committee Median and Distance To Majority Median because I lack necessary data.

Key Interaction	Baseline	Add Controls	Add FEs
Constitutional Cases Can Be Constrained			
Distance to Filibuster Pivot	−0.750*	−0.735*	−0.449*
× Majority Opinion	(.09)	(.08)	(.11)
Distance to Chamber Median	−0.470*	−0.426*	−0.253*
× Majority Opinion	(.08)	(.08)	(.08)
Constitutional Case Implies Unconstrained			
Distance to Filibuster Pivot	−0.430*	−0.423*	−0.284*
× Majority Opinion	(.10)	(.10)	(.10)
Distance to Chamber Median	−0.177*	−0.157*	−0.110
× Majority Opinion	(.06)	(.06)	(.06)

Table A2. Key coefficients from models analogous to those presented in Table 2 in the main text. In all specifications here, XTI scores are used to locate actors in ideological space. In rows three and four, I treat constitutional cases as unconstrained by definition. See text and Table 2 for details. (*: $p < 0.05$.)

inconsistent with the prediction derived from strategic obfuscation theory: As throughout, the consistently negative coefficients indicate that dissenters tend to write less clearly, relative to the majority, when the Court is constrained.

Key Interaction	Baseline	Add Controls	Add FEs
Constitutional Cases Can Be Constrained			
Distance to Filibuster Pivot	−0.997*	−0.998*	−0.607*
× Majority Opinion	(.12)	(.12)	(.16)
Distance to Chamber Median	−0.456*	−0.446*	−0.209
× Majority Opinion	(.12)	(.13)	(.11)
Constitutional Case Implies Unconstrained			
Distance to Filibuster Pivot	−0.795*	−0.797*	−0.548*
× Majority Opinion	(.17)	(.17)	(.18)
Distance to Chamber Median	−0.190	−0.198	−0.122
× Majority Opinion	(.11)	(.11)	(.10)

Table A3. Key coefficients from models analogous to those presented in Table 2 in the main text. In all specifications here, XTI scores are used to locate actors in ideological space and the Court's ideal point is set to that of the Court median. In rows three and four, I treat constitutional cases as unconstrained by definition. See text and Table 2 for details. (*: $p < 0.05$.)

I close this section with a reminder about the interpretation of the negative coefficent on the constraint-majority interaction. The negative coefficient is consistent with exactly three scenarios. First: majority and dissent CLI both increase as a function of majority constraint

(=as the majority becomes more constrained; i.e., moves further from the relevant pivot), but the dissent's increase is steeper. This is empirically the case for almost every model in Table 2 in the main text. Second: majority CLI decreases as a function of majority constraint, but dissent CLI increases. Third: majority and dissent CLI both decrease as a function of majority constraint, but the majority's decrease is steeper. The second and third scenarios obtain for several of the specifications in the Appendix. I emphasize that all three of these scenarios are inconsistent with the predictions of strategic obfuscation theory: in each, the dissenters' response to constraint is closer to the response theoretically predicted for the majority, than the response of the majority itself. Observe also that the second and third scenarios imply that even the original finding about majority obfuscation as a function of constraint is not robust.

Washington v. Recueno: Majority Opinion (Thomas).

Respondent Arturo Recuenco was convicted of assault in the second degree based on the jury's finding that he assaulted his wife "with a deadly weapon." App. 13. The trial court applied a 3-year firearm enhancement to respondent's sentence based on its own factual findings, in violation of Blakely v. Washington, 542 U.S. 296, 124 S. Ct. 2531, 159 L. Ed. 2d 403 (2004). On appeal, the Supreme Court of Washington vacated the sentence, concluding that Blakely violations can never be harmless. We granted certiorari to review this conclusion, 546 U.S. 960, 126 S. Ct. 478, 163 L. Ed. 2d 362 (2005), and now reverse.

I

On September 18, 1999, respondent fought with his wife, Amy Recuenco. After screaming at her and smashing their stove, he threatened her with a gun. Based on this incident, the State of Washington charged respondent with assault in the second degree, i.e., "intentiona[l] assault ... with a deadly weapon, to-wit: a handgun." App. 3. Defense counsel proposed, and the court accepted, a special verdict form that directed the jury to make a specific finding whether respondent was "armed with a deadly weapon at the time of the commission of the crime." Id., at 13. A "firearm" qualifies as a "deadly weapon" under Washington law. Wash.

Rev. Code 9.94A.602 (2004). But nothing in the verdict form specifically required the jury to find that respondent had engaged in assault with a “firearm,” as opposed to any other kind of “deadly weapon.” The jury returned a verdict of guilty on the charge of assault in the second degree, and answered the special verdict question in the affirmative. App. 10, 13.

At sentencing, the State sought the low end of the standard range sentence for assault in the second degree (three months). It also sought a mandatory 3-year enhancement because respondent was armed with a “firearm,” 9.94A.533(3)(b), rather than requesting the 1-year enhancement that would attend the jury’s finding that respondent was armed with a deadly weapon, 9.94A.533(4)(b). The trial court concluded that respondent satisfied the condition for the firearm enhancement, and accordingly imposed a total sentence of 39 months.

Before the Supreme Court of Washington heard respondent’s appeal, we decided Apprendi v. New Jersey, 530 U.S. 466, 120 S. Ct. 2348, 147 L. Ed. 2d 435 (2000), and Blakely, supra. In Apprendi, we held that “[o]ther than the fact of a prior conviction, any fact that increases the penalty for a crime beyond the prescribed statutory maximum must be submitted to a jury, and proved beyond a reasonable doubt.” 530 U.S., at 490, 120 S. Ct. 2348, 147 L. Ed. 2d 435. In Blakely, we clarified that “the ’statutory maximum’ for Apprendi purposes is the maximum sentence a judge may impose solely on the basis of the facts reflected in the jury verdict or admitted by the defendant.” 542 U.S., at 303, 124 S. Ct. 2531, 159 L. Ed. 2d 403 (emphasis in original). Because the trial court in this case could not have subjected respondent to a firearm enhancement based only on the jury’s finding that respondent was armed with a ”deadly weapon,” the State conceded before the Supreme Court of Washington that a Sixth Amendment violation occurred under Blakely. 154 Wn.2d 156, 162-163, 110 P.3d 188, 191 (2005). See also Tr. of Oral Arg. 10-11.

The State urged the Supreme Court of Washington to find the Blakely error harmless and, accordingly, to affirm the sentence. In State v. Hughes, 154 Wn.2d 118, 110 P.3d 192 (2005), however, decided the same day as the present case, the Supreme Court of Washington declared Blakely error to be “‘structural’ erro[r]” which “’will always invalidate the conviction.”’ Id., 154 Wash. 2d, at 142, 110 P. 3d, at 205 (quoting Sullivan v. Louisiana, 508 U.S. 275, 279, 113 S. Ct. 2078, 124 L. Ed. 2d 182 (1993)). As a result, the court refused to apply harmless-error analysis to the Blakely error infecting respondent's sentence. Instead, it vacated his sentence and remanded for sentencing based solely on the deadly weapon enhancement. 154 Wash. 2d, at 164, 110 P. 3d, at 192.

II

Before reaching the merits, we must address respondent's argument that we are without power to reverse the judgment of the Supreme Court of Washington because that judgment rested on adequate and independent state-law grounds. Respondent claims that at the time of his conviction, Washington state law provided no procedure for a jury to determine whether a defendant was armed with a firearm. Therefore, he contends, it is impossible to conduct harmless-error analysis on the Blakely error in his case. Respondent bases his position on Hughes, in which the Supreme Court of Washington refused to "create a procedure to empanel juries on remand to find aggravating factors because the legislature did not provide such a procedure and, instead, explicitly assigned such findings to the trial court." 154 Wash. 2d, at 151, 110 P. 3d, at 209. Respondent contends that, likewise, the Washington Legislature provided no procedure by which a jury could decide at trial whether a defendant was armed with a firearm, as opposed to a deadly weapon.

It is far from clear that respondent's interpretation of Washington law is correct. See State v. Pharr, 131 Wash. App. 119, 124-125, 126 P.3d 66, 69 (2006) (affirming the trial court's imposition of a firearm enhancement when the jury's special verdict reflected a finding that the defendant was armed with a firearm). In Hughes, the Supreme Court of Washington carefully avoided reaching the conclusion respondent now advocates, instead expressly recognizing that "[w]e are presented only with the question of the appropriate remedy on remand –we do not decide here whether juries may be given special verdict forms or interrogatories to determine aggravating factors at trial." Id., 154 Wash. 2d, at 149, 110 P. 3d, at 208. Accordingly, Hughes does not appear to foreclose the possibility that an error could be found harmless because the jury which convicted the defendant would have concluded, if given the opportunity, that a defendant was armed with a firearm.

The correctness of respondent's interpretation of Washington law, however, is not determinative of the question that the Supreme Court of Washington decided and on which we granted review, i.e., whether Blakely error can ever be deemed harmless. If respondent is correct that Washington law does not provide for a procedure by which his jury could have made a finding pertaining to his possession of a firearm, that merely suggests that respondent will be able to demonstrate that the Blakely violation in this particular case was not harmless. See Chapman v. California, 386 U.S. 18, 24, 87 S. Ct. 824, 17 L. Ed. 2d 705 (1967). But that does not mean that Blakely error–which is of the same nature, whether it involves a fact that state law permits to be submitted to the jury or not–is structural, or that we are precluded from deciding that question. Thus, we need not resolve this open question of Washington law. 1

1 Respondent's argument that, as a matter of state law, the Blakely v. Washington, 542 U.S. 296, 124 S. Ct. 2531, 159 L. Ed. 2d 403 (2004), error was not harmless remains open to him on remand.

III

[4] We have repeatedly recognized that the commission of a constitutional error at trial alone does not entitle a defendant to automatic reversal. Instead, "'most constitutional errors can be harmless."' Neder v. United States, 527 U.S. 1, 8, 119 S. Ct. 1827, 144 L. Ed. 2d 35 (1999) (quoting Arizona v. Fulminante, 499 U.S. 279, 306, 111 S. Ct. 1246, 113 L. Ed. 2d 302 (1991)). "'[I]f the defendant had counsel and was tried by an impartial adjudicator, there is a strong presumption that any other [constitutional] errors that may have occurred are subject to harmless-error analysis."' 527 U.S., at 8, 119 S. Ct. 1827, 144 L. Ed. 2d 35 (quoting Rose v. Clark, 478 U.S. 570, 579, 106 S. Ct. 3101, 92 L. Ed. 2d 460 (1986)). Only in rare cases has this Court held that an error is structural, and thus requires automatic reversal. 2 In such cases, the error "necessarily render[s] a criminal trial fundamentally unfair or an unreliable

vehicle for determining guilt or innocence." Neder, supra, at 9, 119 S. Ct. 1827, 144 L. Ed. 2d 35 (emphasis).

2 See Neder v. United States, 527 U.S. 1, 8, 119 S. Ct. 1827, 144 L. Ed. 2d 35 (1999) (citing Johnson v. United States, 520 U.S. 461, 468, 117 S. Ct. 1544, 137 L. Ed. 2d 718 (1997), in turn citing Gideon v. Wainwright, 372 U.S. 335, 83 S. Ct. 792, 9 L. Ed. 2d 799 (1963) (complete denial of counsel); Tumey v. Ohio, 273 U.S. 510, 47 S. Ct. 437, 71 L. Ed. 749, 5 Ohio Law Abs. 159, 5 Ohio Law Abs. 185, 25 Ohio L. Rep. 236 (1927) (biased trial judge); Vasquez v. Hillery, 474 U.S. 254, 106 S. Ct. 617, 88 L. Ed. 2d 598 (1986) (racial discrimination in selection of grand jury); McKaskle v. Wiggins, 465 U.S. 168, 104 S. Ct. 944, 79 L. Ed. 2d 122 (1984) (denial of self-representation at trial); Waller v. Georgia , 467 U.S. 39, 104 S. Ct. 2210, 81 L. Ed. 2d 31 (1984) (denial of public trial); Sullivan v. Louisiana, 508 U.S. 275, 113 S. Ct. 2078, 124 L. Ed. 2d 182 (1993) (defective reasonable-doubt instruction)).

We recently considered whether an error similar to that which occurred here was structural in Neder, supra. Neder was charged with mail fraud, in violation of 18 U.S.C. 1341; wire fraud, in violation of 1343; bank fraud, in violation of 1344; and filing a false income tax return, in violation of 26 U.S.C. 7206(1). 527 U.S., at 6, 119 S. Ct. 1827, 144 L. Ed. 2d 35. At Neder's trial, the District Court instructed the jury that it "'need not consider"' the materiality of any false statements to convict Neder of the tax offenses or bank fraud, because materiality "'is not a question for the jury to decide."' Ibid. The court also failed to include materiality as an element of the offenses of mail fraud and wire fraud. Ibid. We determined that the District Court erred because under United States v. Gaudin, 515 U.S. 506, 115 S. Ct. 2310, 132 L. Ed. 2d 444 (1995), materiality is an element of the tax offense that must be found by the jury. We further determined that materiality is an element of the mail fraud, wire fraud, and bank fraud statutes, and thus must be submitted to the jury to support conviction of those crimes as well. Neder, 527 U.S., at 20, 119 S. Ct. 1827, 144 L. Ed. 2d 35. We nonetheless held that harmless-error analysis applied to these errors, because "an instruction that omits an element of the offense does not necessarily render a criminal trial fundamentally unfair or an unreliable vehicle for determining guilt or innocence." Id., at 9, 119 S. Ct. 1827, 144 L. Ed. 2d 35. See also Schriro v. Summerlin, 542 U.S. 348, 355-356, 124 S. Ct. 2519, 159 L. Ed. 2d 442 (2004) (rejecting the claim that Ring v. Arizona, 536 U.S. 584, 122 S. Ct. 2428, 153 L. Ed. 2d 556 (2002), which applied Apprendi to hold that a jury must find the existence of aggravating factors necessary to impose the death penalty, was a “”watershed rul[e] of criminal procedure" implicating the fundamental fairness and accuracy of the criminal proceeding,"' in part because we could not "confidently say that judicial factfinding seriously diminishes accuracy").

[5] The State and the United States urge that this case is indistinguishable from Neder. We agree. Our decision in Apprendi makes clear that "[a]ny possible distinction between an 'element' of a felony offense and a 'sentencing factor' was unknown to the practice of criminal indictment, trial by jury, and judgment by court as it existed during the years surrounding our Nation's founding." 530 U.S., at 478, 120 S. Ct. 2348, 147 L. Ed. 2d 435 (footnote omitted). Accordingly, we have treated sentencing factors, like elements, as facts that have to be tried to the jury and proved beyond a reasonable doubt. Id., at 483-484, 120 S. Ct. 2348, 147 L. Ed. 2d 435. The only difference between this case and Neder is that in Neder, the prosecution failed to prove the element of materiality to the jury beyond a reasonable doubt, while here the prosecution failed to prove the sentencing factor of "armed with a firearm" to the jury beyond a reasonable doubt. Assigning this distinction constitutional significance cannot be reconciled with our recognition in Apprendi that elements and sentencing factors must be treated the same for Sixth Amendment purposes. 3

3 Respondent also attempts to evade Neder by characterizing this as a case of charging error, rather than of judicial factfinding. Brief for Respondent 16-19. Because the Supreme Court of Washington treated the error as one of the latter type, we treat it similarly. See 154 Wn. 2d 156, 159-161, 110 P. 3d 188, 189-190 (2005) (considering "whether imposition of a firearm enhancement without a jury finding that Recuenco was armed with a firearm beyond a reasonable doubt violated Recuenco's Sixth Amendment right to a jury trial as defined by

Apprendi v. New Jersey, 530 U.S. 466, [120 S. Ct. 2348, 147 L. Ed. 2d 435 (2000)], and its progeny," and whether the Apprendi and Blakely error, if uninvited, could "be deemed harmless").

Respondent attempts to distinguish Neder on the ground that, in that case, the jury returned a guilty verdict on the offense for which the defendant was sentenced. Here, in contrast, the jury returned a guilty verdict only on the offense of assault in the second degree, and an affirmative answer to the sentencing question whether respondent was armed with a deadly weapon. Accordingly, respondent argues, the trial court's action in his case was the equivalent of a directed verdict of guilt on an offense (assault in the second degree while armed with a firearm) greater than the one for which the jury convicted him (assault in the second degree while armed with any deadly weapon). Rather than asking whether the jury would have returned the same verdict absent the error, as in Neder, respondent contends that applying harmless-error analysis here would "'hypothesize a guilty verdict that [was] never in fact rendered,"' in violation of the jury-trial guarantee. Brief for Respondentt 27 (quoting Sullivan, 508 U.S., at 279, 113 S. Ct. 2078, 124 L. Ed. 2d 182).

We find this distinction unpersuasive. Certainly, in Neder, the jury purported to have convicted the defendant of the crimes with which he was charged and for which he was sentenced. However, the jury was precluded "from making a finding on the actual element of the offense." 527 U.S., at 10, 119 S. Ct. 1827, 144 L. Ed. 2d 35. Because Neder's jury did not find him guilty of each of the elements of the offenses with which he was charged, its verdict is no more fairly described as a complete finding of guilt of the crimes for which the defendant was sentenced than is the verdict here. See id., at 31, 119 S. Ct. 1827, 144 L. Ed. 2d 35 (Scalia, J., concurring in part and dissenting in part) ("[S]ince all crimes require proof of more than one element to establish guilt. .. it follows that trial by jury means determination by a jury that all elements were proved. The Court does not contest this"). Put another way, we concluded that the error in Neder was subject to harmless-error analysis, even though the District Court there not only failed to submit the question of materiality to the jury, but also mistakenly concluded that the jury's verdict was a complete verdict of guilt on the charges and imposed sentence accordingly. Thus, in order to find for respondent, we would have to conclude that harmless-error analysis would apply if Washington had a crime labeled "assault in the second degree while armed with a firearm," and the trial court erroneously instructed the jury that it was not required to find a deadly weapon or a firearm to convict, while harmless error does not apply in the present case. This result defies logic. 4

4 The Supreme Court of Washington reached the contrary conclusion based on language from Sullivan. See State v. Hughes, 154 Wn.2d 118, 144, 110 P.3d 192, 205 (2005) ("'There being no jury verdict of guilty-beyond-a-reasonable-doubt, the question whether the same verdict of guilty-beyond-a-reasonable-doubt would have been rendered absent the constitutional error is utterly meaningless. There is no object, so to speak, upon which harmless-error scrutiny can operate"' (quoting Sullivan, 508 U.S., at 280, 113 S. Ct. 2078, 124 L. Ed. 2d 182)). Here, as in Neder, "this strand of reasoning in Sullivan does provide support for [respondent]'s position." 527 U.S., at 11, 119 S. Ct. 1827, 144 L. Ed. 2d 35. We recognized in Neder, however, that a broad interpretation of our language from Sullivan is inconsistent with our case law. 527 U.S., at 11-15, 119 S. Ct. 1827, 144 L. Ed. 2d 35. Because the jury in Neder, as here, failed to return a complete verdict of guilty beyond a reasonable doubt, our rejection of Neder's proposed application of the language from Sullivan compels our rejection of this argument here.

* * *

Failure to submit a sentencing factor to the jury, like failure to submit an element to the jury, is not structural error. Accordingly, we reverse the judgment of the Supreme Court of Washington and remand the case for further proceedings not inconsistent with this opinion.

It is so ordered.

References

Anderson, Robert and Alexander M. Tahk. 2007. "Institutions and Equilibrium in the United States Supreme Court." *American Political Science Review* 101(4): 811–825.

Bailey, Michael A. 2007. "Comparable Preference Estimates Across Time and Institutions for the Court, Congress, and Presidency." *American Journal of Political Science* 51(3): 433–448.

Bailey, Michael A. 2021. "Bridge Ideal Points." Available at: `https://michaelbailey.georgetown.domains/data/`.

Carrubba, Clifford, Barry Friedman, Andrew D. Martin and Georg Vanberg. 2012. "Who Controls the Content of Supreme Court Opinions?" *American Journal of Political Science* 56(2): 400–412.

Owens, Ryan, Justin Wedeking and Patrick Wohlfarth. 2013. "How the Supreme Court Alters Opinion Language to Evade Congressional Review." *Journal of Law and Courts* 1(1): 35–59.

Segal, Jeffrey A. 1997. "Separation-of-powers Games in the Positive Theory of Congress and Courts." *American Political Science Review* 91(1): 28–44.