"Congruence" and "Responsiveness" in the Study of Representation

John G. Matsusaka[*]
University of Southern California

This paper assesses two common methods used to measure representation in empirical research: "congruence," whether a roll-call vote is aligned with constituent opinion; and "responsiveness," the correlation between roll-call votes and constituent opinion. While sometimes viewed as two ways to capture the same underlying concept of "representation," theoretically the measures are not necessarily connected, and responsiveness can produce counterintuitive conclusions. This paper assesses to what extent responsiveness and congruence capture the same thing empirically. Using 4,192 roll-call votes on 33 state laws that were subsequently challenged in a referendum, I calculate both congruence and responsiveness for the same set of votes. The main finding is that responsiveness is an unreliable predictor of congruence, and thus of representation.

December 2024

"Congruence" and "Responsiveness" in the Study of Representation

1. Introduction

Understanding representation is a core issue in public choice research. Democracies operate under the assumption that periodic elections give elected officials the incentive to represent their constituents, an idea formalized in the foundational models of Barro (1971) and Ferejohn (1986). While theory is clear, measuring the actual amount of representation in practice is challenging. In recent years, researchers have made significant progress on this issue, developing new empirical methods and creating rich new data sets. Yet despite this progress, an unresolved tension remains at the heart of the literature, in the way representation is measured. The purpose of this paper is to highlight the tension, suggest that it may be more serious than sometimes believed, and offer suggestions for empirical researchers.

One measure of representation is "congruence," dating back at least to Achen (1977, 1978), which measures representation as the "distance" between what legislators do and what their constituents would like them to do.[1] A general definition is

$$(1) \quad CONGRUENCE = f(Y_n - Y_n^*),$$

where Y is a legislator n's action (e.g., a roll-call vote), Y_n^* is the action preferred by his or her constituents, and $f(x)$ is a function with a maximum at $Y = Y^*$ that is decreasing in distance from $x = 0$. Examples are $f(x) = -|x|$ and $f(x) = -x^2$.

Another measure is "responsiveness," derived from a regression of the form

$$(2) \quad Y_n = \alpha + \beta \cdot Y_n^* + e_n,$$

---

[1] Studies estimating congruence include Matsusaka (2010, 2020), Lax and Phillips (2012), Krimmel et al. (2016), Lax et al. (2019), and Simonovitz et al. (2019), Giger et al. (2020), Caughey and Warshaw (2022). The careful and thorough book by Caughey and Warshaw (2022) is the most comprehensive examination of congruence and responsiveness in recent years, and a good starting point for understanding the issues.

where n indexes a legislator, α and β are coefficients to be estimated, and e_n is an error term. The coefficient β is the "responsiveness." If legislators ignored citizen preferences, then $\beta = 0$, so $\beta > 0$ is a diagnostic for the existence of some amount of representation.[2]

The tension comes from the fact that although congruence and responsiveness are viewed as different ways of capturing the same underlying object – the amount of representation – theory suggests they do not capture the same thing.[3] In particular, higher responsiveness is not necessarily associated with greater representation, and responsiveness is not logically connected to congruence. This matters because a common research strategy is to estimate β for subgroups, and use it to draw conclusions about their relative representation. For example, we could estimate β separately in states with and without direct democracy, and use the two coefficients to infer which states are more representative; we could estimate β separately for Democratic and Republican legislators to determine which represent their constituents better; or we could test if representation has changed over time by comparing responsiveness in different time periods.

Of course, saying responsiveness could be an inaccurate proxy for representation in theory does not mean that it actually is in practice. Both congruence and responsiveness have an intuitive appeal, and feel like they should be related to the object we want to capture. It would give us more confidence in our metrics if turned out that they were highly correlated in practice. This paper investigates whether such an empirical link exists. The research strategy, which I believe is novel, is to compare measures of responsiveness and congruence for votes on exactly the same policy proposals.

The main empirical obstacle is measuring congruence, which requires data on legislator roll-call votes and constituent preferences on the same issues. We have an abundance of polling data about issue preferences, but polling data at the level of an individual legislative district on a specific law is rare. My solution is to use roll-call votes on laws that were subsequently challenged in a referendum: in these cases, we observe the votes of legislators and constituents on the same law. I construct a data set that includes

---

[2] Studies estimating responsiveness include Erikson et al. (1993), Tausanovitch and Warshaw (2014), and Caughey and Warshaw (2018, 2022).

[3] See Romer and Rosenthal (1979), Matsusaka (2001, 2010), and Golder and Stramski (2010).

4,192 roll-call votes, covering 33 laws in 10 states. For each law, a legislator's roll-call vote is counted as congruent if it corresponds with the district's majority position in the referendum. I then calculate responsiveness by regressing roll call votes on constituent opinion, and examine how well the responsiveness coefficients predict congruence. My purpose, then, is not to use these data to explore the determinants of representation (something I pursue elsewhere (Matsusaka, forthcoming)), but to investigate the validity of the methods we use to measure representation.

The main finding is that the responsiveness coefficient is not a reliable predictor of congruence and therefore of representation. I conduct several comparisons between different groups (male versus female legislators, votes in upper versus lower chambers, etc.) and find that the responsiveness coefficients sometimes correctly indicate which group is most congruent, sometimes bear no relation to congruence, and sometimes point in the wrong direction. The unreliable connection between responsiveness and representation calls for a sounder theoretical justification for responsiveness, and suggests that responsiveness should be used as a measure of representation with caution.

2. Theoretical Connection between Representation and Its Measures

The word "representation" has a multitude of meanings (Pitkin 1967), and even within the narrow category of *political* representation, there is a healthy debate about whether representatives should act as delegates/agents or as trustees. Nevertheless, when it comes to empirical research, researchers almost always appear to have in mind that representation "means that the representative must vote as a majority of his constituents would," as Pitkin (1967, pp. 144-145) put it. Given this, the empirical task is to measure to what extent representatives' votes mirror those that would have been cast by their constituents.

This section explores the theoretical connection between the idea of representation, so defined, and the empirical constructs of congruence and responsiveness. For expositional clarity, I work with the following intuitive definition:

Definition. *A group of legislators A are more "representative" than a group of legislators B if the roll-call votes cast by legislators A better mirror the preferences of their constituents than the roll-call votes of legislators B mirror the preferences of their constituents.*

The simplest way to implement this definition is to directly compare a legislator's vote (Y) on a law with the vote preferred by a majority of the legislator's constituents (Y^*). In the case of roll-call voting, where the choice is either yes or no, if legislator n's vote is $Y_n \in \{0,1\}$ and the preferred vote by a majority of constituents is $Y_n^* \in \{0,1\}$, then "congruence" is:[4]

$$(3) \quad CONGRUENCE_n = \begin{cases} 1 & if \quad Y_n = Y_n^*; \\ 0 & if \quad Y_n \neq Y_n^*. \end{cases}$$

The congruence of legislators A is then $CONGRUENCE(A) = \sum_{n \in A} CONGRUENCE_n$. We can compare congruence between two groups A and B, such as Democrats vs. Republicans, or between the same group at different points in time. Theoretically, such a comparison provides a valid way to distinguish which group is more representative.

Observation 1. *If $CONGRUENCE(A) > CONGRUENCE(B)$ then A is more representative than B according to the definition given above.*

Turning to responsiveness, the first thing to note is that responsiveness is not the same as congruence – indeed, as I will explain, there is no logical connection between the two. To see this, consider Figure 1a, which shows a hypothetical distribution of legislators (dots). The x-axis is the percent of the public in a district that supports a policy and the y-axis is the probability that their legislator votes for the policy. Instead of districts, one could study states, cities, counties, etc.; instead of looking across jurisdictions one could study the

---

[4] With dichotomous outcomes, the preferred vote of the majority is also the preferred vote of the median. Alternatively, one could measure congruence as $|Y_n - Y_n^*|$ where Y_n^* is the percentage of constituents in favor of the policy, which case congruence is a continuous variable between 0 and 1. The analysis in this section applies to that case as well.

same jurisdiction over time, which each observation being a different year; and instead of votes one could study policy outcomes. With congruence defined in terms of the majority's preference, the gray line shows observations that are congruent. In this figure, 9 observations are congruent and 3 are noncongruent, for an overall congruence of 75 percent.

Responsiveness is calculated by running a regression line through the data, with the slope of the line denoted "responsiveness." (With dichotomous data, one might prefer a logistic regression; the argument that follows goes through either way.) Note the conceptual difference here: *congruence is the propensity of roll-call votes to match the preferences of citizens; responsiveness is the **slope** of the relation between roll-call votes and preferences.* While congruence can be estimated for individual observations, responsiveness is a parametric characterization that can only be estimated for a set of observations.

If votes are fully congruent, then the responsiveness line slopes up; if legislators completely ignore citizen preferences, then the line is flat. The issue is whether – between these two polar cases – a larger responsiveness coefficient indicates more representation. One could, for example, estimate two regression lines, one for each group (say Democrats vs. Republicans), and compare the slope coefficients. Figure 1b gives a simple example in which greater responsiveness does not mean greater representation. There are two groups, G_1 and G_2, each with two observations. Intuitively, the G_2 observations capture what we think of as "representation" better than the G_1 observations (votes in G_1 are congruent for only one of two observations, while those in G_2 are congruent for both observations). However, consider the estimated regressions indicated by dashed lines between the two points in each group. The slope is larger for G_1 than G_2, so G_1 is more "responsive" than G_2.

Figure 1(a)

Figure 1(b)
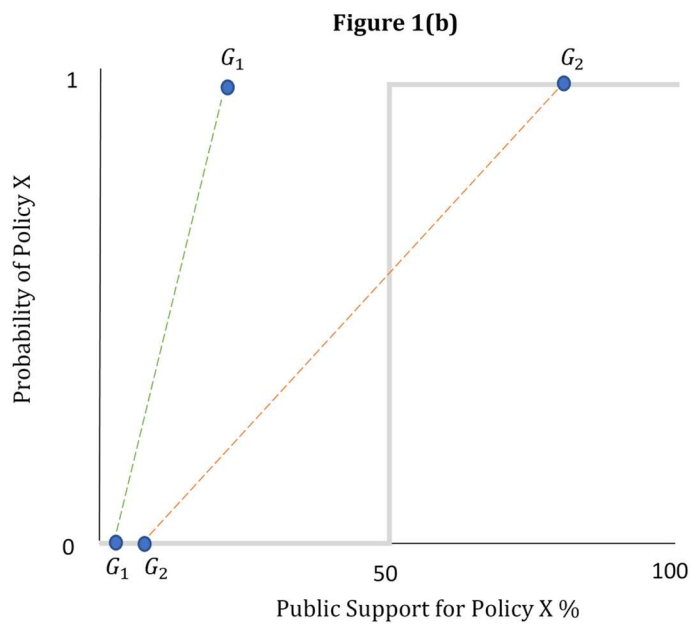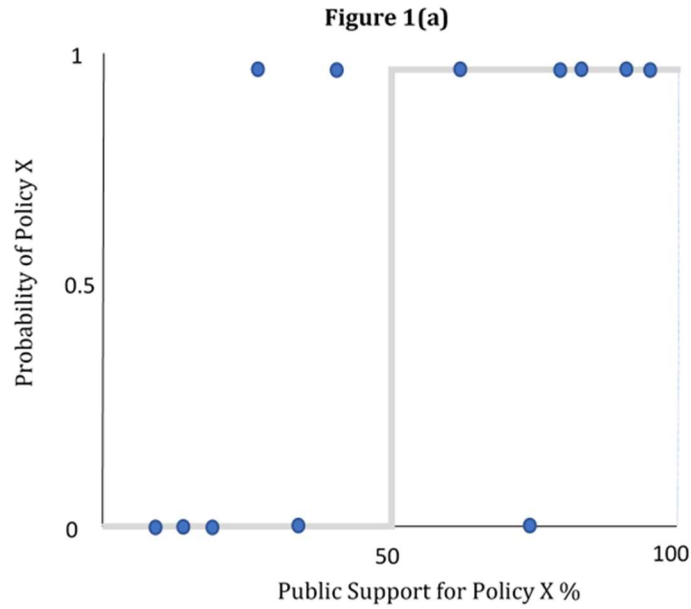
One can construct examples in which the responsiveness coefficient points in the correct direction, as well as examples in which the slopes are the same but congruence is different, and conversely. The general point is that there is no logical connection between β and the amount of representation, as we intuitively envision it.

Observation 2. *If $\beta_A > \beta_B$ then A could be more, less, or equally representative as B.*

This theoretical point is much broader than the simple example. First, there is nothing special about using linear estimates; the point would stand for more general functional forms. Second, there is nothing special about the policy being dichotomous; essentially the same arguments hold if policy is a continuous variable (see Matsusaka (2001) for an example). Third, as mentioned above, nothing hinges on defining congruence in terms of majority rule – the same argument applies to any definition of representation that hinges on citizen preferences.

The limitations of responsiveness can also be seen from a simple verbal example, motivated by Caughey and Warshaw (2022, p. 97): "In a perfectly responsive democracy, a given change in citizens' preferences (e.g., from a $10 minimum wage to an $11 one) would be followed by a policy change of the same magnitude (a $1 increase in the minimum wage)." This is true in a perfectly representative democracy, but the intuition does not carry through in a partially representative democracy. Suppose that the minimum wage is $11; initially citizens prefer a minimum wage $W^* = \$10$ and their preferred policy changes to $W^* = \$11$. It would actually be more representative, in terms of the definition above, to keep the minimum wage at $11 than to increase it to $12. Responsiveness in this case *reduces* the representativeness of the policy.

The same logic extends to comparisons. Suppose the initial minimum wage is $11; there are two group of citizens A and B that initially prefer $W_A^* = \$10$ and $W_B^* = \$8$; and each group experiences a +1 change in its preferences, that is, to $W_A^* = \$11$ and $W_B^* = \$9$. Now suppose the government of A does not respond and holds the minimum wage to $11, while the government of B increases it to $12. Here the government of A has been entirely "unresponsive" ($\beta_A = 0$) while the government of B has been perfectly "responsive" ($\beta_B = 1$) but the government of A is clearly better representing voter preferences than the government of B. If we were to assess the representativeness of the two groups based on responsiveness alone, we would wrongly celebrate B and condemn A.

Given that "responsiveness" lacks a consistent theoretical link to the underlying concept of representation, the natural question is how to justify its use as a metric for representation? One possible justification is empirical: even though there is no necessary theoretical link, it may be true generally that there is an empirical link: responsive legislators may be representative in practice. Having such a link would provide a

foundation for using responsiveness in research, but to the best of my knowledge there is no evidence on whether such a relation holds in practice. The rest of this paper examines whether such an empirical connection exists in a large sample of roll-call votes. Relying on the observation the congruence reliably captures representation, I examine if responsiveness is correlated with congruence.

Observation 3. *If, in practice, responsiveness captures representation, then we would observe empirically that groups with a high responsiveness would have a high congruence, and conversely.*

3. Data and Methods

My approach is to calculate responsiveness using (2), congruence using (3), and then assess how well responsiveness proxies for congruence. The challenge in executing such a comparison is the need for data on both legislator roll-call votes and the preferences of their constituents on the same laws. My research strategy is to focus on laws that were challenged by a referendum, for which we can observe roll call votes and citizen preferences in each district on exactly the same policy.

At present, 23 American states allow citizens to use the referendum process to challenge state laws approved by elected officials (i.e., passed by the legislature and signed by the governor). Implementation details differ, but in all cases if sponsors collect a predetermined number of signatures from fellow citizens, an election is held in which voters approve or repeal the law.[5] I use district-level referendum election returns to measure the preference of the majority (= median) of voters in each legislator's district.

To construct the sample, I began by identifying all state-level referendums during the period 2000-2022 using the Initiative and Referendum Institute database. From this list

---

[5] For example, in California, petitioners have 90 days after approval of a law to collect signatures from eligible voters equal to 5 percent of the number of votes cast in the previous gubernatorial election (546,651 signatures as of 2024). For institutional details, see Gerber (1999).

of ballot measures, I searched for district-level returns from each state's election division.[6]
The bill associated with each referendum was then identified, and roll call votes on that bill
were drawn from legislative records. The final sample after dropping abstentions contains
4,192 roll-call votes on 33 laws in 10 states.[7]

The laws are listed in Table 1, with the year of legislative approval indicated. The
referendums took place in Alaska, California, Maine, Maryland, Michigan, North Dakota,
Ohio, South Dakota, and Washington. These states represent a mix of urban and rural, and
include both "blue" and "red" states. The laws covered fiscal, political, and social issues, and
included hot-button topics of national interest such as same-sex marriage, as well as issues
of local interest such as Alaska's law allowing aerial hunting of wolves and North Dakota's
law allowing the state university to discontinue use of the "Fighting Sioux" nickname for its
mascot. The ideological orientation of the laws was also mixed, some proposing to move
policy in a liberal direction (e.g., allowing same-sex marriage or allowing the state to use
affirmative action/racial preferences) and others proposing a conservative change (e.g.,
allowing charter schools or limiting collective bargaining by public employees). Voters
approved 55 percent of the laws.

An alternative approach would be to measure district preferences by opinion
surveys instead of referendum election returns.[8] A significant limit of referendum data is
that the number of such elections is limited; opinion data are much more abundant, and
using modern MRP (multilevel regression and poststratification) methods can be extended
to jurisdictions that were not surveyed in the first place. A relative advantage of
referendum data is that they reveal voter preferences on exactly the law approved by the
legislature, while opinion surveys usually ask a law in general terms – and poll results can
vary dramatically with minor changes in question wording (Caughey and Warshaw, 2022,

---

[6] Referendums were excluded if a state did not report sufficiently disaggregated data, if a state changed its
district lines between the time of the roll call vote and the time of the referendum election, if individual roll-
call votes were unavailable, or if a referendum was abandoned by its sponsors after qualifying for the ballot.

[7] Matsusaka (forthcoming) provides more details and summary information about the dataset.

[8] Studies using opinion data include Krimmel et al. (2016), Lax et al. (2019), and Caughey and Warshaw
(2022). Portmann et al. (2012), Stadelmann et al. (2013, 2014), and Giger et al. (2020) use ballot measure
returns to calculate congruence.

Table 1. List of Laws			
State	Description	Outcome	Year
Alaska	Permits hunters to use airplanes to hunt wolves	Repealed	2000
California	Permits Pala tribe to operate video lottery terminals	Approved	1998
California	Allows third parties to sue insurance companies	Repealed	1999
California	Allows third parties to sue insurance companies (alt)	Repealed	1999
California	Requires companies to provide health care coverage	Repealed	2003
California	Authorizes gambling compact with Pechanga tribe	Approved	2007
California	Authorizes gambling compact with Morongo tribe	Approved	2007
California	Authorizes gambling compact with Sycuan tribe	Approved	2007
California	Authorizes gambling compact with Agua Caliente tribe	Approved	2007
California	Allows North Folk tribe casino in Central Valley	Repealed	2013
California	Bans plastic bags in grocery stores	Approved	2014
California	Ends use of cash bail	Repealed	2018
Colorado	Joins state in national popular vote compact	Approved	2019
Maine	Replaces health insurance claims tax with beverage tax	Repealed	2008
Maryland	Changes voting procedures	Approved	2006
Maryland	Allows illegal immigrants to pay in-state tuition rates	Approved	2011
Maryland	Congressional redistricting plan	Approved	2011
Maryland	Allows same-sex marriage	Approved	2012
Michigan	Allows hunting of mourning doves.	Repealed	2004
North Dakota	Discontinues "Fighting Sioux" college nickname	Approved	2011
North Dakota	Allows corporate dairy and swine farms	Repealed	2015
Ohio	Limits interest rate charged by payday lenders	Approved	2008
Ohio	Limits collective bargaining by public employees	Repealed	2011
South Dakota	Bans smoking in restaurants and bars	Approved	2009
South Dakota	Increases barriers for independent candidates for office	Repealed	2015
South Dakota	Reduces minimum wage for teenagers	Repealed	2015
Washington	Increases taxes for unemployment insurance	Repealed	2002
Washington	Allows charter schools	Repealed	2004
Washington	Prohibits insurers from denying certain claims	Approved	2007
Washington	Grants domestic partners same rights as married people	Approved	2009
Washington	Allows same-sex marriage	Approved	2012
Washington	Allows government to use racial preferences	Repealed	2019
Washington	Requires schools to provide sex education	Approved	2020

p. 17). Another advantage of referendum data is that referendum elections make law –
citizen opinions may be more informed when they have real consequences and have been
exposed to campaigns than when they give an off-the-cuff response to a pollster. Both
referendum data and opinion surveys suffer from having a nonrepresentative sample of
issues – the issues that come to a referendum or that pollsters ask about tend to be more

controversial than others. Having said this, the two approaches are complementary, and having evidence from both sources helps form a full picture.
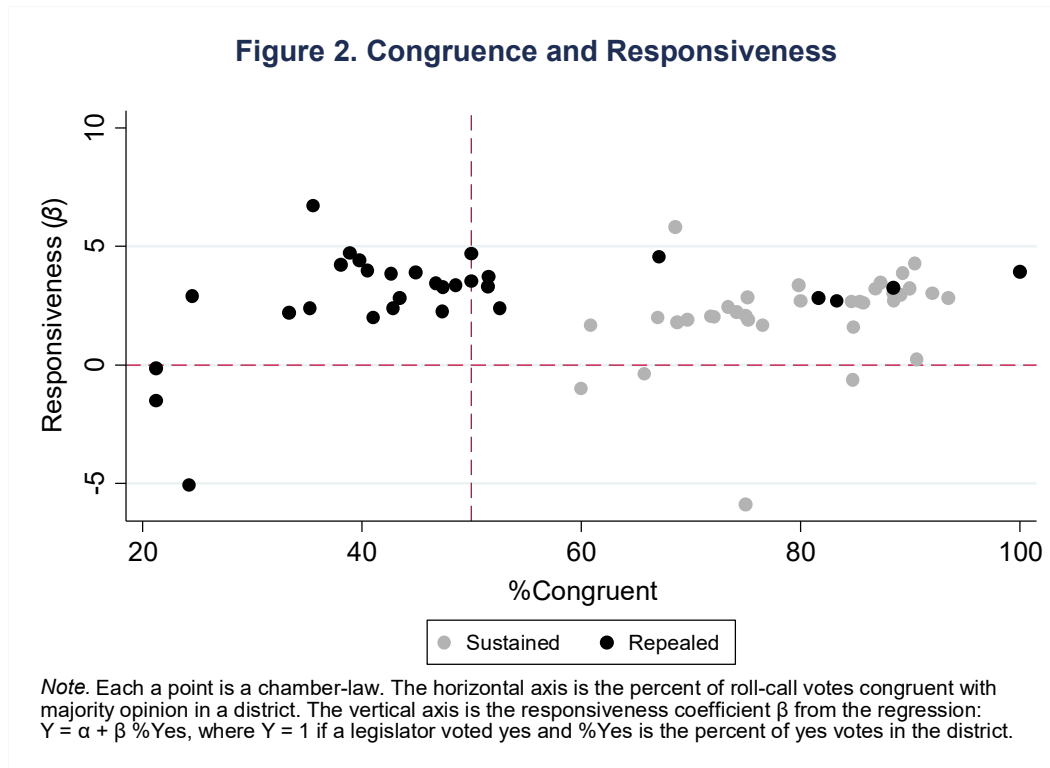
4. Findings

Congruence was 66 percent for the entire sample of 4,192 observations, meaning that two-thirds of roll-call votes were cast in accordance with majority opinion in the district. Since congruence would be 50 percent with random voting, anything over this threshold indicates some degree of representation. Responsiveness for the full sample was 1.08, meaning that a 1 percent increase in public support for a law was associated with a 1.08 percentage point greater chance that a legislator voted in support.

To investigate how well responsiveness tracks congruence, I begin by estimating responsiveness and congruence separately for each law and each chamber of each state, producing a total of 66 points of comparison. Figure 2 is a scatterplot of congruence against responsiveness. If responsiveness predicts congruence, we should see a positive relation between the two measures, and the responsiveness coefficient should be negative when congruence is less than 50 percent and positive when congruence exceeds 50 percent. Neither of these patterns is visible in the figure. A regression that uses β to predict congruence for these data has an R^2 of 0.006.[9] If a researcher was interested in which legislatures were more likely to vote in accordance with majority opinion in their districts, knowing the responsiveness coefficient would not help.

A few examples help illustrate the point: The chamber with the highest congruence (100 percent), California's senate in 2003, had a responsiveness of $\beta = 3.93$, while the chamber with one of the lowest congruences (38.9 percent), California's senate in the second insurance law of 1999, had a *higher* responsiveness coefficient of $\beta = 4.71$. The chamber with the highest responsiveness coefficient of $\beta = 6.71$, North Dakota's Senate in 2015, had a low congruence of 35.6 percent, while the chamber with the lowest responsiveness coefficient of $b = -5.88$, California's senate in 1998, had a higher congruence of 75.0 percent.

---

[9] The coefficient on responsiveness is a statistically insignificant $\beta = 0.83$ (standard error 1.34).

Figure 2. Congruence and Responsiveness

Note. Each a point is a chamber-law. The horizontal axis is the percent of roll-call votes congruent with majority opinion in a district. The vertical axis is the responsiveness coefficient β from the regression: Y = α + β %Yes, where Y = 1 if a legislator voted yes and %Yes is the percent of yes votes in the district.

We might expect the laws that voters approved to have been more representative than those they repealed.[10] Figure 2 shows the two groups in gray and black, respectively. Consistent with this intuition, all of the laws that voters sustained were more than 50 percent congruent, and most of the laws that they repealed were less than 50 percent congruent. In this sense, congruence "works" as a measure of representation. Responsiveness, on the other hand, does not. There is no reliable relation between a law's responsiveness coefficient and whether or not voters approved it.

Table 2 provides comparisons of congruence and responsiveness between several natural groups. The question is: to what extent do differences in responsiveness coefficients predict congruence differences between the groups? Congruence numbers are calculated by summing over all legislators in a given group. Responsiveness coefficients are calculated by estimating model (2) with chamber-law fixed effects using all legislators in the group. Fixed effects allow the propensity to support a law to vary by chamber and law.

---

[10] This is plausible but not necessary. If districts are gerrymandered, even a fully congruent legislature (that is, a legislature in which every member votes according to majority opinion in his or her district) can choose a policy that is not supported by a majority of voters statewide (Gilligan and Matsusaka, 2006).

Table 2. Congruence and Responsiveness by Outcome, Chamber, Term Limits, Year				
Panel A	Approved by Voters	Repealed	Difference	
Congruence (%)	79.6	48.8	30.8	[.001]
Responsiveness (β)	2.69 (0.12)	3.20 (0.18)	-0.51	[.019]
N	2,382	1,810		
Panel B	Lower House	Upper House	Difference	
Congruence (%)	66.3	66.4	0.1	[.992]
Responsiveness (β)	2.93 (0.13)	2.66 (0.15)	0.27	[.191]
N	2,922	1,270		
Panel C	Term Limits	No Term Limits	Difference	
Congruence (%)	61.4	71.3	9.9	[.057]
Responsiveness (β)	3.17 (0.14)	2.67 (0.13)	0.50	[.010]
N	2,119	2,073		
Panel D	Years 1998-2010	Years 2011-2020	Difference	
Congruence (%)	63.9	69.0	-5.2	[.330]
Responsiveness (β)	2.68 (0.21)	2.95 (0.09)	-0.28	[.236]
N	2,186	2,006		

Note. Congruence is calculated using all legislators in a group. Responsiveness is calculated from regression (2) with the addition of chamber-law fixed effects. Standard errors, clustered at the chamber-law level, are in round parentheses. Differences are calculated from regressions that pool all observations, with standard errors clustered at the chamber-law level; *p*-values are shown in square brackets.

Panel A compares approved and repealed laws. Congruence was 79.6 percent for approved laws and 48.8 percent for repealed laws, a difference of 30.8 percent that is statistically different from zero in a regression.[11] Responsiveness was positive for both groups, but higher for votes on the repealed than the approved laws. The two responsiveness coefficients are statistically different from each other at the 5 percent level.

---

[11] Here and below, comparisons of congruence or responsiveness between groups is done by estimating a single a regression pooling all observation with different coefficients for each group. Standard errors are clustered by chamber-law.

Responsiveness incorrectly implies that voting on the repealed laws was more representative in the sense used here.

Panel B distinguishes between the lower and upper chambers of the legislature. This is an interesting comparison because lower chambers are designed to be more closely connected to popular opinion than upper chambers by requiring members to stand for election more frequently and by having fewer constituents per district.[12] Despite the design intention, congruence was essentially identical between lower and upper chambers; the 0.1 percent difference is not statistically different from zero. Similarly, the responsiveness coefficients are similar and not statistically distinguishable. In this case, the responsiveness coefficients tell the same story as congruence – that representation was similar in the two chambers.

Panel C compares states that did and did not have term limits on legislators.[13] There are offsetting theoretical predictions about how term limits affect representation.[14] Proponents argue that term limits bring about the election of "citizen legislators" whose views are more consonant with their constituents, replacing career politicians, leading to superior representation. Skeptics argue that term-limited legislators have less incentive to follow constituent preferences when they no longer face re-election incentives. For the votes studied here, congruence was 9.9 percent lower in term-limit states than other states, with a p-value of 0.057. The responsiveness coefficients point in the opposite direction, with a lower β coefficient for term-limit than non-term-limit states, although they are not statistically distinguishable. Again, the responsiveness coefficients lead to an incorrect conclusion about representation.

Panel D compares representation across time. With the growth of polarization, one might wonder if representation has fallen over time. I calculate congruence and responsiveness for two periods, 1998-2010 and 2011-2020, that approximately divide the observations in half. Congruence increased 5.2 percentage points between the periods,

---

[12] All states in the sample have more members in the lower than upper chamber. Alaska, California, Michigan, Ohio, and Washington have longer terms for members of the upper than lower chamber.

[13] The term limit states were California, Colorado, Maine, Michigan, Ohio, and South Dakota.

[14] See Besley and Case (2003), Kousser (2008), Mooney (2009).

Table 3. Congruence and Responsiveness by Legislator Party, Gender, Vote Margins				
Panel A	Democrat	Republican	Difference	
Congruence (%)	71.4	60.3	11.1	[.257]
Responsiveness (β)	1.06 (0.24)	1.09 (0.25)	-0.55	[.263]
N	2,291	1,900		
Panel B	Female	Male	Difference	
Congruence (%)	67.9	65.7	2.2	[.234]
Responsiveness (β)	2.49 (0.14)	2.93 (0.11)	-0.10	[.457]
N	1,213	2,979		
Panel C	Roll-call margin < 10%	Roll-call margin > 10%	Difference	
Congruence (%)	68.8	65.7	3.1	[.662]
Responsiveness (β)	3.20 (0.11)	2.72 (0.13)	0.48	[.006]
N	872	3,320		
Panel D	District vote margin < 10%	District vote margin > 10%	Difference	
Congruence (%)	60.4	67.7	-7.3	[.016]
Responsiveness (β)	2.80 (0.31)	2.82 (0.11)	0.18	[.251]
N	762	3,430		

Note. Congruence is calculated using all legislators in a group. Responsiveness is calculated from regression (2) with the addition of chamber-law fixed effects. Standard errors, clustered at the chamber-law level, are in round parentheses. Differences are calculated from regressions that pool all observations, with standard errors clustered at the chamber-law level; *p*-values are shown in square brackets.

although the difference is not statistically significant. The responsiveness coefficients point in the same direction: larger in the second period but statistically insignificant. In this case, responsiveness tracks congruence.

Table 3 reports additional comparison, including some in which the groups are defined based on legislator-specific characteristics. Panel A compares Democratic (plus Green) vs. Republican legislators. While there is no theoretical reason to expect one party to be more representative than the other, some evidence suggests there may be partisan differences in how accurately legislators perceive constituent preferences (Broockman and Skovron 2018). Democratic legislators voted congruent with their constituents 11.1 percent more often than Republican legislators, a sizeable but not statistically significant

difference. The responsiveness coefficients, calculating by pooling all member of the group and including chamber-law fixed effects, are similar for the two parties. In a regression pooling all observations, with a different coefficient on congruence in the two groups, the responsiveness difference is tiny and statistically insignificant. The responsiveness coefficients thus conform to the pattern shown by congruence.

Panel B compares congruence of female and male legislators. A growing literature investigates whether female or male legislators behave differently. Anzia and Berry (2011) found that female members of Congress delivered more federal spending to their districts than male members; Stadelmann et al. (2014) reported that female legislators voted more congruently than male legislators in Switzerland; while Ferreira and Gyourko (2014) found no difference in the policy choices of female and male mayors. For the votes in my sample, female legislators were 2.2 percent more likely to cast a congruent vote than male legislators, a small and statistically insignificant difference. The responsiveness coefficient is also larger for female than male legislators, but not statistically different from zero.

Panel C distinguishes legislators according to whether the roll-call vote was close or lopsided, defined by a vote margin less or greater than 10 percent (the patterns are similar with other cutoffs). Snyder and Groseclose (2000) argue that legislative leaders are more likely to put pressure on regular members when a roll-call vote is close than when it is lopsided because it's costly to persuade a legislator to vote against constituent interests, and there is no need to buy votes when the outcome is a foregone conclusion. Contrary to this argument, congruence was essentially the same for close and lopsided roll-call voters.[15] The responsiveness coefficient is larger for close than lopsided votes, and statistically different from zero in the pooled regression, suggesting that representation was higher for close votes, contrary to theory and the congruence evidence.

Finally, Panel D compares legislators that represented competitive versus noncompetitive districts. Some activists believe that competition brings about better representation. I classify a district as competitive if the margin of victory between the top

---

[15] Although this does not support the theory, Matsusaka (forthcoming) offers what is arguably a more parsimonious test, finding that congruence was higher on close votes if the sample is restricted to legislators who were ideologically opposed to their constituents on an issue.

two candidates in the previous election was less than 10 percent (the patterns are similar with other cutoffs). Legislators representing noncompetitive districts were actually 7.4 percent more congruent than those representing competitive districts, a difference significant at the 5 percent level. The responsiveness coefficients reveal no meaningful differences. Responsiveness thus fails to reveal an apparent difference in representation.[16]

5. Discussion

This study aims to advance the study of representation by probing – both theoretically and empirically – two leading methods used to measure representation, "congruence" and "responsiveness." The focus is not on the substantive basis of representation, but on the methods we use to measure it. Although sometimes viewed as different ways of capturing the same underlying concept, only congruence is theoretically justified as a measure of representation. Responsiveness – the correlation between roll-call votes and constituent preferences – has been justified only with brief verbal intuitions, and it is easy to construct examples for which those intuitions fail in the sense that group A has a higher responsiveness than group B even though group A is clearly *less* representative than group B. In some sense this merely restates a conclusion reached by Achen (1978, p. 475) almost 50 years ago ("correlations are virtually uninterpretable within the framework of democratic theory") that has not been widely appreciated.

Despite the lack of a theoretical justification, it is possible that responsiveness coefficients might work in practice. Unfortunately, this does not appear to be the case. This paper's core contribution is to evaluate empirically – I believe for the first time – if responsiveness in fact tracks representation for a large sample of roll-call votes. As a proxy for representation, I use congruence since its theoretical justification is secure. I find that responsiveness is not a reliable predictor of congruence; sometimes it works but sometimes it does not. Perhaps the most dramatic failure in my sample comes from comparing laws that voters approved with laws they repealed: congruence was much

---

[16] Matsusaka (forthcoming) shows that legislators from noncompetitive districts cast more congruent votes because they were more likely to be ideologically aligned with their constituents.

higher on approved than repealed laws, as one would expect, but the estimated responsiveness coefficients were lower for approved than repealed laws.

The usual caveat applies: this is only one study, and further evidence along the same lines is needed before drawing strong conclusions. The particular votes and laws examined here may or may not be representative of other contexts. But if congruence and responsiveness were tapping the same concept, they should exhibit an empirical connection for any set of laws. The failure of responsiveness in one context raises difficulties for its application in all contexts.

To the extent that the findings here prove to be general, several implications follow. First, congruence should be preferred over responsiveness in studies of representation. Congruence clearly conforms to our intuitions about representation – there are no theoretical issues with the concept. The challenge with using congruence has been the lack of data on constituent preferences. Fortunately, with the advent of new data sources and methods, especially MRP, we are now in a position to calculate congruence over a much wider range of issues. It is encouraging to see a burgeoning literature do just that.[17]

Nevertheless, there may be circumstances where the data necessary to calculate congruence are unavailable, but calculating responsiveness is feasible. A second implication is that studies using responsiveness should be sensitive to its potential limitations. It is not unusual for an empirical study to begin with only a brief verbal justification for using responsiveness before moving to the data. Simply asserting that responsiveness captures representation seems inadequate given what we now know. Going forward, it is desirable for studies using responsiveness to grapple with the known limitations, and offer more compelling justifications for its use. Ideally, responsiveness estimates would be coupled with congruence estimates as in Caughey and Warshaw (2022). As for the existing literature that uses responsiveness to draw conclusions about representation, some caution seems advisable when considering those claims; and it would be informative for researchers to revisit their claims using congruence as the measure. A

---

[17] Examples: Matsusaka (2010), Lax and Phillips (2012), Portmann et al., (2012), Brunner et al. (2013), Stadelmann et al. (2013, 2014), Krimmel et al., (2016), Kousser et al. (2018), Becher and Menendez Gonzalez (2019), Lax et al. (2019), Simonovitz et al. (2019), Matsusaka (2020, forthcoming).

good example: Bashir (2015), Enns (2015), and Branham et al. (2017) revisited the well-known finding of Gilens and Page (2014) that policy was more responsiveness to rich than poor Americans, and found that congruence was essentially the same if not reversed for the two groups.

Third, more research is needed on the theoretical underpinnings of responsiveness as a measure of representation. Responsiveness appeals to our sense that representation should consist of a correlation across time or space between roll-call votes and constituent preferences. Since we have specific examples where responsiveness does not "work," it would be useful to know under what conditions responsiveness does and does not accurately capture representation, especially a set of conditions under which responsiveness is likely to be a valid proxy. Ideally, those conditions would be amenable to empirical implementation, giving researchers a menu of auxiliary of tests to determine if it is appropriate to use responsiveness in their context.

References

Achen, Christopher H., "Measuring Representation: Perils of the Correlation Coefficient," *American Journal of Political Science*, November 1977, Vol. 21(4), 805-815.

Achen, Christopher H., "Measuring Representation," *American Journal of Political Science*, August 1978, Vol. 22(3), 475-510.

Anzia, Sarah F. and Christopher R. Berry, "The Jackie (and Jill) Robinson Effect: Why Do Congresswomen Outperform Congressmen?," *American Journal of Political Science*, July 2011, Vol. 55(3), 478-493.

Barro, Robert J., "The Control of Politicians: An Economic Model," *Public Choice*, Spring 1973, 14(1): 19-42.

Bashir, Omar S, "Testing Inferences about American Politics: A Review of the `Oligarchy' Result, *Research and Politics*, October-December 2015, 1-7.

Becher, Michael and Irene Mendendez Gonzalez, "Electoral Reform and Trade-offs in Representation," *American Political Science Review*, 2019, Vol. 113(3), 694-709.

Branham, J. Alexander, Stuart N. Soroka, and Christopher Wlezien, "When Do the Rich Win?," *Political Science Quarterly*, 2017, Vol. 132(1), 43-62.

Broockman, David E. and Christopher Skovron, "Bias in Perception of Public Opinion among Political Elites," *American Political Science Review*, August 2018, Vol. 112(3), 542-563.

Brunner, Eric, Stephen L. Ross, and Ebonya Washington, "Does Less Income Mean Less Representation?," *American Economic Journal: Economic Policy*, May 2013, Vol. 5(2), 53-76.

Caughey, Devin and Christopher Warshaw, "Policy Preferences and Policy Change: Dynamic Responsiveness in the American States," *American Political Science Review*, 2018, 112(2), 249-266

Caughey, Devin and Christopher Warshaw, *Dynamic Democracy: Public Opinion, Elections, and Policymaking in the American States*, Chicago, IL: University of Chicago Press, 2022.

Enns, Peter K., "Relative Policy Support and Coincidental Representation," Perspectives on Politics, December 2015, Vol. 13(4), 1053-1064.

Erikson, Robert S., Gerald C. Wright, and J. P. McIver, *Statehouse Democracy: Public Opinion and Policy in the American States*, Cambridge, UK: Cambridge University Press, 1993.

Ferejohn, John, "Incumbent Performance and Electoral Control," *Public Choice*, 1986, 50: 5-25.

Ferreira, Fernando and Joseph Gyourko, "Does Gender Matter for Political Leadership? The Case of U.S. Mayors," *Journal of Public Economics*, 2014, Vol. 112, 24-39.

Gerber, Elisabeth R., *The Populist Paradox: Interest Group Influence and the Promise of Direct Legislation*, Princeton, NJ: Princeton University Press, 1999.

Giger, Nathalie, Heike Klüver, and Christopher Witko, "Electoral Vulnerability, Party Affiliation, and Dyadic Constituency Responsiveness in U.S. Legislatures," *American Politics Research*, 2020, Vol. 48(4), 484-491.

Gilens, Martin and Benjamin I. Page, "Testing Theories of American Politics: Elites, Interest Groups, and Average Citizens," *Perspectives on Politics*, 2014, Vol. 12, 564-581.

Gilligan, Thomas W. and John G. Matsusaka, "Public Choice Principles of Redistricting,"
  Public Choice, December 2006, Vol. 129(3-4), 381-398.

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

Kousser, Thad, "Term Limits and State Legislatures," in *Democracy in the States:
  Experiments in Election Reform*, edited by Caroline Tolbert, Todd Donovan, and
  Bruce Cain, Washington D.C.: Brookings Institution Press, 2008.

Kousser, Thad, Justin Phillips, and Boris Shor, "Reform and Representation: A New Method
  Applied to Recent Electoral Changes," *Political Science Research and Methods*,
  October 2018, Vol. 6(4), 1-19.

Krimmel, Katherine, Jeffrey R. Lax, and Justin H. Phillips, "Gay Rights in Congress: Public
  Opinion and (Mis)representation," *Public Opinion Quarterly*, December 2016, Vol.
  80(1-4), 888-913.

Lax, Jeffrey R. and Justin H. Phillips, "The Democratic Deficit in the States," *American Journal
  of Political Science*, January 2012, Vol. 56(1), 148-166.

Lax, Jeffrey R, Justin H. Phillips, and Adam Zelizer, "The Party or the Purse? Unequal
  Representation in the US Senate," *American Political Science Review*, 2019, Vol.
  113(4), 917-940.

Matsusaka, John G., "Problems with a Methodology Used to Test the Responsiveness of
  Policy to Public Opinion in Initiative States," *Journal of Politics*, November 2001, Vol.
  63, 1250-1256.

Matsusaka, John G., "Popular Control of Public Policy: A Quantitative Approach," *Quarterly
  Journal of Political Science*, 2010, Vol. 5, 133-167.

Matsusaka, John G., *Let the People Rule: How Direct Democracy Can Meet the Populist Challenge*, Princeton, NJ: Princeton University Press, 2020.

Matsusaka, John G., "When Do Legislators Represent Their Constituents? Evidence from Roll-Call and Referendum Votes," *Journal of Politics*, forthcoming.

Mooney, Christopher Z., "Term Limits as a Boon to Legislative Scholarship: A Review," *State Politics and Policy Quarterly*, Summer 2009, Vol. 9(2), 204-228.

Pitkin, Hanna Fenichel, *The Concept of Representation*, Berkeley, CA: University of California Press, 1967.

Romer, Thomas and Howard Rosenthal, "The Elusive Median Voter," *Journal of Public Economics*, October 1979, Vol. 12, 143-170.

Simonovits, Gabor, Andrew M. Guess, and Jonathan Nagler, "Responsiveness without Representation: Evidence from Minimum Wage Laws in the U.S. States," *American Journal of Political Science*, April 2019, Vol. 63(2), 401-410.

Stadelmann, David, Marco Portmann, and Reiner Eichenberger, "Quantifying Parliamentary Representation of Constituents' Preferences with Quasi-Experimental Data," *Journal of Comparative Economics*, 2013, Vol. 14, 170-180.

Stadelmann, David, Marco Portmann, and Reiner Eichenberger, "Politicians and Preferences of the Voter Majority: Does Gender Matter?," *Economics and Politics*, November 2014, Vol. 26(3), 355-379.

Tausanovitch, Chris and Christopher Warshaw, "Representation in Municipal Government," *American Political Science Review*, August 2014, Vol. 108(3), 605-641.