4

Measuring Political Inequality[*]

Larry M. Bartels

Democracy has something to do with equality – but what, exactly? How should we gauge the extent of inequality in democratic political systems? What sorts of inequality are objectionable from the standpoint of democratic theory and why?

In an influential essay on "Measuring Representation," Achen (1977: 806) argued that "The central difficulty is not statistical, but conceptual. Rarely is a measure of representativeness related to the ideas of liberal democratic theory – for example, citizen equality and popular sovereignty. Instead, measures have been plucked from the statistical shelf and employed without much theoretical interpretation." More than forty years later, much the same could be said of the scholarly literature on political inequality. Scholars purporting to measure inequality deploy a variety of very different analyses, perhaps justified with a sentence or two gesturing to democratic theory. They often employ similar terms – "representation," "responsiveness," "congruence," "alignment," "association," "influence" – to describe different analyses and different terms to describe similar analyses. As a result, what appear to be substantive disagreements are often instances of scholars simply talking past each other, not noticing or not caring that they are talking about different things.

This chapter provides a conceptual and methodological roadmap of research on political inequality, with particular emphasis on the grounding of empirical analyses in "the ideas of liberal democratic theory." Like all roadmaps, mine is subjective, with some routes emphasized and others portrayed as backroads or even dead ends. However, my aim is not to resolve normative or empirical disagreements in the field – merely to make the disagreements more productive by clarifying what they are about.

Political inequality has been a subject of scientific study since the time of Aristotle, who classified regimes based on the relationship between political power and economic wealth. In the United States, studies of unequal political power – perhaps most famously, Dahl's (1961) *Who Governs? Democracy and Power in an American City* – were a hallmark of the mid-twentieth-century "behavioral revolution" in political science. However, the pluralist research program embodied in this and other studies of "who actually governs" bogged down in methodological and political controversies, and analyses of inequality increasingly came to focus on narrower but more tractable issues, as with the monumental studies of political participation published by Verba and colleagues over a span of forty years (Schlozman, Verba, and Brady 2012; Verba and Nie 1972; Verba, Schlozman, and Brady 1995).

In the twenty-first century, political scientists have once again aspired to gauge political inequality directly – this time, with the precision of systematic quantitative analysis. The roots of this work lie in two distinct threads of research on political representation: one relating the policy choices of individual elected officials to the preferences of their constituents as measured by survey data, and the other relating policy outcomes to aggregate public opinion across issues or over time.[1] In each case, the key analytical innovation was quite simple: to relate policy choices or outcomes to the distinct preferences of separate subgroups of citizens rather than to the preferences of the public as a whole.

Given this intellectual lineage, contemporary studies of political inequality have inherited much of the conceptual framework – and attendant complexities and confusions – of scholarship on political representation, while adding further complexities and confusions stemming from the application of this framework to a new set of questions. My aim here is to survey the most significant complexities and confusions of both sorts.

CONGRUENCE: SATISFYING PREFERENCES

Perhaps the most straightforward way to gauge the relationship between citizens and elected officials is by assessing the extent of *congruence* between citizens' preferences and policymakers' actions. In her seminal theoretical account of political representation, Pitkin (1967: 163–164) suggested that political leaders "must not be found persistently at odds with the wishes of the represented without good reason":

What the representative must do is act in his constituents' interests, but this implies that he must not normally come into conflict with their will when they have an express

---

[1] My own research on unequal responsiveness in Congress (Bartels 2016: Ch. 8) was grounded in a voluminous scholarly literature elaborating upon the pioneering work of Miller and Stokes (1963) on congressional representation. Gilens (2012: xiii) cited the influence of Monroe (1979), "the first to assess democratic representation by relating public preferences to government policy outcomes across large numbers of issues."

will…. Thus, when a representative finds himself in conflict with his constituents' wishes, this fact must give him pause. It calls for a consideration of the reasons for the discrepancy; it may call for a reconsideration of his own views.

Political theorists sometimes castigate empirical researchers – especially those who do "'large-N,' statistical work" – for adopting a "simplistic normative model of democracy whereby democratic majorities are to get whatever they want, on every issue, and in short order" (Sabl 2015: 345–346). I think a fairer characterization would be that most empirical researchers view the relationship between citizens' preferences and policy outcomes in much the same spirit as Pitkin. Consider, for example, the nuanced statement framing the most influential recent empirical analysis of disparities in representation (Gilens 2012: 47–48):

The quality of democratic governance in any society must be judged on a range of considerations. Are elections free and fair? Do citizens have access to the information necessary to evaluate their political leaders and competing candidates? Do government agencies perform their duties in a competent and unbiased manner? In this book I concern myself with only one aspect of democratic governance—the extent to which government policy reflects the preferences of the governed…. In documenting the ways in which policy fails to reflect (or reflect equally) the preferences of the public, I do not mean to imply that a perfect (or perfectly equal) responsiveness to the public is best.

There are good reasons to want government policy to deviate at times from the preferences of the majority: minority rights are important too, and majorities are sometimes shortsighted or misguided in ways that policymakers must try to recognize and resist…. Particular segments of the public may hold preferences on particular issues that are harmful to the community, violate important democratic values, or are misinformed and detrimental to the interests of those citizens themselves.

From this perspective, as in Pitkin's account, a pattern of significant discrepancies between citizens' preferences and policy outcomes "calls for a consideration of the reasons." The bases and coherence of citizens' preferences are amenable to empirical research and indeed have generated voluminous analysis and debate. Principles of justice and their application have mostly been treated by empirical researchers as topics beyond their remit, suitable for normative rather than empirical analysis.

Assessments of congruence evaluate representatives as "delegates" rather than "trustees," to employ a venerable theoretical distinction. Rehfeld (2009: 219) suggested that "Empirical scholars may favor delegate views of representation because they are easier to measure: one need only compare roll-call votes of representatives with public opinion surveys, or election outcomes with votes cast, to evaluate whether 'good' representation in this sense is achieved." While "empirical scholars" of representation may chafe at the phrase "one need only," there *is* an appealing conceptual simplicity to the notion that policy outcomes should, at least presumptively, correspond with public preferences. Alas, that conceptual simplicity breaks down rather quickly in practice.

One vexing set of problems turns on the measurement of citizens' preferences. Even when those preferences are not "incoherent" in a common-language sense, they may be subject to vagaries that complicate the task of assessing the correspondence between preferences and policies. Opinion surveys may frame policy issues in ways the call to respondents' minds some relevant considerations rather than others. For example, Americans have much more negative views regarding government spending on "welfare" than on "assistance to the poor." Many more would "not allow" a communist to make a speech than would "forbid" him from doing so. In instances like these, it seems hard to say exactly what the preferences are that representatives should be weighing (Bartels 2003).

Even if citizens' preferences are clearly captured by surveys or other data, assessing congruence requires us to decide whether the behavior of policymakers is consistent with those preferences. When policy choices are framed in dichotomous terms, congruence with any given citizen's preference may be thought of as an all-or-nothing matter. The citizen either favors or opposes adding a prescription drug benefit to a government health program, and policymakers do or don't comply. In many cases, this is straightforward enough; but sometimes assessing congruence may be a difficult matter of judgment. Is *any* prescription drug benefit enough to count?[2]

In other cases, policy outcomes may be arrayed along a continuum, making it natural to think of congruence as a measure of the "distance" between any citizen's preferred policy and the one her government adopts. Spending preferences are often portrayed in this way, since the corresponding policy outcomes are conveniently quantifiable. However, this formalization, too, may sometimes do considerable violence to reality when, for example, a citizen who wants her government to spend more on "healthcare" sees the money go to insurers and pharmaceutical companies rather than to clinics and nursing homes.

Even greater complexities arise in comparing the positions of citizens on general ideological scales with the positions adopted by or attributed to political elites. Citizens' understanding of ideological term is often shallow or confused (Converse 1964; Kinder and Kalmoe 2017). Even when they are splendidly well informed, it requires a good deal of optimism to assume that one person's "7" on a zero-to-ten "left-right" scale means the same thing as another's, or as a member of parliament's, or as a country expert's assessment of a party's

---

[2] Gilens (2012: 63) reported that coders agreed whether a proposed policy change had occurred 91 percent of the time (after excluding some partial change codes), but he did not discuss the nature of disagreements or how they were resolved. Bartels (2012) examined some of Gilens' specific cases of responsive policymaking, concluding that "it is seldom straightforward to classify policies as responsive or unresponsive to public preferences" and that, as a result, "responsiveness is a partial and often problematic standard for assessing the role of citizens' preferences in democratic policymaking."

position on the same scale. This is especially true in times and places when the meaning of ideology is contested or changing due to the emergence of new political issues and cleavages.[3]

Regardless of how policy positions are measured, the notion of congruence seems to require that they be measured identically for citizens and policymakers, or somehow reconciled, in order to allow for comparison between them. In practice, analysts must often make do with imperfect comparisons, relying on assumptions to overcome the limitations of available data. In his work revisiting Miller and Stokes's classic study of congressional representation, Achen (1978: 481, 484–485) acknowledged "some question about comparability" between opinion scales constructed from separate surveys of constituents and representatives. "Although the topics covered were essentially identical," he noted, "the congressional questionnaire was more specific, making reference to specific programs and proposals in some cases." Nonetheless, "For present purposes, one has little choice but to inspect the distribution of opinion on the scales among both Congressmen and constituents, and if no anomalies appear (none do), to follow Miller in standardizing the two scales to the same range and treating them as comparable."

In an ambitious cross-national study of congruence, Lupu and Warner (2022a: 279) applied a similar strategy on a much broader scale. They compiled data on the preferences of citizens and political elites in 565 country years from a wide variety of surveys employing a variety of scales. "To make these responses comparable," they reported, "we rescale them to range from –1 to 1." With this sort of wholesale normalizing, it seems very hard to know whether any resulting pair of citizens' and elites' responses is indeed "comparable," and thus very hard to gauge the extent of congruence or incongruence between them. Alas, concessions of this sort are common, given the scarcity of directly comparable measures of citizens' and policymakers' preferences.[4]

Even in cases where directly comparable measures of mass and elite preferences are available, difficult conceptual issues sometimes arise in comparing them. In legislative systems with single-member districts, we may be

---

[3] Powell (2019) provided detailed analyses and discussion of ideological congruence in parliamentary democracies. Brady (1985) explored the "perils" involved in statistical analysis of "interpersonally incomparable" survey data. Zechmeister (2006) documented substantial variation in the meaning of "left" and "right" among citizens in Mexico and Argentina, which she attributed to different national contexts, "elite packaging," and levels of political sophistication.

[4] Lupu and Warner added, "our analyses control for the scale used in each mass and elite survey and for the differences between the scales provided to elite and mass respondents in each country-year"; but there is no reason to expect measurement error in congruence introduced by incompatible scales to be eliminated, or even mitigated, by including fixed effects for scale formats. Nor is it necessarily the case that biases in measured congruence for distinct income groups will be subject to similar errors (for example, on issues where low-income citizens are generally to the "left" and high-income citizens are generally to the "right" of legislators).

interested in the correspondence between each individual representative's policy choices and the preferences of her own constituents, but the extent of *dyadic* representation sheds little light on the correspondence between citizens' preferences and overall policy outcomes (Weissberg 1978). In electoral systems without single-member districts, scholars have typically compared the preferences of rank-and-file supporters of each party with the preferences of the party's parliamentarians, as in Esaiasson and Holmberg's (1996) remarkably detailed study comparing the views of citizens and members of parliament in Sweden. But here, too, the relationship between party representation and policy outcomes may be complex and variable, depending on legislative institutions (the distribution of agenda-setting rights and resources), party cohesion, and the role of the president or prime minister, among other factors.

Golder and Stramski (2010: 95) distinguished between "absolute citizen congruence," measured by the average absolute distance between the preferences of citizens and those of a single representative, government, or policy outcome, and "many-to-many congruence" based on comparing overall *distributions* of opinion among citizens and legislators. They motivated attention to the latter, in part, by referring to "the importance of having a representative body whose preferences accurately correspond to those of the nation as a whole." However, they noted that "many-to-many congruence" between citizens and legislators is neither necessary nor sufficient to produce congruence between citizens' preferences and policy outcomes. A legislature that is, collectively, splendidly representative of the distribution of public opinion may nonetheless adopt policies that fail to comport with most citizens' preferences – for example, because a governing party or coalition representing one set of views dominates the policymaking process. Thus, it is crucial to distinguish, as Lupu and Warner (2022a: 277) put it, between "congruence or opinion representation – the process of generating a body of representatives that reflects the preferences of the electorate" and "the process by which these representatives generate policies that reflect citizens' preferences."

Even if congruence with majority preferences was a foolproof benchmark for assessing representation, additional conceptual difficulties would arise in adapting it to serve as a benchmark for assessing political inequality. A representative (or, more broadly, a political system) reflecting the preferences of majorities will fail to reflect the preferences of minorities. Thus, individuals who persistently find themselves in the minority will have their preferences satisfied less often than those who are generally in the majority. Some observers may consider this a justifiable form of political inequality because it is produced by the mechanism of majority rule, a familiar feature of democratic political systems, and one with a variety of desirable properties. As is often the case in discussions of inequality, a result that is splendidly egalitarian from one perspective (everyone's preferences count equally in gauging the will of the

majority) is plainly unequal and arguably invidious from a different perspective (some people routinely get their way and others do not).[5]

There is also a more prosaic arithmetic problem with attempts to measure differential congruence using aggregated tabulations of group preferences. The fact that policy outcomes are closer to the *average* preference of Group A than of Group B does not necessarily imply that congruence is greater for the *individuals* in Group A than for those in Group B, even on average. In the terminology proposed by Achen (1978: 481–488), congruence depends not only on the "centrism" of policy outcomes relative to a group's average preference, but also on the variance of those preferences. There is little reason to think that "centrism" (relative to the average preferences of a group) is an intrinsic good when the notional "group" is merely a convenient analytical fiction. Thus, in the context of assessing congruence, it seems very hard to attach any real significance to tabulations involving average group preferences.[6]

EQUAL INFLUENCE OVER POLICY

So far, I have surveyed a variety of complications involved in measuring inequalities in congruence between the preferences of citizens and the attitudes or choices of policymakers. But I have not addressed what should be a logically prior question – why care about congruence?

The most obvious answer is that we want our political system to give us what we want. But do we? As we have already seen, Pitkin (1967: 163–164, emphasis added) argued that "What the representative must do is act in his constituents' *interests*." Finding himself "in conflict with his constituents' *wishes*" is not in itself a dereliction of his duty as a representative, though it might "call for a reconsideration of his own views" if constituents' wishes are "normally" a good guide to discerning their interests.[7]

If our wishes are only relevant as indicators of our interests, then preference satisfaction itself is not an intrinsic good from the standpoint of democratic theory. Thus, a political philosopher (Kolodny 2023: 300) considered but rejected the view that "Each of us has a correspondence interest in the satisfaction of his or her policy preferences as such." But in that case, tabulations of

---

[5] Alternative procedures create analogous difficulties. For example, if policy choices are made by citizens chosen at random, everyone's preferences will be equally influential *ex ante*, but those whose views are popular among their fellow citizens will still get their way more often than those whose views are unpopular.

[6] The mean squared distance between a policy outcome and the preferences of group members can be decomposed into two terms – (1) the squared distance between the policy outcome and the average preference of group members and (2) the variance of preferences. Even if the first term is smaller for Group A than for Group B, their sum may be larger for Group A if the variance of preferences in Group A is sufficiently larger than in Group B.

[7] On the relationship between preferences and interests – and the daunting normative and analytical complexities involved in measuring political interests systematically – see Bartels (1990).

inequality in congruence, without careful additional consideration of the correspondence between preferences and interests, are of little normative relevance.

What justice demands, Kolodny (2023: 323, 320, 87–145) argued, is not equality of preference *satisfaction* but equality of *influence* over policy outcomes. "Equal Influence," he wrote, "is satisfied insofar as any individual who is subject to superior untampered power and authority [that is, to the power of the state] has as much opportunity as any other individual for informed, autonomous influence over decisions about how that power and authority are to be exercised." Equal influence is intrinsically good, Kolodny reasoned, because "If someone is to have influence, then everyone should have equal influence, lest the inequality convey, or be taken to convey, something disparaging about those with less." In the context of his broader "philosophy of social hierarchy," a demand for equal influence is an instance of "claims against inferiority." Disparities in influence that are correlated with economic and social inequalities seem especially problematic if our concern is about real or perceived "social hierarchy."

Kolodny's emphasis on equal influence as the foundation of just collective decision-making resonates with Dahl's analysis of political equality. Dahl (2006: 4, 9) grounded his normative argument for democracy in the "assumption" that "the moral judgment that all human beings are of equal intrinsic worth, that no person is intrinsically superior to another, and that the good or interests of each person must be given equal consideration" in the determination of public policy. The phrase "equal consideration" seems to imply something like equal weight in the determination of policy, rather than equal probability of winning or equal satisfaction with policy outcome – in the language proposed here, equal *influence* rather than equal *congruence*. That interpretation is bolstered by the fact that Dahl went on to list a series of necessary procedural conditions for "an ideal democracy." The most relevant of these, "Equality in voting," stipulated that "When the moment arrives at which the decision will finally be made, every member must have an equal and effective opportunity to vote, and all votes must be counted as equal." Here, too, the emphasis is on *procedures* rather than *outcomes*; once all votes are counted as equal, presumably some will win and some will lose.

Of course, most policy decisions in real democracies are made not directly by popular vote, but by elected or appointed officials. The closest Dahl (2006: 9) came to addressing this fact was to stipulate that "policies of the association would always be open to change by the demos, if its members chose to do so." But, even leaving aside the vagueness of how that would work, what about all those policies the demos does not choose to decide directly? For those cases, we need a conception of "equal consideration" that does not hinge on the mechanics of casting and counting votes.

The conception of "equal consideration" or "equal influence" animating contemporary empirical research on political inequality has its roots in the same "behavioral revolution" that inspired Dahl's study of *Who Governs?* a

half-century earlier. Dahl (1957), Harsanyi (1962), Simon (1953), and other prominent mid-century social scientists contributed to a substantial theoretical literature focusing on the concepts of power and influence. The most important upshot of that work, codified in Nagel's (1975) book, *The Descriptive Analysis of Power*, is that power entails a positive causal relationship between an actor's preferences and outcomes. Nagel proposed using statistical models to represent relationships of this sort. In the context of collective decision-making, we might model a policy outcome as a function of the preferences of various relevant political actors, including citizens, parties, interest groups, and elected or unelected government officials.[8] Contemporary studies of political inequality employing regression analyses relating policy outcomes to citizens' preferences instantiate exactly this approach – or attempt to.

As with attempts to measure *congruence* between opinions and policy, attempts to measure *influence* may be more or less cogent. But the challenges to persuasive measurement are different in kind. One significant advantage of focusing on influence rather than congruence is that the opinions of citizens and the choices of policymakers need not be measured on commensurate scales, as long as the opinions being measured appropriately reflect citizens' relevant policy preferences. Analyses of responsiveness in the United States have employed survey data on ideological self-placements, views on specific issues, and even election returns as measures of citizens' preferences. In the comparative literature, levels of social spending have been related to broad support for the government's role in providing jobs and reducing income differences as well as to preferences for increases or decreases in spending on specific government programs.

While analyses of political influence may be less demanding from the standpoint of measurement than analyses of congruence, taking seriously the notion that influence entails a causal relationship between preferences and policy outcomes raises a host of daunting complications – essentially the same complications that arise in any attempt to make causal inferences based on statistical associations. One problem is that measured public opinion may be an *effect* as well as a *cause* of policy outcomes. This is especially likely to be the case in cross-sectional analyses of relatively stable policies and opinions. For example, Brooks and Manza's (2007: 56) study of *Why Welfare States Persist* tracked public attitudes toward the welfare state in a variety of affluent democracies using broad questions about the government's responsibility to provide jobs and reduce income differences between the rich and the poor. They showed that responses to these questions were strongly correlated with countries' welfare state spending. But did "the policy preferences of national populations

---

[8] Bartels (1985) sketched a statistical framework for analyzing situations involving both power (defined as the impact of actors' preferences on outcomes) and influence (the impact of actors' preferences on other actors' preferences); but that complication has generally been ignored in empirical analyses of political inequality.

strongly influence aggregated welfare state spending," as Brooks and Manza surmised, or did long-standing differences in the scope of countries' welfare states shape their citizens' views about the appropriate role of government?[9]

Another concern is that analyses of political influence may be sensitive to the specification of how citizens' preferences matter. Many studies of inequality focus on disparities in responsiveness to the preferences of affluent, middle class, and poor people, assigning separate regression coefficients to people in each tercile of the income distribution or to preferences imputed to people at a few specific points in the income distribution. As Achen (1978: 480) argued in the context of studies of congressional representation, "estimating a distinct influence coefficient for every individual would be computationally infeasible and theoretically uninteresting." Thus, analyses of this sort implicitly assume that everyone in the same income subgroup is equally influential. But subgroups may be more or less heterogeneous, and the implications of the tradeoff between bias (from treating heterogeneous individuals as identical) and imprecision (from treating them as distinct) deserve careful attention.[10]

Heterogeneity in political influence is almost surely greatest for high-income subgroups. Given the distribution of income in capitalist societies, the long upper tail has its own long upper tail, which has its own long upper tail, *ad infinitum*. Thus, if political influence is proportional to income, a simple average of the policy preferences of people in the top one-third or one-fifth of the income distribution may be a poor approximation of their *effective* preferences weighted by political influence. No one has managed to measure the political preferences of rich people with sufficient precision across space, time, or political issues to produce a systematic analysis of their impact on policy outcomes. However, scholars have gathered more limited descriptive data on the preferences of rich people and have used those data to speculate about the political power of the wealthy (Page, Bartels, and Seawright 2013; Page, Seawright, and Lacombe 2019).

It is also worth bearing in mind that even the most careful delineation of citizens' preferences along one dimension may be misleading if it overlooks other bases of inequality. Most contemporary research has focused on the translation of economic inequality into political inequality; but in some settings, differences in income may be less consequential than racial, ethnic, or other social distinctions. Moreover, the effects of distinct but correlated bases of inequality may easily be confounded. Are poor people underrepresented because they are poor, or because they are disproportionately women and members of racial and ethnic minority groups?

---

[9] Kenworthy (2009) noted that cross-national differences in welfare state effort are quite stable over long periods of time, making it very difficult to discern whether supportive public attitudes are a cause or an effect of government policy.

[10] On the statistical considerations arising in pooling disparate observations, see Bartels (1996a).

More broadly, policy outcomes are shaped by a wide variety of factors besides citizens' preferences. Kingdon's (1989) study of roll call voting in the U.S. Congress portrayed constituents' opinions as one among several important considerations shaping members' voting decisions.[11] But while it may be possible to construct a general list of potentially important actors in policymaking, the specific factors that may confound any particular analysis are likely to vary from case to case. Public employee unions loom large in some local policy domains, developers and business interests in others; ignoring these groups will make it hard to get sensible estimates of political influence (Anzia 2022). In setting defense budgets, policymakers are likely to be sensitive to the magnitude of external security threats. Those threats may also affect citizens' defense spending preferences, producing a spurious correlation between citizens' preferences and policy outcomes even if policymakers act solely on the basis of their own strategic judgments (Hartley and Russett 1992). Once we approach the problem of measuring political inequality as a problem of causal inference, the variety of potentially relevant factors to be considered is no less complex than the policymaking process itself.

One ubiquitous potential confounding factor in analyses of this sort is the preferences of the policymakers themselves. Perhaps affluent citizens only *appear* to be influential because their preferences happen to coincide with what policymakers were going to do anyway. Elkjær (2020: 2232, 2238) related Danish government spending in a variety of policy domains to the preferences of affluent, middle-class, and poor citizens. He found that "political representation appears to increase monotonically with income"; but his interpretation of that finding was that high-income groups have preferences that better reflect current economic and political circumstances. Accordingly, when governments pursue standard macroeconomic policies, such as stabilizing fiscal policies, these short-term policy changes more closely reflect the preferences of high-income groups. But the bias is coincidental, driven by better information, rather than a substantive overrepresentation of the "interests of the rich."

A direct test of this interpretation would require adding measures of policymakers' own preferences to Elkjær's "influence" analyses and seeing whether the apparent impact of high-income preferences was reduced or eliminated. Unfortunately, analysts of responsiveness rarely have access to reliable measures of policymakers' own preferences.[12] A more feasible approach would be

---

[11] Kingdon (1989: 18) tabulated members' spontaneous mentions of various actors in explaining their decisions on a series of specific roll call votes. Constituencies were mentioned in 37 percent of the cases, fellow members in 40 percent, interest groups in 31 percent, and the administration in 25 percent, with party leaders, staff, and "reading" mentioned less frequently.

[12] Some analysts have employed rough proxies for policymakers' own preferences, such as partisanship or statements in party manifestos. Examining the roll call votes cast by US senators, Bartels (2016: 235–249, 347) interpreted substantial differences in the voting behavior of Democrats and Republicans representing similar constituencies as reflections of "partisan ideologies," concluding that "the specific policy views of citizens, whether rich or poor, have less impact in the policy-making process than the ideological convictions of elected officials."

to augment the analysis with measures of government partisanship, macroeconomic conditions, and other factors potentially relevant to spending decisions. If those factors are consequential and positively correlated with the preferences of high-income citizens, then accounting for them would indeed reduce the apparent influence of high-income citizens' preferences on government spending.

In another article, Elkjær and Iversen (2020: 269–270) related long-run social spending in twenty-one affluent democracies to average support for redistribution in different income classes. They interpreted their results as "point[ing] to the critical role of the middle class" and indeed as "suggest[ing] that the level of redistribution is largely decided by the middle class." However, adding a measure of average government partisanship in each country produced a much better fit to the data, while the apparent impact of middle-income preferences evaporated, suggesting that the preferences of political elites were more consequential than those of the middle class – and mostly *not* themselves accounted for by the preferences of the middle class.[13]

Of course, the impact of government partisanship on policy is likely to vary significantly by country and policy domain. One advantage of analyses focusing on specific policy domains, like Elkjær and Iversen's, is that they facilitate assessing the direct impact on policy outcomes of partisanship and other factors correlated with but distinct from citizens' preferences. Capturing these effects in catch-all analyses including dozens of different policies will generally be much more difficult. For example, Mathisen and colleagues in this volume explore the impact of government partisanship on linkages between citizens' preferences and policy outcomes, but the main effects of "left government" in their analyses capture general orientations for or against policy change, not the leeway of governments to promote or block specific policies based on their own ideological proclivities. An additional complexity, addressed by Becher and Stegmueller in this volume, is that governments' own ideological proclivities may be shaped, in part, by citizens' preferences through both electoral selection and lobbying.

The ubiquity of concerns regarding potential confounding factors in analyses of political influence is daunting; as Wlezien (2017: 562) observed in surveying research on political responsiveness, "It is simply hard to demonstrate causality in observational studies." It is no more likely that analysts will agree about the theoretical and statistical assumptions required to make persuasive causal inferences in this realm than in any other. Thus, there is good reason to be modest about our conclusions. Yet that is no good reason to refrain from drawing conclusions, with due allowance for uncertainty – or to use the difficulty of the task as an excuse for pretending that simpler analyses will suffice.

---

[13] Cross-national analyses of changes in social spending using similar data (Bartels 2017: 57–59) likewise found most of the variation accounted for by factors other than citizens' preferences, though the estimated effects of high-income preferences were also, in several cases, substantial.

MULTICOLLINEARITY, PREFERENCE
DIVERGENCE, AND INEQUALITY

Having sketched in general terms the significance of congruence and influence as dimensions of potential political inequality, it may be helpful to consider some examples of how these concepts have been employed in the scholarly literature. One common bugaboo in analyses of this sort is that the policy preferences of distinct subgroups of citizens are often highly correlated across time or space. From the standpoint of assessing congruence, that is not really a problem, though it can be a source of confusion when analysts mistake correlation for similarity. As Gilens (2015b: 1068) noted, "even a strong correlation between two groups' preferences need not imply similar levels of congruence between preferences and outcomes." In studies of social spending, for example, the preferences of distinct income subgroups are often highly correlated across countries or over time, but with substantial, ubiquitous preference gaps between subgroups producing greater congruence for some subgroups than others.

From the standpoint of assessing political influence, multicollinearity is both a real problem and a pseudo-problem. Statistically, the effect of multicollinearity is to produce less precise estimates of the impact of each subgroup's preferences. For some purposes, that is a substantial disadvantage, for others not so much. If our scientific interest is really in *inequality* rather than in the extent of responsiveness to each group considered separately, it may be feasible to recast our analyses (by redefining our explanatory variables) to focus directly on the impact of *differences* in subgroup preferences, which are less likely to be highly correlated. Schakel, Burgoon, and Hakhverdian (2020) and Mathieson et al. (in this volume) provide examples of that approach.[14]

But aside from its statistical implications, multicollinearity has also produced a good deal of conceptual confusion and misdirection. While *perfect* collinearity between two (or more) explanatory variables in a multiple regression analysis makes it impossible to distinguish their separate effects, high levels of collinearity short of this extreme violate none of the standard assumptions of regression analysis; neither the regression parameter estimates

---

[14] These analyses, like Gilens' employ estimated preferences of citizens at the 90th, 50th, and 10th percentiles of the income distribution, denoted P90, P50, and P10. While P90, P50, and P10 are likely to be highly correlated, P90 can be rewritten as (P90–P50)+P50 and P10 can be rewritten as P50–(P50–P10). Relating policy outcomes to P50, (P90–P50), and (P50–P10) rather than to P50, P90, and P10 captures the same information about preferences, but isolates the *differential* impact of affluent and poor citizens' preferences *relative* to those of middle-income citizens. The parameter estimate for P50 in this analysis reflects a combination of the influence of all three groups, so is no longer directly interpretable as the impact of middle-income preferences. Analyses with only two explanatory variables, P50 and one of (P90–P50), (P50–P10), or (P90–P10), will also be difficult to interpret, since they impose implausible constraints on the estimated influence of one or more of the three groups.

nor their standard errors are biased.[15] The standard errors will be larger than they would be with less-correlated regressors – just as the standard errors will be larger than they would be with more observations. In either case, if the results are too imprecise to answer the questions being asked, the solution is straightforward: find more data.

Unfortunately, finding more data can be hard. Thus, scholars have sometimes attempted to sidestep the problem of having too little data by resorting to statistical shortcuts. Soroka and Wlezien (2010: 161–165), for example, proposed a model in which annual changes in government spending in each of several policy domains are related to the spending preferences of subgroups of citizens (differentiated by party, education, or income), with distinct weights translating each subgroup's preferences into policy change. "Applying this approach here," they wrote, "is complicated by very high multicollinearity" among preferences for change in the distinct subgroups. "To assess differential responsiveness, therefore, we separately model the effect of each group's preferences."

It is hardly surprising that regression analyses with fifteen to thirty-three slow-moving annual observations of preferences and spending are insufficient to estimate disparities in responsiveness to a variety of distinct subgroups. Unfortunately, there is no reason to think that the alternative of comparing parameter estimates from separate models focusing on each subgroup's preferences in isolation can shed any reliable light on the question of "whether policy responds more to the preferences of some groups than others." Each of these mutually contradictory analyses is biased by the omission of other subgroups' preferences (aside from any other factors) from the set of relevant explanatory variables. Moreover, the higher the correlations among the subgroup preferences are, the more severely biased the bivariate regression parameter estimates will be. There is simply nothing useful to be learned from analyses of this sort about disparities in political influence.

The implications of correlated subgroup preferences are further muddled by a tendency to mistake statistical imprecision for evidence in favor of null hypotheses. Using spending and survey data from the United States, Wlezien and Soroka (2011: 299, 302, 298) assessed income-group differences in dynamic representation across thirty-five years and six different policy domains. Only three of the resulting eighteen estimates of responsiveness (for each of the three income groups in each of the six domains) were "statistically significant," and the authors concluded that "it is difficult to distinguish responsiveness to particular groups." So far, so good. However, by the end of their chapter,

---

[15] The notion that "Gilens and Page's analyses are questionable based on concerns about collinearity among the independent variables" (Branham, Soroka, and Wlezien 2017: 58) is sometimes attributed to Bashir (2015), overlooking fatal flaws in Bashir's simulation analysis noted by Gilens (2016). Winship and Western (2016) provided a Bayesian analysis of how multicollinearity can exacerbate biases stemming from misspecification, but no reason to think that omitting relevant variables would mitigate those biases.

this statistical uncertainty was somehow transmuted into substantive equality: policymakers, they concluded; "appear to be guided as much by the median voter as anyone else. This is about all that we would expect if people had equal weight in the policymaking process." In fact, their estimates of responsiveness to the rich, averaged across policy domains, were almost 50 percent larger than those for the "median voter," while the average estimated responsiveness to low-income people was slightly negative.[16] Given the limitations of the data and analysis, this is certainly not conclusive evidence of unequal influence, but it is even less indicative of "equal weight in the policymaking process."

Another way to generate inconclusive statistical results is to limit the analysis to small subsets of cases. Branham, Soroka, and Wlezien (2017: 60, 56) analyzed 185 of Gilens' 1,779 proposed policy changes,[17] those where majorities of affluent and middle-income people disagreed. The result of truncating the sample was to inflate the standard errors of the key parameter estimates by a factor of four or five, leading the authors to conclude that "it is nearly a coin flip as to which group wins," a result they interpreted as "more encouraging (normatively speaking) than recent scholarship." Statistical analyses that are too underpowered to shed light on quantities of interest are not "encouraging," they are simply uninformative.

Why focus on cases in which majorities of income subgroups disagree? According to Branham, Soroka, and Wlezien (2017: 56, 60), "We know that disagreement in policy preferences is a necessary condition for differential representation. If majorities in different income cohorts prefer the same policy, we cannot distinguish whose preferences are being represented." "Differential representation" here seems to mean differential congruence between preferences and policy outcomes. But clearly, disagreement between subgroup majorities is not a necessary condition for differential congruence. If a policy is adopted with 80 percent support from one subgroup and 51 percent support from another subgroup, clearly more people in the first subgroup than the second got their way. Nor does agreement between subgroup majorities imply equality of influence. Indeed, when the authors examined cases where majorities of affluent and middle-income people agreed, they found strong evidence of unequal influence.[18]

---

[16] The average responsiveness estimates were 0.187 for the high-income group, 0.128 for the middle-income group, and –0.034 for the low-income group. Elsewhere in the same edited volume, Bhatti and Erikson (2011: 241) provided a rather more nuanced interpretation of ambiguous empirical results, writing that "Conclusive statistical evidence could not be found in favor of the differential representation hypothesis."

[17] Gilens (2012), for the most part, and Gilens and Page (2014) focused on 1,779 policy questions asked in U.S. opinion surveys between 1981 and 2002, relating the opinions of survey respondents at various points in the income distribution (imputed from the quadratic relationship between preferences and reported incomes for each survey question) to subsequent changes in policy.

[18] In 1,594 cases with coincident majorities, the estimated impact of "Rich Preferences" (from a structural equation model taking account of measurement error in subgroup preferences) was

Some analysts have focused on cases of preference divergence in the apparent hope that doing so would mitigate statistical biases resulting from employing mutually contradictory bivariate analyses of influence. An analytical shortcut in Gilens' book seems to have served as an encouraging example in this respect. His most persuasive evidence of unequal influence was derived from regression analyses simultaneously incorporating the preferences of affluent, middle-class, and poor people and allowing for correlated measurement error in the estimated preferences of the three income subgroups (Gilens 2012: 85–87, 256).[19] However, in much of his book, he presented the results of simpler bivariate statistical analyses relating policy outcomes to the preferences of each income subgroup separately, first for his entire sample of 1,779 policy questions and then for subsets of issues where the subgroups' preferences differed. He was clear about the inferential limitations of the latter approach. "To assess the ability of citizens at different economic levels to influence government policy," he wrote (2012: 78), "we need to know not the strength of the overall preference/policy link for each income group, but rather the strength of this association net of the impact of other income groups." Nonetheless, he offered parallel analyses of subsets of issues where subgroups' preferences diverged as "an alternative to multivariate analysis," noting that "this technique produces results comparable to a multivariate model when the multivariate approach is feasible."

The similarity to his more sophisticated statistical findings notwithstanding, I know of no reason to think that limiting analyses to cases of preference divergence will overcome the bias resulting from misspecified bivariate models. While sample selection may reduce the correlation between subgroup preferences, and thus the bias resulting from misspecification, the bias would only be eliminated if that correlation were reduced to zero – and in that case, the cost in precision of including multiple subgroups in the analysis would also be eliminated, so there would still be no reason to prefer a bivariate model.

Gilens' shortcut was relatively benign, in that the key results of his bivariate analyses were corroborated by more sophisticated analyses, either in the appendix of his book or in subsequent work by Gilens and Page (2014). However, there is no comparable corroborating evidence for many other bivariate analyses of subsets of issues on which the average preferences of income subgroups diverge, either in absolute terms or in the sense that a majority of

---

0.757 (with a standard error of 0.079); the estimated impact of "Middle Preferences" was 0.032 (with a standard error of 0.082).

[19] Gilens' correction for measurement error employed estimates of error variances and covariances derived from the subset of cases in which substantively similar policy questions were asked of independent survey samples in the same calendar year. The persuasiveness of his results was bolstered by careful examination of a variety of potential alternative explanations for his findings of unequal influence, including differences across income subgroups in the reliability, intensity, and homogeneity of policy preferences and in levels of education.

one subgroup favored a proposed policy change that a majority of another subgroup opposed.[20]

Sometimes, bivariate analyses have been presented not just as shortcuts for assessing disparities in political influence, but as significant in their own right. For example, Enns (2015: 1055) proposed "relative policy support" as a benchmark for assessing representation, arguing that a positive correlation between the strength of a subgroup's support for various policies and the probability that they are adopted constitutes "straightforward – perhaps even axiomatic … evidence of representation." But it is very hard to see why subgroup members should be gratified by a correlation that implies neither congruence nor influence. This is a conception of representation with little apparent grounding in any theory of democracy.[21]

In other cases, it is unclear whether bivariate statistical associations are supposed to be measuring congruence, influence, or something else. In their study of the relationship between support for redistribution and levels of social spending in twenty-one democracies, Elkjær and Iversen (2020: 267–268) estimated "simple bivariate responsiveness models to examine how well social spending aligns with the preferences of each income class." They found that the bivariate relationship was "strongest for the middle class, suggesting that the middle class is instrumental in setting the level of redistribution." This sounds like a simple conflation of "alignment" with influence. However, Elkjær (2020: 2228) separately offered a different-sounding interpretation of "policy alignment": "Unequal policy responsiveness should be disaggregated into two concepts: policy alignment and policy influence. Policy alignment conceptualizes the extent to which policies correspond to subgroup preferences, whereas policy influence conceptualizes the degree of independent influence of subgroup preferences on policies." Here, "policy alignment" seems intended to capture something like congruence, distinct from influence. But what? The estimated slopes from bivariate regression analyses – indeed, from *any* regression analyses – shed no light on how well policy outcomes satisfy any individual's or subgroup's preferences.

On the other hand, if the bivariate "alignment" between policy outcomes and subgroup preferences is supposed to be significant in its own right, as with Enns's notion of "relative policy support," the logic is equally murky. Why should a person living in any one of Elkjær and Iversen's twenty-one democracies be expected to care how closely spending policies in other countries "align" with the average preferences of people in the corresponding income groups in those countries? If "alignment" is not a measure of congruence or influence, it seems to be a statistical measure looking for a theoretical rationale.

---

[20] Bowman (2020) provided a comprehensive assessment of analyses of various subsets of Gilens' data employing alternative "preference gaps" and "preference thresholds."

[21] On the logic of "relative policy support," see Gilens (2015b: 1066–1068).

CONGRUENCE, INFLUENCE, AND
COINCIDENTAL REPRESENTATION

Even if analysts of political inequality could agree about how to conduct their empirical analyses, they would still be left to wrestle with the implications for democracy of findings regarding congruence and influence. Gilens' data from the United States revealed substantial disparities in apparent influence across income groups, but only modest differences in the extent to which citizens got the policy outcomes they preferred. Parallel analyses of European data by Mathisen and colleagues (this volume) reveal a similar pattern, as do a variety of other studies employing different research designs. As Soroka and Wlezien (2008: 325) wrote of the first wave of such studies, "we take that research to imply that policy would represent the median voter *only* because the preferences of people with middling income are much like the preferences of those with high incomes. From this perspective, representation of the middle would be indirect."

These findings raise two distinct issues, one empirical and the other normative. The empirical issue turns on the prevalence of what Soroka and Wlezien referred to as "indirect" representation and Gilens and Page (2014: 573) termed "democracy by coincidence, in which ordinary citizens get what they want from government only because they happen to agree with elites or interest groups that are really calling the shots." Soroka and Wlezien (2008: 325) acknowledged that "there are differences in preferences across income levels in some important policy domains," but argued that "regardless of whose preferences policymakers follow, differences across income groups are often rather small, and policy will end up in essentially the same place." Gilens (2015b: 1070, 1065) was more pessimistic, acknowledging that "'democracy by coincidence' is an important feature of contemporary American politics," but emphasizing specific "important and highly salient issues on which the power of the affluent and interest groups has pushed policy away from the preferences of the majority."[22]

Statistical analyses aggregating hundreds of distinct policy issues tend to occlude detailed consideration of differences among them, including differences

---

[22] In some cases, scholars have employed selective citation to bolster broad claims that policy disagreement between income subgroups is "relatively rare." For example, Elkjær and Iversen (2020: 257, 258) argued that "unequal representation is naturally quite limited on most policies with no redistributive aim, since class preferences barely diverge." In support of this claim, they cited Soroka and Wlezien's (2008: 319) tabulations of responses to eight spending questions in the United States over twenty-four years, ignoring Gilens' (2009: 339) response documenting substantial gaps between the average preferences of income subgroups across hundreds of survey questions drawn from a wide range of policy domains, including not only social welfare, taxes, and economic policies, but also moral issues and foreign policy and national security. Similar preference gaps appear elsewhere; for example, European survey data reveal significant differences between income subgroups in attitudes toward gay rights, the role of science in addressing environmental problems, trust in the legal system, and other issues.

in the similarity of preferences across subgroups and potential differences in the influence of specific actors in different policy domains. Gilens' examinations of variation across policy domains (2012: ch. 4) and political contexts (2012: ch. 6–7) are a notable exception in this regard, but much more work of this sort will be necessary to clarify the empirical significance of "democracy by coincidence."

The normative significance of coincidental representation is an equally important issue, but much harder for empirical analysts to adjudicate. Gilens (2015b: 1070) argued that "democracy by coincidence is a debased and conditional form of democracy (if it is a form of democracy at all)." Kolodny (2023: 304) reached a similar conclusion on philosophical grounds, arguing for "a democratic ideal not of correspondence, but instead of influence: not of satisfying the People's policy preferences, but instead of ensuring the People's control over policy." For the most part, however, and despite its seeming prevalence, "democracy by coincidence" has received rather little attention from theorists of democracy.

CONCLUSION

As Gilens (2012: 47) observed, "There is no single right way to assess something as complex as government responsiveness to public preferences; alternative approaches offer different sets of trade-offs and limitations." From the standpoint of research design, studies in which the units of analysis are distinct policy proposals – like those described by Gilens, and by Mathisen and colleagues in this volume – rest on rather different assumptions and offer rather different analytical opportunities than those focusing on temporal or cross-national variation (or both) in a single policy domain. Cross-sectional studies relating citizens' preferences to the preferences or choices of specific policymakers or parties may help to overcome ubiquitous data limitations, but they require careful attention to the question of how policymakers' choices are aggregated into policy outcomes.

No one analytical template will or should monopolize the study of political inequality. However, in designing research, it behooves us to be as clear as possible about what we hope to learn, how, and why. My focus here has been on two key aspects of political inequality – *congruence* and *influence*. Each of these concepts has a (relatively) coherent theoretical pedigree with (relatively) unequivocal methodological implications. While I do not mean to suggest that these two concepts exhaust the ways in which we might study political inequality, alternative approaches have yet to find comparable grounding in democratic theory. Attaching significant-sounding labels to measures "plucked from the statistical shelf and employed without much theoretical interpretation," as Achen (1977: 806) put it more than forty years ago, is unlikely to produce much real insight.

For analysts aspiring to measure inequality in the extent of congruence between citizens' preferences and policy outcomes, the key challenge will be to

calibrate preferences and policies, either by coding policy outcomes to harmonize with existing survey data (the approach taken by Gilens and by Mathisen and colleagues) or by employing survey data that take the policy status quo as an explicit point of reference (as in studies of governmental spending). Both of these approaches suggest that the preferences of affluent citizens are better satisfied than those of poor citizens, though the differences are often modest in magnitude.

If our interest is in measuring differences in political power or influence, we will succeed to the extent that we can produce credible inferences regarding the impact of citizens' preferences on policy outcomes. The potential pitfalls here are of two broad sorts. On the one hand, there is the temptation to evade substantive difficulties by oversimplifying. As in most realms of social research, bivariate analyses are not a promising basis for inferring causality. Analyses representing the policymaking process as a simple contest among the preferences of distinct subgroups of citizens will generally be somewhat more informative, though still less credible than more sophisticated analyses taking account of political parties, interest groups, and other salient actors in the policymaking process. Analyses that also take account of the potential indirect influence of citizens via parties, interest groups, and other salient actors will be most persuasive of all.[23]

On the other hand, there is the temptation to evade substantive difficulties by imposing unrealistic standards of perfection on our data analyses. While experimental research has occasionally shed valuable light on responsiveness, its utility in this realm is likely to be limited, given the scale and complexity of the political processes involved.[24] For the most part, we will have to do the best we can with empirical analyses that reflect the policymaking process sensibly rather than precisely, producing inferences that are never wholly persuasive. Given the rudimentary state of knowledge in the field, even experienced scholars will often disagree about the persuasiveness of any specific analysis. Disagreement is to be expected, a natural feature of the scientific process of criticism and successive approximation. Nonetheless, we can hope that results from multiple studies with distinct strengths and weaknesses in different political contexts will gradually produce a clearer picture of the unequal distribution of political influence in contemporary democracies.

When Gilens and Page's (2014) analysis was published, I argued that "their findings should reshape how we think about American democracy."[25] That

---

[23] Of course, citizens' preferences are also shaped by parties, interest groups and other salient actors, raising additional normative and empirical complexities that are generally ignored in this literature.

[24] Butler (2014) and Kalla and Broockman (2016b) used field experiments to assess biases in the responsiveness of congressional offices to constituents' requests for assistance and access, respectively.

[25] Larry Bartels, "Rich People Rule!" *Washington Post*, Monkey Cage, April 8, 2014 (www .washingtonpost.com/news/monkey-cage/wp/2014/04/08/rich-people-rule/).

assessment may have been too modest. Subsequent research on other countries suggests that substantial disparities in political influence are ubiquitous in affluent democracies (Bartels 2017; Elsässer, Hense, and Schäfer 2021; Mathisen and colleagues in this volume; Schakel, Burgoon, and Hakhverdian 2020). Those findings imply that political inequality is *not* primarily attributable to specific features of the US system, such as permissive campaign finance regulations, weak unions, and a policymaking process with myriad veto points. Its roots apparently lie much deeper in the social and political soil of democracy than even pessimistic analysts have supposed.

Political science, like politics, involves a lot of slow boring of hard boards. In the past two decades, the scientific study of political inequality has advanced considerably. Nonetheless, we have only begun to scratch the surface of the problem, and much more work will be necessary to confirm and extend our understanding of the magnitude and bases of inequality in putative democracies. The challenges are formidable, but it is difficult to think of a more vital set of questions.