

**Testing Spatial Models of Politics:
The Perils of Subtraction-Based Measures of Subjective Ideological Proximity**

Michael J. Nelson

Assistant Professor
Department of Political Science
The Pennsylvania State University
232 Pond Laboratory
University Park, PA 16802
United States
mjn15@psu.edu

James L. Gibson

Sidney W. Souers Professor of Government
Department of Political Science
Washington University in St. Louis
Campus Box 1063
St. Louis, MO 63130-4899
United States
jgibson@wustl.edu

Fellow, Centre for Comparative and International Politics
Professor Extraordinary in Political Science
Stellenbosch University (South Africa)

Version 140, May 20, 2017

©

* We greatly appreciate the support of Steve Smith and the Weidenbaum Center for this research. One survey on which this article is based was made possible by a grant from TESS: Time-Sharing Experiments for the Social Sciences (NSF Grant 0818839, Jeremy Freese and James Druckman, Principal Investigators), to whom we are much indebted. We also acknowledge the assistance of Patrick Tucker in preparing these data for analysis, and we appreciate the comments of Pamela Corley, Chris Fariss, Charles Crabtree, Deborah Beim, Markus Neumann, Jeffrey Ziegler, and Miguel Maria Pereira on an earlier version of portions of this analysis.

Abstract

Spatial theories of politics fuel our understanding of a great deal of political behavior. Tests of these theories require measures of ideological proximity that are valid and reliable, yet little attention has been paid to the amount of measurement error in widely-used measures of subjective ideological proximity. The literature on political knowledge, which finds that much of the public is ignorant about broad swaths of American politics, suggests that these measures rely upon heroic assumptions. This, in turn, implies that indicators derived from such procedures may have significant validity limitations. Using a recent controversy over the legitimacy of the U.S. Supreme Court as a venue, we examine the accuracy of several common measures of subjective ideological proximity. We find that 40% of respondents reject estimates derived for them, and the measurement error is systematic enough to confound the relationship between ideological proximity and support for the U.S. Supreme Court.

6498 Words

Spatial models of politics rely on the axiom that individuals prefer political candidates, policies, or institutions whose ideological positions are closer to their own than those things that are farther away from them in ideological space. This fundamental assumption drives Black's (1948) median voter theory, our understanding of voter behavior in elections (Downs 1957), models of the policy process (Krehbiel 1998), attitudes toward political institutions (Bartels and Johnston 2013), elite behavior (Krehbiel 1991), and political representation (Segal, Cameron, and Cover 1992). These theories suggest that ideological proximity fuels innumerable political phenomena.

Tests of these theories require reliable and valid measures of ideological proximity. The predominant way to measure this concept is to acquire separate indicators of the perceived ideological locations of the various individuals, institutions, policies, or candidates and then simply subtract them, creating a measure of ideological distance. Such a subtraction-based procedure is routine in applications as varied as that of the relationship between the U.S. Supreme Court and Congress (Segal 1997; Owens 2010), voters' decisions in congressional elections (Adams et al. 2011), Senate actions on bureaucratic nominations (Ostrander 2016), and models of the policy process (Krehbiel 1998).

But other theories of politics—especially in the arena of political behavior—rely not on *objective* measures of ideological proximity but instead on *subjective* measures of this concept (Enelow and Hinich 1984). For example, Bartels and Johnston (2013) argue that individuals of every ideological stripe may rightfully perceive the U.S. Supreme Court as out of step with their views. As a result, *subjective*, rather than *objective*, disagreement with the Court's policymaking fuels attitudes about institutional legitimacy. Such subjective measures are commonplace in

studies of presidential voting (Boatright 2008) and congressional elections (Adams, Bishin, and Dow 2004). Perhaps surprisingly, while scholars have assessed the reliability and validity of various objective measures of the political positions of actors for decades (Poole and Rosenthal 2007; Epstein and Mershon 1996), the psychometrics of subjective measures of the same concept has received much less attention.

Measuring ideological locations is a challenging task, and it is especially arduous when testing spatial theories of politics with subjective measures of proximity. In these applications, the predominant mode of data collection is a survey, and researchers ask respondents to give their perceptions of various actors (including themselves) on a single, liberal-conservative continuum. The demanding nature of these questions suggests that respondents may not be able to answer the questions, or may generalize based on a few scattered pieces of information, guess entirely, or weigh some pieces of information more heavily than others (Bartels and Johnston 2013). Because respondents may rate themselves and institutions on different aspects of the ideological dimension or even fail generally to subscribe to a unidimensional policy space for these ratings (Treier and Hillygus 2009), the extent to which a subtraction-based measure of proximity is valid is unclear.

It is surprising, then, that so few studies have sought to validate subjective measures of ideological proximity. Scholars typically assume (largely implicitly) that these measures have limited random and systematic measurement error, rather than testing these assumptions directly (Gibson, Pereira, and Ziegler 2017). If these measures of ideological proximity suffer from both systematic and random measurement error, then doubt would be cast on the conclusions derived

from such measures. Hence, assessing the nature of the measurement error in individual-level indicators of subjective ideological proximity is a vital task.

We take up that challenge here, relying on an application that has been the source of recent debate in the study of judicial politics. Briefly, despite a conventional wisdom that suggests that the legitimacy of the U.S. Supreme Court is little moved by performance dissatisfaction among the public (Caldeira and Gibson 1992), new evidence seems to suggest that ideological dissatisfaction with the Court’s policies represents a substantial threat to the legitimacy of the institution (Bartels and Johnston 2013; Christenson and Glick 2015).¹ Unfortunately, in addition to disagreeing over the substantive issue—whether subjective ideological proximity/dissatisfaction² affects legitimacy—scholars have also debated how to measure the key explanatory variable: ideological proximity (e.g., Gibson and Nelson 2015). Our hope is that this debate in judicial politics can inform the larger issue of how to measure spatial locations in other applications and substantive contexts.

Our analysis relies upon a novel approach to assessing the validity of these measures: we calculated respondents’ levels of ideological proximity to the U.S. Supreme Court during their interview using the traditional subtraction-based approach, and then asked them directly about the degree to which they accepted our calculated assessment as accurate. Moreover, we apply this technique using the variety of differing Court-location measures that have been used to

¹ For summaries of Legitimacy Theory, see Tyler (2006) and Gibson and Nelson (2014).

² This quantity is sometimes referred to as “proximity,” and sometimes as “dissatisfaction.” We use both terms interchangeably.

assess the relationship between ideological proximity and legitimacy, enabling us to evaluate the validity of subtraction-based measures of ideological proximity across a variety of different question wordings involved in the judicial politics debate.

Our conclusions are telling. About 40% of our respondents *reject* the measure of ideological proximity calculated for them, regardless of which question wording is used. Perhaps not surprisingly, those respondents who are most likely to reject their classification are the least politically sophisticated, suggesting that measurement error inherent in this subtraction approach is systematic. Indeed, we show that the direction of the estimated relationship between ideological proximity and legitimacy *is entirely different* for those respondents who are well-classified by their measure of ideological proximity compared to those who are not. Finally, we propose a simpler way to measure subjective ideological proximity. Our findings have implications for a broad swath of research on topics as varied as vote choice, representation, and elite political behavior.

Subtraction and Opinion

When testing spatial models of politics at the individual level, scholars often rely on the subjective ideological placements of survey respondents. In these surveys, proximity is not directly expressed by the respondents but is instead derived by the analyst by calculating the difference between the respondent's own ideological position and that of the candidate, policy, or institution she is asked to rate. Then, if the analyst is interested in the degree of ideological proximity rather than the direction, he or she will calculate the absolute value or the square of the

difference between the respondent's stated ideological location and the perceived location of the candidate, policy, or institution. If respondents reliably and validly provide the two ratings, then the measure of ideological proximity that results is an appropriate one for scholars to use. If, however, those ratings are error prone, then there is reason to doubt their use, and, by extension, the substantive conclusions they generate.³

Indeed, there are reasons to doubt the extent to which a simple subtraction-based procedure results in an accurate measure of ideological proximity. Most obviously, respondents may guess, rationally satisfice by relying on heuristics to judge institutional performance, or generalize from a few salient decisions to their assessments of an institution's global performance (Bartels and Johnston 2013). This likelihood raises two important concerns. First, respondents' subjective beliefs about institutional performance may well differ from an objective measure of institutional performance, especially when respondents care more about some issues than others.⁴ For example, a conservative respondent might judge the 2015 U.S. House of Representatives as very liberal because they did not successfully repeal the Affordable Care Act even though the institution was, by objective measures, fairly conservative. Second, it suggests

³ We acknowledge the debate among scholars about the conditions under which objective and subjective measures of distance are appropriate (e.g., Gershtenson 2009). We focus our efforts here on the best way to measure proximity subjectively.

⁴ Differences between subjective and objective measures have been noted in other applications, such as citizens' perceptions of economic performance (e.g. Nadeau and Lewis-Beck 2001).

that respondents' views of the institution may well not fall neatly on a simple unidimensional space. Indeed, survey researchers continue to debate the unidimensional nature of ideology, with Treier and Hillygus (2009) finding that "failing to account for the multidimensional nature of ideological preferences can produce inaccurate predictions" of political behavior (697). Moreover, respondents may respond that an institution is "liberal" based off of its action on a single issue of importance (and salience) to the respondent, even though the respondent is aware of other, countervailing decisions that happen to be of a lesser degree of importance to them at the time of the interview.

Additionally, these measures rely on the assumption that respondents are rating themselves and the institution or policy they are asked to rate on the *same* unidimensional ideological scale. A good deal of survey research (e.g., Krosnick and Presser 2010), has shown that even the most minute changes to survey design, such as changing the endpoints of a survey scale, can have drastic effects on the responses given by respondents, in large part because it affects the anchor from which respondents formulate their answer. It is unclear, when respondents are asked to rate the ideological position of a candidate, whether their judgment is based upon a comparison between themselves and the candidate, between the candidate and the opposing candidate, between the candidate and the institution, or between the candidate and something else. Yet, subtraction-based measures of disagreement uniformly assume the former.

To compound the issue, the extent to which respondents actually employ a simple, subtraction-based distance calculation is unclear. The literature on directional voting (e.g. Rabinowitz and Macdonald 1989), for example, suggests that respondents do not engage in a

simple distance calculation when deciding for whom to cast their vote. Instead, the theory argues, voters privilege direction over distance. Yet, despite this potential for additional systematic measurement error, subtraction-based measures of ideological proximity abound.

Thus, the central hypothesis guiding this research is that subtraction-based measures of subjective ideological proximity are contaminated with measurement error and therefore threaten the substantive conclusions scholars draw about spatial models of politics.

Subjective Ideological Proximity and Diffuse Support

The substantive application for our investigation concerns the relationship between institutional support for the U.S. Supreme Court (diffuse support) and evaluations of the outputs of the institutions (specific support). Conventional wisdom, based upon Legitimacy Theory, holds that the relationship between the two concepts is negligible; legitimacy is “sticky,” and the Court’s legitimacy fails to respond directly and quickly to short-term fluctuations in performance satisfaction (Caldeira and Gibson 1992; Gibson and Nelson 2015).

Recent studies challenge this conventional wisdom, positing that there is a “potent ideological foundation” to the Supreme Court’s legitimacy (e.g., Bartels and Johnston 2013, 193). These studies present evidence that the relationship between subjective performance satisfaction and legitimacy is significant and meaningful, with individuals who become dissatisfied with the Court’s ideological tilt withdrawing legitimacy from the institution (Christenson and Glick 2015). This finding, of course, has considerable consequences for understanding the role of the Supreme Court in the American political process.

Other studies are not so certain. Finding little relationship between performance evaluations and institutional support, Gibson and Nelson (2015) conclude that “[t]he legitimacy of the Court is not overly dependent upon perceptions and evaluations of its performance . . . institutional support is primarily grounded in more fundamental and obdurate democratic values, and is therefore resistant to change” (163). This view is consistent with the notion of a “reservoir of goodwill” – existing support for the Court is not dislodged much by individual judicial decisions, even if sustained dissatisfaction can undermine support (as Gibson and Caldeira 1992 report happening with African Americans).

Common among all of these studies is their use of a subtraction-based measure of subjective ideological proximity with the U.S. Supreme Court as a key explanatory variable for legitimacy. But, unfortunately, the measure of ideological proximity used both differs across studies and is closely associated with the substantive conclusions the study reaches. The two studies that find a strong relationship between ideological proximity and legitimacy, Bartels and Johnston (2013) and Christensen and Glick (2015), both rely on the following question: “Judging by its recent decisions, do you think the Supreme Court is generally liberal, generally conservative, or *is it making decisions more on a case-by-case basis?*” (emphasis added). Gibson and Nelson (2015) criticize this question, arguing that it creates a heterogeneous middle category, housing respondents who believe that the Court is legalistic rather than ideological (and therefore decides cases on a “case-by-case” basis), those respondents who believe that the Court’s policymaking is ideological but moderate, as well as those who simply do not know

where the Court stands and use the middle category on the response set to disguise their ignorance.

Gibson and Nelson (2015), who find that the relationship between ideological proximity and institutional support is, at best, very weak, drew their measure from the following question: “Thinking about the United States Supreme Court in Washington and the decisions that it has been making lately, would you say that the Supreme Court is a very liberal court, a somewhat liberal court, a somewhat conservative court, or a very conservative court.” This question has also come under fire for its omission of a middle category, the lack of which fails to provide respondents a stated option for “moderate” policymaking (see Bartels, Johnston, and Mark 2015, 777, footnote 13).

Given that those scholars who claim a strong linkage between perceived ideological location and legitimacy have relied upon the “case-by-case” question while those who have failed to find such a strong connection have relied upon a question that does not offer respondents a middle category, this controversy provides a valuable testing ground for examining the use of “subtraction-based” measures of ideological dissatisfaction. By examining the accuracy of the dissatisfaction measure across alternative question wordings (including those that remedy the criticisms of the original measures), we are able to (1) assess the utility of subtraction-based measures of ideological proximity and (2) determine whether these concerns about measurement error are exacerbated or mitigated by question-wording effects. As such, our analysis both helps to resolve a salient debate surrounding institutional support for the U.S.

Supreme Court while also having broader implications for testing spatial models of politics more generally.

Research Design

This paper relies on data collected by The American Panel Study (TAPS), a monthly online survey. Panelists were first recruited as a national probability sample with an address-based sampling frame in the fall of 2011 by Knowledge Networks for the Weidenbaum Center at Washington University. Individuals without internet access were provided a laptop and internet service at the expense of the Weidenbaum Center. More technical information about the survey is available at taps.wustl.edu.

The respondents in the TAPS survey were randomly assigned to one of four unequally-sized conditions: (1) the three-category, “case-by-case” version of the ideological location question (and follow-ups): N = 678. (2) the four-category version (and follow-ups): N = 342. (3) the “case-by-case” question and then the four-category version (with follow-ups): N = 257. (4) the four-category question, followed by the “case-by-case” question (with follow-ups): N = 276. For analytical purposes, we combine some of these categories of respondents. We only do so when the respondents were asked the same questions in the same order to eliminate question-order effects.⁵

⁵ Chi-squared yielded no evidence of question-order effects. For the four-category question, the p-value is .61; for the “case-by-case” question, the p-value is .58. There are no significant pairwise differences by condition.

An important innovation of our research design is the use of a question in the survey instrument to directly validate the measure of ideological proximity. After the respondents were queried about both their placement of the Court in ideological space and their own ideological self-placement, the survey software created a *difference score* by subtracting the respondent's own self-placement from the perceived location of the Court. Using these difference scores, respondents were asked the following question:

According to your earlier answers to our questions about you and the U.S. Supreme Court, we determine that you regard the Court as [FILL IN THE CALCULATION]. How strongly would you say the statement that the Court is [FILL IN THE CALCULATION] actually represents your views?

The respondents were asked whether that placement “represents my views extremely well,” “represents my views fairly well,” “does not represent my views very well,” or “definitely does not represent my views.” The respondents’ answers to this question constitute their *reported self-validation*. Because this technique queries respondents directly about the validity of the measure, we can determine (1) the extent to which respondents judge subtraction-based measures of ideological proximity to be accurate, (2) whether particular types of respondents are especially likely to object to their calculated difference score, and (3) whether misclassification using this method might affect substantive conclusions scholars draw. Below, we investigate all three issues.

Measures of Key Concepts

Subjective Ideological Proximity

We used five different measures of the respondents’ placement of the U.S. Supreme Court, and,

from those measures, we construct five indicators of ideological proximity. Three of the measures are those used by Bartels and Johnston (2013), Christenson and Glick (2015), and Gibson and Nelson (2015). The other two draw upon follow-up questions that respond to the criticisms of the question wordings noted above. The five measures are: (1) the four-category “case-by-case” (Bartels and Johnston) measure, which ranges from “strong agreement” to “strong disagreement,” (2) a five-category measure that asks respondents a follow-up on the middle, “case-by-case” response to the Bartels and Johnston question, (3) a seven-point measure that mirrors the Christenson and Glick measure with the use of follow-up questions to the original Bartels and Johnston item, (4) the four-point Gibson and Nelson scale, and (5) a five-category measure that relies upon a follow-up to the “somewhat liberal” and “somewhat conservative” answers to the Gibson and Nelson question.⁶ To reiterate, respondents were randomly assigned to their question wording, thus enabling us to draw clear conclusions about question-wording effects.⁷

⁶ We have scored all variables in this analysis to range from 0 to 1.

⁷ Conditions 3 and 4 asked the follow-up questions necessary to create the multi-point scales *after* the respondents were asked the opposing question, while Conditions 1 and 2 asked no potentially confounding questions before respondents were asked the ideological location questions and its follow-ups. Because we are especially cautious about the potential for question-order effects, we rely on respondents from Conditions 1 and 2 when we use these measures of ideological proximity, but the full set of respondents when we analyze the difference score.

The measures of ideological dissatisfaction with the Court are a function of the respondent's own ideological position and her or his perception of the ideological location of the Court. We replicate Bartels and Johnston's methodology precisely when using that question wording.⁸ The comparison of this approach with our respondents and theirs is revealing: strong disagreement, 27.0% versus 27.8%; moderate disagreement, 12.9% versus 13.3%; tacit agreement, 27.2% versus 32.9%; and strong agreement, 32.9% versus 26.1%, for our survey versus their survey, respectively. Thus, our findings regarding the distribution of ideological dissatisfaction are remarkably similar to theirs, providing some evidence of stability at the aggregate level in the American people's perceptions of their ideological disagreement with the Court between 2005, when the Bartels and Johnston survey was conducted, and 2014, when the TAPS survey was fielded. Though journalists and scholars have documented a conservative turn in the Court's jurisprudence over early part of the past decade (Liptak 2010), it appears that the public sees itself as about as equally distant from the Court today as it did a decade ago.

⁸ Bartels and Johnston place respondents into four categories based upon their answers to the three-point ideology and Supreme Court ideology questions. Those who give the same answers to both questions are in "Strong Agreement," those who identify as liberal or conservative but believe the Court is moderate are in "Tacit Agreement," those who are moderate but believe the Court to be liberal or conservative are in "Moderate Disagreement," and those who are liberal but believe the Court is conservative (and vice-versa) are in "Strong Disagreement."

Measuring ideological proximity with the other four measures is relatively easy. We use the absolute value of the difference between the Court and self-ideological placement responses for the Gibson and Nelson and Christenson and Glick measures. One might be concerned that this procedure attenuates correlations; to this end, we have also re-estimated all models that employ the absolute value measure with a squared measure (distance²). The results are identical.

Institutional Support

We have measured the diffuse support of our respondents (e.g., Gibson, Caldeira, and Spence 2003) so as to be able to correlate that support with various measures of ideological distance from the U.S. Supreme Court and other variables. Appendix A reports the indicators of institutional support and their univariate frequencies, and evidence of reliability.

Our survey also included the conventional measures of specific support: assessments of how well the Court is doing its job (performance satisfaction) and another question asking about respondents' global judgments about whether the Court's decisions are "just right" (as opposed to "too liberal" or "too conservative"). Most of the respondents (65%) judged the Court to be doing at least a "pretty good job," although only 42% rate the Court's decisions as "about right." The two measures of specific support are moderately related at .35. Assessments of how well the Court is doing its job are correlated with the diffuse support index at .40.

Validating Reported Self-Validation

Our research design relies upon a comparison between a derived, subtraction-based measure of subjective ideological proximity and respondents' answers to the self-validation question. Some may be concerned with the use of the self-validation as a comparison – in essence, as a criterion validity variable. Perhaps this measure itself suffers from systematic measurement error, suggesting that any differences we find between the measures are due to error in our criterion variable rather than in the calculated measure of ideological distance. We defend this decision on both theoretical and empirical grounds.

First, as Bartels and Johnston (2013) have ably documented, the key concept these studies seek to measure is respondents' *perceptions* of ideological distance. Because subjective and objective judgments may differ, validating the calculated measure of ideological distance with an objective measure of the concept would not achieve our ends (for a study with that goal, see Jesse and Malhotra 2013).⁹

Second, our approach is analogous to experiments that use a manipulation check to gauge whether an experimental stimulus was received as it was intended. The “stimuli” in our validation approach are the separate calculated measures of ideological placement, and our

⁹ Similarly, Christenson and Glick focus on subjective assessments of the Court's location. In their study, conservatives could easily view the Court as moving to the right as a result of the decision and therefore increase their support for the Court.

“manipulation check” is the reported self-validation question.

Third, we offer several empirical tests of the validity of the reported self-validation measure based on a widely-employed measure of subjective ideological proximity as a criterion variable. Scholars of institutional legitimacy have long included in their surveys a simple question that asks respondents whether they believe the Court’s policies are “too liberal,” “too conservative,” or “about right” (e.g., Gibson and Caldeira 1992; Jessee and Malhotra 2013). This measure has the virtue of asking respondents directly about their degree of ideological proximity rather than relying on many of the assumptions outlined above, and is one simple question rather than the multiple questions required by the subtraction-based method. The question directly measures the concept: the ideological proximity of the Court relative to the position of the respondent.

We compare the strength of the relationship between respondents’ answers to the trichotomy question and the derived proximity measure *within self-validation categories*. If the self-validation measure is valid, the relationship between the respondent’s proximity score and the trichotomy measure should be stronger for those respondents with higher levels of reported self-validation. That is, individuals whose proximity score indicates that the Court is “a great deal more conservative” than they are should be more likely to answer the trichotomy by saying the Court is “too conservative” if they have high levels of reported self-validation than respondents who have low levels of reported self-validation.

Across a variety of tests, this is exactly what we find. First, in the bivariate case, the correlation between the calculated difference score and the trichotomy is .52 for respondents who

say the calculated classification “definitely does not” represent them. For respondents who say the score “does not” represent them very well, the correlation increases to .59, while the correlation balloons to .71 for those who say the calculated score “somewhat” represents them. For those saying the score “definitely” represents them, the correlation skyrockets to .83. In other words, the correlation between the difference score and the trichotomy answer is much higher for respondents who say they are well-classified by the difference score than those who reject the classification.

The same pattern holds in the multivariate case. Here, we compare the squared prediction errors from separate linear regression models run at each level of reported self-validation. If the trichotomy performs better at higher levels of reported self-validation, then the mean squared error from a regression model should be lower for models run for respondents with higher levels of reported self-validation. Model estimates are available in Appendix B.¹⁰

Again, the results support the conclusion that the trichotomy is a valid measure of subjective ideological proximity. The mean squared errors decrease as reported self-validation increases. As Figure 1 demonstrates, the mean squared error is twice as large for those respondents who reject their classification as it is for those who say it describes them very well. This difference is highly statistically significant ($p < .001$). Similarly, the standard deviation of the squared errors decreases from .54 to .31 as reported self-validation varies across its range,

¹⁰ An analysis, reported in Appendix B, relying on the percentage of trichotomy answers predicted correctly from a series of ordinal logistic regression models provides additional evidence supporting this measure.

and the predictive power of the linear regression models increases with reported self-validation. All of these metrics suggest that the trichotomy is a valid measure of ideological proximity.

[INSERT FIGURE 1 ABOUT HERE]

Results

Reported Self-Validation

To what extent do “subtraction-based” measures of subjective ideological proximity accurately classify respondents’ positions? Our first finding is that, overall, only 60.2% of the respondents say the calculated measure fits them as least fairly well, while 39.8% say the measure does not fit them at least fairly well. This is a surprisingly low level of ratification of the score derived from questions measuring the respondent’s position and the Court’s location. Indeed, 10.1% of the respondents claim that the score “definitely” does not describe their ideological relationship to the Supreme Court.

The widespread rejection of the subtraction-based classifications persists regardless of whether the measure of subjective ideological proximity was derived using the “case-by-case” or the four-category questions. Indeed, a difference-of-means test reveals that the scores do not vary by the type of question used to measure the Court’s location ($p = .190$, $N = 1,504$); nor do they vary by the condition of the experiment to which the respondent was assigned, which we might observe if question-order effects were at play ($p = .667$).

We draw two conclusions from this analysis. (1) In general, the calculated measure of ideological dissatisfaction is far from overwhelmingly accepted by the respondents: *from the*

points-of-view of the respondents, there is considerable measurement error in the ideological dissatisfaction measure. (2) The degree to which the respondents accept the calculated score does not vary significantly according to the question used to score the Supreme Court's position.

Who is Misclassified?

Having established that nearly 40% of respondents reject the accuracy of a subtraction-based measure of subjective ideological proximity, we ask whether misclassification is systematic. To this end, we conducted an ordered logistic regression analysis. The dependent variable is the four-category reported self-validation (with higher values indicating more reported validity). We assess the extent that reported self-validation is related to (1) ideological disagreement and partisanship, (2) political sophistication, including level of education and Supreme Court knowledge, (3) the respondent's belief in legal realism, to account for the possibility that individuals who reject ideological judicial decisionmaking in favor of a legalistic approach will report less validity from the measure, and (4) standard demographic factors.¹¹

The results are shown in Table 1. The most obvious initial finding from Table 1 is the fact that reported self-validity is related primarily to political sophistication.¹² The extent to

¹¹ Technical details about the measurement of belief in legal realism and Supreme Court knowledge are available in Appendix A.

¹² The results shown in Table 1 are robust to the inclusion of a variable to indicate whether the difference score was derived from the "case-by-case" or the four-category question ($p = .342$), providing further evidence that question-wording effects raise no worrisome issues.

which political sophistication, here measured by education and knowledge of the U.S. Supreme Court, predicts reported self-validation is quite large. Indeed, increasing knowledge and education from the 25th quartile to the 75th quartile triples (from .06 to .18) the probability that the respondent reports that the calculated measure “represents my views extremely well,” while decreasing the probability that a respondent reports that the measure “definitely does not represent my views” from .14 to .05. What is more, for a respondent at the first quartile on both political sophistication measures, the probability she rejects the measure’s reported self-validation is .51, suggesting that the respondent is more likely than not to reject the classification! This result dovetails well with Jessee and Malhotra’s (2013) finding that more sophisticated respondents were better able to locate the ideological position of the U.S. Supreme Court accurately and highlights the primacy of political sophistication in conditioning the level of systematic measurement error inherent in these measures.

[INSERT TABLE 1 ABOUT HERE]

Table 1 also provides some evidence of an ideological effect. Respondents who were estimated to believe that the Court is more conservative than they are tend to believe they are better classified than those who were scored that the Court is more liberal than they are. All else equal, a respondent with a difference score that indicates the Court is much more liberal than the respondent has a .50 probability of agreeing with the stated classification. For respondents who are calculated as in complete agreement with the Court, the same probability is .61. That probability rises to .71 for respondents who have a calculated difference score at its maximum (indicating that the Court is much more conservative than the respondent).

One might worry that many Americans reject the view that the Court is ideological and therefore reject their proposed classification because they also reject legal realism. We have considered this possibility, including in the model a measure of a respondent's belief in legal realism; if it were the case that a belief in legal realism leads respondents to systematically reject their classification, this variable's regression coefficient should be negative and statistically significant. That is not the case. In this model, there is no relationship between one's belief in legal realism and her self-validation score. Thus, there is no evidence that those who are not legal realists (in other words, those who believe that the Court decides cases entirely on the basis of legal factors) tend to reject attempts to classify the Court in ideological space.

Additionally, demographic characteristics have no statistically significant relationship with reported self-validation, providing no evidence that such measures are problematic for specific subgroups of respondents. Instead, political sophistication reigns supreme.

We draw one major conclusion from this analysis. As one might expect given the cognitively difficult nature of many of the assumptions in these measures, subtraction-based measures of subjective ideological proximity contain systematic measurement error. Indeed, our analysis suggests that *the most unsophisticated respondents are more likely than not to reject the subtraction-based classification!*

Does Misclassification Matter?

If misclassification affects about 40% of the respondents and is correlated with political sophistication and ideological placement, it is reasonable to wonder about the substantive

ramifications of misclassification. Put differently, does the relationship between subjective ideological proximity and Supreme Court legitimacy—the substantive debate at stake—vary depending on how valid respondents judge the measure? To consider this question, we conducted several regression analyses. The dependent variable in each analysis is the diffuse support index, and the independent variables are (1) the difference score, (2) the measure of the reported self-validation of the calculated score, (3) the interaction of the difference score with reported self-validation.¹³ Full model results, including results by question wording, are provided in Appendix B (Table B3).

Because we aim to determine whether the effect of subjective ideological proximity on legitimacy varies based upon reported self-validation, Figure 2 plots the marginal effect of ideological proximity across the four-point reported self-validation measure. In other words, each bar represents the effect of ideological proximity on diffuse support for a respondent with a particular level of reported self-validation.

[INSERT FIGURE 2 ABOUT HERE]

The figure demonstrates that the relationship between ideological disagreement and legitimacy increases in magnitude as reported self-validation increases. For those respondents who are dissatisfied with their score, ideological disagreement has *no effect*. But, for those

¹³ Appendix B contains additional analyses that use each of the individual question wordings as the measure of ideological proximity. Across all models, the conclusion is the same: the effect of ideological disagreement is strongest for respondents who have high reported self-validation.

respondents who have high levels of reported self-validation, ideological disagreement has a negative and statistically significant effect on diffuse support.

Thus, this validation exercise (1) suggests that subtraction-based measures of subjective ideological proximity are not valid for a substantial proportion of the respondents whose views they purport to represent and (2) provides some new insights into why dissatisfaction has such minor consequences for Supreme Court legitimacy. To wit, the fact that the relationship between ideological proximity and diffuse support increases in magnitude with respondents' reported self-validation suggests that a measure with less systematic measurement error might lead to a stronger relationship between the two concepts.

Discussion and Concluding Comments

Our analysis leads us to a series of conclusions. First, and most importantly, our data raise concerns about the validity of subtraction-based measures of subjective ideological proximity. Across a variety of different question wordings, we find that about 10% of respondents say that such measures “definitely” do not describe them, while about 40% of respondents reject to at least some degree their classification using such a measure. These results, which rely upon a novel survey experiment that queried respondents directly about measurement validity, suggest that the validity of these measures is far from certain.

Second, we show that the measurement error that arises from these invalid classifications is far from random, a necessary condition for a conclusion that these measures suffer from validity threats. In this sense, our results echo Alvarez and Franklin's (1994) finding that low

sophistication respondents perform poorly on ideological placement items.¹⁴ Our results amplify this conclusion, predicting that a respondent at the lowest quartile of political sophistication is *more likely than not* to judge the measure of subjective ideological proximity as inaccurate. In other words, these measures likely do a particularly poor job of judging subjective ideological proximity between individuals with low levels of political information and attention than they do the opinions of political sophisticates.

Third, we demonstrate that misclassification has important substantive consequences, with the effect of subjective ideological disagreement on diffuse support failing to reach statistical significance for those respondents who have low levels of reported self-validation. This suggests that one reason that subjective ideological proximity is so poorly related to diffuse support is measurement error in the independent variable used in earlier analyses.

If two-fifths of the respondents are misclassified by a subtraction-based measure of subjective ideological proximity, and individuals who are misclassified differ significantly from those who are correctly classified, is there a better way to measure ideological proximity? We conclude by suggesting that a simpler (and more cost-effective) measure of the concept may be superior for many purposes. Indeed, the three-category trichotomy measure we used to validate the reported self-validation measure may well be a better measure of subjective ideological

¹⁴ Following Alvarez and Franklin's (1994) results, studies seeking to use aggregated survey responses as measures of party location often rely on the judgements of the most educated respondents (e.g. Golder and Stramski 2010). Our results provide some additional evidence that this approach improves the validity of these estimates.

proximity than the subtraction-based measures discussed throughout this paper. By asking respondents to make the comparison directly rather than relying on opinions expressed across items, this new measure both relies on fewer assumptions about respondent preferences and saves space on questionnaires. Moreover, the simple trichotomy question we favor could be easily followed up with a supplemental question(s) to expand the precision of the measure of ideological proximity.

There are important limitations to our analysis. For example, there may also be measurement error in the reported self-validation score, an issue that our experimental design allows us only some leverage to detect. More importantly, though we sought to present evidence of generalizability across a variety of question wordings, we analyze only a single substantive application: the relationship between subjective ideological proximity and the legitimacy of the Supreme Court. Within this context, we conclude that question wording has only minor, at most, effects on the measurement of the concept, providing evidence to settle a major debate in the literature. Moreover, we established that respondents' rejection of their classification was unrelated to their belief in legal realism, suggesting that these findings are not hampered by our use of a legal institution to which some respondents may hesitate to ascribe an ideological position.

As far as generalizability across applications is concerned, some might argue that the debate over the attitudinal versus legal policymaking tendencies of the Court might make our findings *sui generis*. We have acknowledged this possibility though the inclusion of a measure of belief in legal realism throughout the analysis. Indeed, we tested for such differences, finding no

evidence that those respondents who rejected the notion of legal realism had systematically lower levels of self-validation. Thus, the analysis gives us no reason to conclude that the legal nature of the Court's policymaking in the eyes of some undermines the generalizability of our findings.

At a minimum, our findings raise questions about the validity of a variety of findings in diverse areas of political science and presented a potential solution to mitigate the issues we raise. Spatial models of politics are fundamental to the study of political science, and tests of these models are vital for our understanding of voting behavior, representation, and public attitudes toward political institutions. Thus, obtaining estimates of respondents' positions with minimal measurement error is as vital of a task for subjective measures of proximity as it is for objective measures of the concept. This paper provides a step toward this goal, raising questions and providing a solution that is applicable to a wide range of research agendas.

References

- Adams, James, Benjamin G. Bishin, and Jay K. Dow. 2004. "Representation in Congressional Campaigns: Evidence for Discounting/Directional Voting in U.S. Senate Elections." *Journal of Politics* 66(2): 348-373.
- Adams, James, Samuel Merrill III, Elizabeth N. Simas, and Walter J. Stone. 2011. "When Candidates Value Good Character: A Spatial Model with Applications to Congressional Elections." *Journal of Politics* 73(1): 17-30.
- Alvarez, R. Michael and Charles H. Franklin. 1994. "Uncertainty and Political Perceptions." *Journal of Politics* 56(3): 671-89.
- Bartels, Brandon L., and Christopher D. Johnston. 2013. "On the Ideological Foundations of Supreme Court Legitimacy in the American Public." *American Journal of Political Science* 57: 184-99.
- Bartels, Brandon L., Christopher D. Johnston, and Alyx Mark. 2015. "Lawyers' Perceptions of the U.S. Supreme Court: Is the Court a 'Political' Institution?" *Law and Society Review* 49 (#3): 761-794.
- Black, Duncan. 1948. "On the Rationale of Group Decision-making." *Journal of Political Economy* 56(1): 23-34.
- Boatright, Robert Guy. 2008. "Who Are the Spatial Voting Violators?" *Electoral Studies* 27: 116-125.
- Burden, Barry C., Gregory A. Caldeira, and Tim Groseclose. 2000. "Measuring the Ideologies of U.S. Senators: The Song Remains the Same." *Legislative Studies Quarterly* 25(2): 237-

258.

Caldeira, Gregory A., and James L. Gibson. 1992. "The Etiology of Public Support for the Supreme Court." *American Journal of Political Science* 36: 635-64.

Christenson, Dino P., and David M. Glick. 2015. "Chief Justice Roberts's Health Care Decision Disrobed: the Microfoundations of the Supreme Court's Legitimacy." *American Journal of Political Science* 59 (#2, April): 403-418.

Converse, Phillip E. 1964. "The Nature of Belief Systems in Mass Publics." In *Ideology and Discontent*, eds. David Apter. New York: Free Press, 206-261.

Delli Carpini, Michael and Scott Keeter. 1996. *What Americans Know About Politics and Why It Matters* New Haven, CT: Yale University Press.

Downs, Anthony. 1957. *An Economic Theory of Democracy*. Harper and Row.

Enelow, James M. and Melvin J. Hinich. 1984. *The Spatial Theory of Voting: An Introduction*. New York: Cambridge University Press.

Epstein, Lee and Carol Mershon. 1996. "Measuring Political Preferences." *American Journal of Political Science* 40(1): 261-294.

Gershtenson, Joseph. 2009. "Candidates and Competition: Variability in Ideological Voting in U.S. Senate Elections." *Social Science Quarterly* 91(1): 117-133.

Gibson, James L., and Gregory A. Caldeira. 1992. "Blacks and the United States Supreme Court: Models of Diffuse Support." *Journal of Politics* 54: 1120-45.

Gibson, James L., Gregory A. Caldeira, and Lester Kenyatta Spence. 2003. "Measuring Attitudes toward the United States Supreme Court." *American Journal of Political Science* 47:

354-67.

Gibson, James L., and Michael J. Nelson. 2014. "The Legitimacy of the U.S. Supreme Court: Conventional Wisdoms and Recent Challenges Thereto." *Annual Review of Law and Social Science* 10: 201-219.

Gibson, James L., and Michael J. Nelson. 2015. "Is the U.S. Supreme Court's Legitimacy Grounded in Performance Satisfaction and Ideology?" *American Journal of Political Science* 59: 162-74.

Gibson, James L., Miguel M. Pereira, and Jeffrey Ziegler. 2017. "Updating Supreme Court Legitimacy: Testing the 'Rule, Learn, Update' Model of Political Communication." *American Politics Research*, forthcoming.

Golder, Matt and Jacek Stramski. 2010. "Ideological Congruence and Electoral Institutions." *American Journal of Political Science* 54(1): 90-106.

Jessee, Stephen. 2016. "(How) Can We Estimate the Ideology of Citizens and Political Elites on the Same Scale?" *American Journal of Political Science* 60(4): 1108-1124.

Jessee, Stephen, and Neil Malhotra. 2013. "Public (Mis)perceptions of Supreme Court Ideology: A Method for Directly Comparing the Positions of Citizens and Justices." *Public Opinion Quarterly* 77:619-34.

Johnston, Christopher D., D. Sunshine Hillygus, and Brandon L. Bartels. 2014. "Ideology, The Affordable Care Act Ruling, and Supreme Court Legitimacy." *Public Opinion Quarterly* 78(4): 963-973.

Krehbiel, Keith. 1991. *Information and Legislative Organization*. Ann Arbor, MI: University of

- Michigan Press.
- Krehbiel, Keith. 1998. *Pivotal Politics: A Theory of U.S. Lawmaking*. Chicago, IL: University of Chicago Press.
- Krosnick, Jon A. and Stanley Presser. 2010. "Question and Questionnaire Design." In *Handbook of Survey Research*. 2nd Ed. Bingley, UK: Emerald Group Publishing: 263-313.
- Liptak, Adam. 2010. "Court Under Roberts is Most Conservative in Decades." *New York Times* 24 July. <http://www.nytimes.com/2010/07/25/us/25roberts.html?_r=0> (Accessed 18 May 2016).
- Lupia, Arthur. 2015. *Uninformed: Why People Know So Little About Politics and What We Can Do About It*. New York, NY: Oxford University Press.
- Malhotra, Neil and Stephen A. Jessee. 2014. "Ideological Proximity and Support for the Supreme Court." *Political Behavior* 36: 817-46.
- Nadeau, Richard and Michael S. Lewis-Beck. 2001. "National Economic Voting in U.S. Presidential Elections." *Journal of Politics* 63(1): 158-181.
- Ostrander, Ian. 2016. "The Logic of Collective Inaction: Senatorial Delay in Executive Nominations." *American Journal of Political Science* 60 (4): 1063-1076.
- Owens, Ryan J. 2010. "The Separation of Powers and Supreme Court Agenda Setting." *American Journal of Political Science* 54: 412-427.
- Pew Research Center. 2015. "Negative Views of Supreme Court at Record High, Driven by Republican Dissatisfaction." Washington, DC: Pew Research Center, July 29, 2015. <<http://www.people-press.org/files/2015/07/07-29-2015-Supreme-Court-release.pdf>>.

(Accessed 18 May 2016).

Poole, Keith T. and Howard L. Rosenthal. 2007. *Ideology and Congress*. 2nd ed. Transaction Publishers.

Rabinowitz, George and Stuart Elaine Macdonald. 1989. "A Directional Theory of Issue Voting." *American Political Science Review* 83(1): 93-121.

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., Charles M. Cameron, and Albert D. Cover. 1992. "A Spatial Model of Roll Call Voting: Senators, Constituents, Presidents, and Interest Groups in Supreme Court Confirmations." *American Journal of Political Science* 36(1): 96-121.

Treier, Shawn, and D. Sunshine Hillygus. 2009. "The Nature of Political Ideology in the Contemporary Electorate." *Public Opinion Quarterly* 73: 679-703.

Tyler, Tom R. 2006. "Psychological Perspectives on Legitimacy and Legitimation." *Annual Review of Psychology* 57: 375-400.

Table 1 Ordered Logistic Regression Analysis of Respondents' Reported Self-Validation

	b	s.e.	p-value
Difference Score	.11	.04	.00
Belief in Legal Realism	.07	.29	.82
Supreme Court Knowledge	1.59	.21	.00
Party Identification	.27	.18	.14
Age	.16	.32	.62
Gender	-.02	.10	.82
Black	-.32	.18	.08
Hispanic	.14	.15	.35
Education	1.06	.19	.00
Own Home	.14	.12	.25
Religious Attendance	.03	.17	.88
Cutpoint 1	-.34	.26	.18
Cutpoint 2	1.56	.25	.00
Cutpoint 3	4.19	.28	.00
BIC	3426.05		
N	1,447		

Note: The dependent variable in the analysis is reported self-validation. Higher values of the dependent variable indicate that respondents ascribe more accuracy to the measure. All independent variables range from 0 to 1 (see Table A2, Supplementary Appendices, for the summary statistics).

Fig. 1 Mean Squared Error, by Reported Self-Validation. The estimates are from separate linear regressions run for each level of reported self-validation. The dependent variable is the trichotomy measure of ideological proximity. The error bars show 95% confidence intervals.

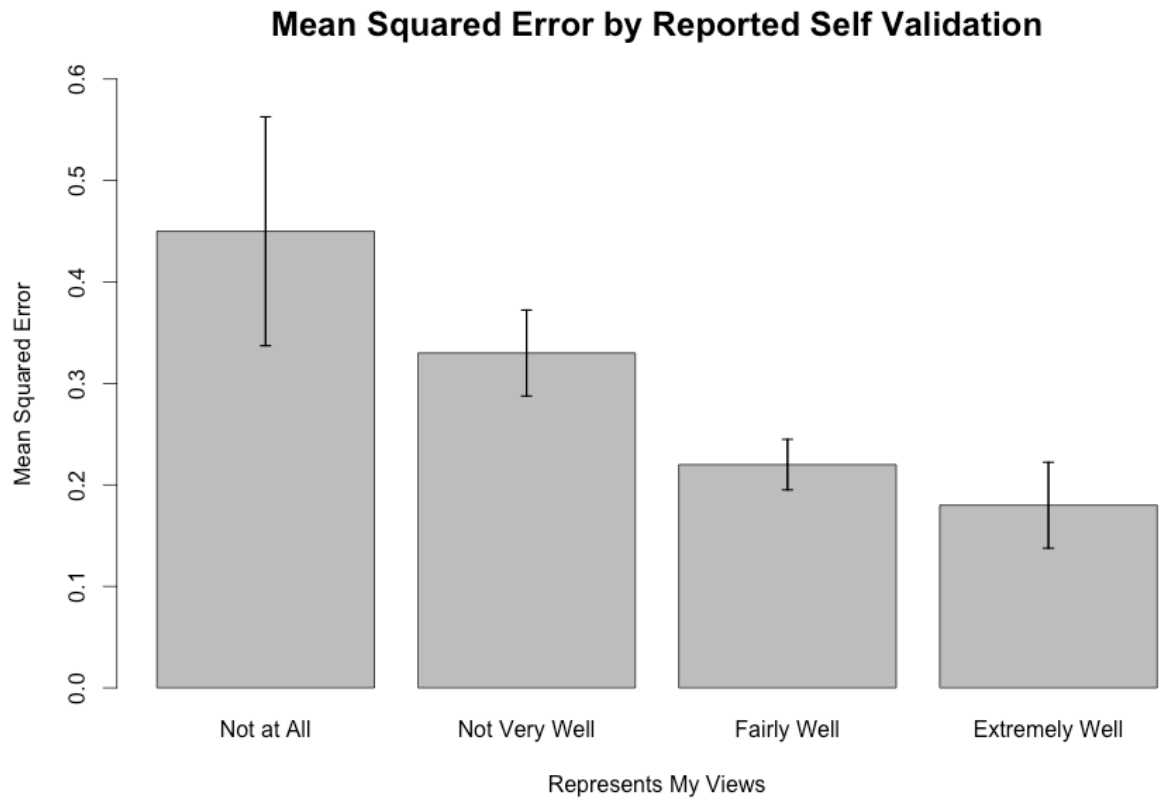


Fig. 2 The Marginal Effect of Ideological Disagreement on Diffuse Support, by Reported Self-Validation. Full model results are available in Appendix B. The measure of ideological proximity is the difference score. The error bars show 95% confidence intervals.

