Districting Principles and Democratic Representation

Thesis by

Micah Altman

In Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

California Institute of Technology

Pasadena, California

Mar. 31, 1998

Districting Principles and Democratic Representation

Acknowledgements

I thank Morgan Kousser for continuous guidance and wise advice throughout the course of the dissertation, as well as for teaching me to write. I wish to thank Scott Page for his numerous insightful comments and assistance during the course of the dissertation. I thank Mike Alvarez and Rod Kiewiet for their comments and assistance during the course of the dissertation.

I also thank Gary King, Daniel Lowenstein and Ken McCue for their comments on individual chapters.

Abstract

Redistricting is always political, increasingly controversial, and often ugly. Politicians have always fought tooth-and-nail over district lines, while the courts, for most of their history, considered the subject a thicket too political even to enter.

Three decades ago the courts finally entered the political thicket, ruling in *Baker* v. *Carr* (1962) that redistricting was justiciable. A decade ago, the court showed signs that it wanted to chop the thicket down, ruling in *Davis* v. *Bandemer* (1986) that partisan gerrymanders were actionable. But, in fact, few suits followed this potentially momentous decision. Just five years ago, however, the court took its ax to the thicket in earnest: In a line of cases starting with *Shaw* v. *Reno* in 1993, and continuing through the

1996-97 term of court in *Abrams* v. *Johnson* (1997), the Court has made a strong bid to outlaw what it terms “racial gerrymandering.” In this attempt to eliminate gerrymandering, the Court has placed an extreme emphasis on what they term “traditional districting principles,” which are primarily formal, measurable criteria such as population equality, compactness, and contiguity.

This extreme emphasis threatens to radically change the redistricting process in the United States. Justice Souter, in a dissent in *Vera* in which Justices Ginsburg and Breyer joined, argued that the logic of the *Shaw* line of cases can lead only to one of two outcomes: Either “the Court could give primacy to the principle of compactness,” or it could radically change traditional districting practice -- eliminating it or “replacing it with districting on some principle of randomness...”

In this dissertation, I examine “traditional districting principles,” and their implications for representation. I am motivated by, and attempt to answer, the following questions: What theories of representation are implicit in the Court's recent line of cases? Where do “traditional districting principles” come from, and are they really traditional? Are the formal standards that the Court wishes to adopt judicially manageable? Are they theoretically consistent? What effect will using these principles have on politics? Can we eliminate politics in redistricting by automating the process?

Contents

The dissertation is organized into six chapters. In Chapter 1, I discuss the legal debates over redistricting principles, and how this dissertation, and political science in general, can shed light on this debate. In Chapters 2 and 3 I define measures for and gather data about historical and modern districts. In Chapter 4 I develop a model to predict the partisan effects of applying strict compactness standards. In Chapter 5, I analyze the theoretical and practical limitations of mechanically applying any formal districting principles. Finally, in Chapter 6, I apply statistical models to determine the effects of traditional districting principles on recent elections.

- Unprincipled Limitations on Gerrymandering: The Supreme Court's Tempestuous Use of Traditional Districting Principles
- The Consistency and Effectiveness of Mandatory District Compactness Rules
- Traditional Districting Principles: Judicial Myths vs. Reality
- Predicting the Electoral Effects of Mandatory District Compactness on Partisan Gerrymanders
- Is Automation the Answer? -- The Computational Complexity of Automated Redistricting
- Do Traditional Districting Principles Matter?

1. Unprincipled Limitations on Gerrymandering: The Supreme Court's Tempestuous Use of Traditional Districting Principles

Theories created in the absence of fact are fantasies, and decisions made in the absence of theory are impulses. In its latest opinion on redistricting, *Bush* v. *Vera*, the Supreme Court produces both. It is a truism that judicial principles emerge from the consideration of individual cases, and we do not expect theories of representation to spring from the court like Athena from the head of Zeus, fully-formed. After more than three decades of redistricting cases, however, the Court should be able to give a consistent answer to redistricting's central legal question: What constitutional harm does gerrymandering cause?

In an effort to avoid considerations of politics, the Court has turned to “traditional districting criteria.” I show that the Court's use of these districting principles distorts the history of districting, exaggerates the political importance of these principles, and ignores theories of political representation.

2. The Consistency and Effectiveness of Mandatory District Compactness Rules

As the technology for drawing districts has become more sophisticated, and as the Supreme Court has grown more critical of district lines, academics and politicians have renewed their interest in evaluating and regulating legislative districts. In the field of redistricting, one of the most significant and controversial claims is that gerrymandering can be easily eliminated by requiring districts to be “compact.”

Compactness criteria attempt to measure the irregularity of a district's shape; in other words, they capture its ugliness. Political scientists and geographers have measured compactness in many different ways, and some of these measurements have been used to

investigate isolated district plans. There is, however, no scholarly consensus on which compactness measure, if any, is best. Furthermore, while scholars have debated the merits of compactness measures in general terms, most of this debate has been based only upon hypothetical or isolated examples. Political scientists have done little formal modeling of or empirical research into this issue. Many important questions remain open: What, exactly, are all of these compactness criteria measuring? Are these measures consistent with each other -- does it matter, really, which one we use? Which measures are best?

In this chapter, I answer these questions by using formal analysis and by exhaustively analyzing small sets of districts. First, I find that many compactness criteria can, in fact, contradict each other; contrary to the claims of some previous researchers, it matters which measure we choose. Second, I find that some measures are, indeed, better than others -- though the existence of a single best measure is doubtful.

3. Traditional Districting Principles: Judicial Myths vs. Reality

Compactness, contiguity, respect for electoral boundaries and population equality have been hailed as “traditional districting principles.” Proponents bemoan their decline, and blaim modern techniques for gerrymandering and the creation of majority-minority of districts. Are these principles traditional? Are they in decline, and if so, why? In this chapter, I examine this question in the light of historical evidence from all district plans 1789 and 1913, and decadal redistricting plans from 1923 to 1993.

I find that historical districts were more likely to be both regularly shaped and to follow natural boundaries than are modern districts. Most of the decline in district compactness, however, directly followed the Court's decision to impose strict equal population standards on districts, far preceding the creation of majority-minority districts and modern use of computers in redistricting. Moreover, a study of historical Congressional debates shows that formal districting principles such as contiguity were subordinate to the main purpose of redistricting -- expressing representational values.

4. Predicting the Electoral Effects of Mandatory District Compactness on Partisan Gerrymanders

Proponents of a compactness standard have offered it as a politically neutral solution to the problem of gerrymandering. But are such standards, in general, electorally neutral?

In this chapter, I examine the effects of compactness standards on political representation when some political groups are geographically concentrated. By treating redistricting formally as a combinatorial optimization problem, I examine the neutrality of compactness standards and the ability of such standards to prevent gerrymandering. Since these problems cannot, in general, be solved exactly, I use Monte-Carlo techniques, simulated annealing, genetic algorithms, and other simulation techniques to solve them approximately.

These simulations reveal that compactness standards, when strictly applied, do constrain electoral manipulation, but that they are not electorally neutral. The particular effects of compactness depend on both the distribution of voting groups and the political institutions under which districts are created. If compactness attains primacy over other districting principles, large geographically concentrated minority groups will benefit. On the other hand, where redistricting is primarily a partisan process constrained by compactness, the party which relies on such groups will be relatively weakened by compactness constraints.

5. Is Automation the Answer? – The Computational Complexity of Automated Redistricting

Over the last three decades, many academics, politicians, and judges have called for redistricting to be automated in order to prevent gerrymandering and promote electoral fairness. Automated redistricting has been offered as a general-purpose, unbiased, and value-free method for creating districts. While proponents have consistently expressed optimism about its feasibility and benefits, the results of automated redistricting systems have fallen short of these optimistic expectations.

In this chapter, I explain the failure of automated redistricting: I show that for any computer program to find the "best" district, it must solve a mathematical problem that is *computationally complex*; redistricting belongs to a class of problems that many computer scientists believe to be impossible to solve precisely and efficiently.

I argue that it may be impossible to design an automated redistricting system that both is assured to find optimal districts and is "value-free." Because of the difficulty of the redistricting problems, automated redistricting methods may always contain biases in the types of districts they create or assumptions about the values to be used in the redistricting process.

6. Do Traditional Districting Principles Matter?

In recent cases, the Supreme Court has given geographical compactness and other “traditional districting criteria” a large role in its “strict scrutiny” of majority-minority districts. In the future, it is possible that formal measures of district shape will become as pervasive in the design of district plans as formal measures of district population equality are at present. Yet a central empirical question remains unanswered: Do these principles ultimately affect elections? Do these bizarre districts, as opponents argue, cause “expressive harms” to voters? In this paper I use multiple measures to evaluate congressional districting plans, and maximum-likelihood models to analyze the relationship between the “traditional” properties of modern and historical district plans electoral outcomes.

I find that while different population-equality measures, even those with poor theoretical properties, produce very similar evaluations of plans. On the other hand, different compactness measures fail to agree over the compactness of most districts and plans.

In effect, the courts can use any convenient measure of population equality and obtain similar results, while the courts' choice of compactness measures will significantly change the evaluations in each case. Since there is no single generally accepted measure of compactness, this disagreement among measures raises concerns about whether

compactness is a readily operationalizable notion, to use a social scientific formulation, or a judicially manageable one, to employ terms from law.

Furthermore, my results indicate that, in modern elections, traditional districting principles do not have many of the virtues attributed to them. Although reductions in malapportionment may reduce partisan bias, the addition of district compactness has little effect on partisan bias or responsiveness. The only detectable effect of shape was on turnout. Moreover, I could find little evidence that bizarre districts cause "expressive harms."

Chapter 1. Unprincipled Limitations on Gerrymandering: The Supreme Court's Tempestuous Use of Traditional Districting Principles

> Full fathom five thy father lies;
>
> Of his bones are coral made;
>
> Those are pearls that were his eyes:
>
> Nothing of him that doth fade
>
> But doth suffer a sea-change
>
> Into something rich and strange.
>
> Sea-nymphs hourly ring his knell
>
> -- The Tempest, Act I, Sc. II

1.1. The Consequences of Judicial Fantasy: A Tempest

Theories created in the absence of fact are fantasies, and decisions made in the absence of theory are impulses. In its latest opinions on redistricting, the Supreme Court produces both. It is a truism that judicial principles emerge from the consideration of individual cases, and we do not expect theories of representation to spring from the Court like Athena from the head of Zeus, fully-formed. After more than three decades of redistricting cases, however, the Court should be able to give a consistent answer to redistricting's central legal question: What harm does gerrymandering cause to constitutional rights or fundamental representational values, and why?

During the three-and-a-half decade plunge into the "political thicket" of districting, different Courts have attempted to shape different answers to these questions using empirical observations of politics, the history of representation in the U.S, and implicit and explicit theories of representation. The current Court, disliking previous Courts' conclusions, has instead constructed a judicial fantasy based upon so-called "traditional districting principles" that have little to do with the realities of politics, history or representation. In this chapter, I evaluate the Court's use of traditional districting principles (t.d.p.'s) in the light of reality.

For most of its history, the Court considered redistricting a thicket too political to enter. Three decades ago the courts finally entered the political thicket, ruling in *Baker* v. *Carr* (1962*)* that redistricting was justiciable. A decade ago, the Court showed signs that it wanted to chop the thicket down, ruling in *Davis* v. *Bandemer* (1986) that partisan gerrymander's were actionable. Little action followed this potentially momentous decision. Just five years ago, however, the Court took its axe to the thicket in earnest: In a line of cases starting with *Shaw* v. *Reno* in 1993, and continuing through year in *Abrams* v. *Johnson* (1997), the Court has made a strong bid to outlaw what it terms "racial gerrymandering." In this attempt to eliminate gerrymandering, the Court has placed an extreme emphasis on what they term "traditional districting principles" — primarily mathematically measurable criteria such as population equality, and compactness.

This extreme emphasis threatens to change radically the redistricting process in the United States. Justice Souter, in a dissent in *Vera* in which Justices Ginsburg and Breyer joined, argued that the logic of the *Shaw* line of cases can lead only to one of two outcomes: Either "the Court could give primacy to the principle of compactness," or radically change traditional districting practice — eliminating it or “replacing it with districting on some principle of randomness...” (*Vera* at 2009-2010).

The Court’s discovery and elevation of “traditional districting principles” raises many empirical and positive questions: Which districting principles really are “traditional”? How do we measure them? Who would benefit if traditional districting principles (t.d.p.’s) dominated the redistricting process? In the following sections, I examine the history of t.d.p.’s, their political effects, and their normative value. I then propose some ways in which the Court can extend its treatment of districting principles in order to make it more consonant with political history, science, and philosophy.

1.2. The Historical and Judicial Basis of Traditional Districting Principles – Suffering a Sea Change

One of the few things that is clear about the current set of redistricting decisions is that “traditional districting principles” are important — all of the recent redistricting cases agree on this point, at least. But, what are traditional districting principles? How do we know one when we see it? Why are they important? The Court’s opinions are imprecise and contradictory on such details.

1.2.1. Out of Nowhere – Traditional Districting Principles in **Shaw** *and its Successors*

Within the context of judicial opinions, “traditional districting principles” come from nowhere.[1] “Traditional districting principles,” as such, first appeared in the Supreme Court’s decision *Shaw* v. *Reno* (*Shaw I,* henceforth), and the phrase has multiplied through subsequent decisions. Symbiotically, the principles that the Supreme Court deems “traditional” have multiplied along with the phrase that refers to them.

The Court’s first mention of “traditional districting principles” appears in connection with Judge Voorhees’s lower Court dissent. In this debut, the Court attributed these principles to *UJO* v. *Carey:* “Chief Judge Voorhees agreed that race-conscious redistricting is not *per se* unconstitutional but dissented from *the* rest of the majority’s equal protection analysis. He read Justice White’s opinion in UJO to authorize race-based reapportionment only when the State employs traditional districting principles such as compactness and contiguity”(*Shaw I*).

How does the Court know that compactness and contiguity are traditional districting principles? The *UJO* opinion, in fact, did not refer to “traditional” districting principles

---

[1] Both a close reading of the major redistricting and malapportionment cases and a full-text search of all modern and major historical Supreme Court decisions (Infosynthesis 1998) fail to reveal any use of the phrases “traditional districting principles” or “traditional districting criteria.”

at all, but to “sound” (430 U.S. at 168) redistricting principles.[2] Moreover, the *UJO* court neither suggests that traditional and sound principles are the same, nor mentions contiguity as an example of either type.

Besides their brief reference to *UJO*, the Court gives no hint as to the origin of these traditional principles. They cite neither law nor history. If “traditional” principles do not come from *UJO*, as the Court implies, do they come from other precedents, from the Constitution, or from other historical sources? O’Connor, delivering the judgment in *Vera*, explicitly denies that the Constitution guarantees traditional districting principles, but neither *Vera* nor its predecessors refer to other sources. Are traditional districting criteria those that were used around the development of the Constitution, at the time of the 14[Th] amendment, or after? Are traditional criteria those that were followed by all states, by a majority of states, or by the particular state in question? Are traditional criteria those that were historically mandated or those actually used? When traditions conflict, which should we choose? Are any traditions invalid, *a priori*? New York, for example, has traditions of malapportionment, gerrymandering, ill-compactness, and non-contiguity, stretching back from before the Constitution. Are these to be honored as “traditional districting criteria”? Our strong impulse is to answer: “surely not,” but the

---

[2] *UJO* is by no means the first case in which *sound* or *rational* districting principles are recognized. See the discussion of *Gaffney* below, and the opinion in *Gaffney* itself at 749, for a discussion of earlier cases.

Court's failure to explain where traditional districting principles come from leaves us without a legal basis for this denial.

1.2.2. Destination Unknown – The Multiplication of Principles and their Definitions

Although no general rule for identifying traditional districting criteria emerges from the Court's decisions, examples of individual criteria multiply throughout these cases. In *Shaw I*, the Court first identifies compactness and contiguity as traditional districting principles, then further in the opinion identifies "respect for political subdivisions" as also belonging to this august class. In *Miller*, "respect for... communities defined by actual shared interests" (excluding race) makes its debut as a traditional districting criterion, without further commentary. In *Vera*, O'Connor refers to "natural geographic boundaries" and "regularity" for the first time as distinct traditional districting criteria.[3]

---

[3] The Court in *Shaw* (and repeated in *Miller*) cited *Gomillion* on the subject of the irregularity of a district plan: "In some exceptional cases, a reapportionment plan may be so highly irregular that, on its face, it rationally cannot be understood as anything other than an effort to –segregat(e)... voters- on the basis of race." Several other times in these opinions, the Court uses these terms interchangeably. In *Vera*, O'Connor appears to have discovered a distinction between "regularity" and "compactness" and treats them as separate, although related, criteria: "appellants do not deny that District 30 shows substantial disregard for the *traditional principles of compactness and regularity"* and

In *Abrams* v. *Johnson* (1997), the Court reveals that traditional districting principles included maintaining "district cores," precinct boundaries, "corner districts," and even "an urban black majority district" – at least in Georgia.

Rather than identifying traditional districting principles exhaustively or systematically describing the rules for deciding whether a principle is traditional, the court has chosen to reveal individual t.d.p.'s piecemeal. This approach leaves lower courts that are faced with plausible, but as yet unconsecrated, districting principles, unable to answer important questions: Which historical principles are important? How old does a principle have to be before it becomes "traditional?" How should lower courts weigh these against other goals of redistricting, traditional and otherwise?

Had the Court created an exhaustive list of all t.d.p.'s, lower courts and state governments would still be faced with the problem of how to assess whether a plan followed these rules. In Chapter 3, I show that even apparently straightforward theoretical principles, such as contiguity, preservation of county lines, and population equality, are subject to a variety of different practical interpretations and measurements.[4]

"Fifty percent of the district population is located in a *compact, albeit irregularly* shaped, core in south Dallas..." (emphasis added).

[4] While compactness has received the most recent scholarly attention, other principles suffer from similar problems of definition: For example, what is "irregularity," and how does it differ from compactness? O'Connor's only reference to measuring irregularity cites

It is particularly difficult to measure meaningfully the most widely discussed principle, compactness. Geographic compactness has never been defined in a way that was both mathematically precise and politically meaningful. Precision in defining these criteria is not the Court's strong suit: For example, in *Abrams* the Court disparaged the ACLU's alternative plan for its resemblance to an iguana, without an attempt to further define compactness. Not including such Rorschach tests, there are dozens of different ways of measuring compactness. In Chapter 2 I show that these measures are not

---

Pildes & Niemi (1993), but Pildes & Niemi do not define a measure of "irregularity" per se, but instead use the term as a synonym for compactness, as did the Court in *Shaw.* What types of geographic boundaries and political subdivisions will the Court recognize as legitimate? The Court, in *Vera*, criticizes the Texas districts for splitting voter tabulation districts and in *Abrams* similarly criticizes the splitting of precincts — but voter tabulation districts are far from "traditional" in the normal sense of the word: they are artificial units created by the Census to approximate precincts (U.S. Dept. of Commerce 1992), the precincts themselves are often changed, they often do not coincide with more historical political boundaries such as those dividing counties and cities, and they are not designed to represent any communities of interest. In fact, precincts are often created solely for the administrative convenience in holding elections — often by the same state government that is responsible for redistricting.

consistent with each other in theory, and in Chapter 6 I show that they conflict in practice.[5]

1.2.3. What are Historical *Districting Principles?*

While "traditional" districting principles remain difficult to identify, the historical record does suggest a number of candidates, at least for congressional districting. In Chapter 3, "Traditional Districting Principles – Judicial Myths vs. Reality," I delve into historical evidence, examining the congressional record surrounding districting legislation, and measuring compactness, contiguity, boundary-integrity, and

---

[5] Richard Pildes claims that "various measures of compactness tend to converge at extreme cases" (Pildes 1997, 2513). Although Pildes' claim is probably mathematically correct for most measures of compactness (i.e., it is possible to construct a bizarre shape such that multiple methods of measuring compactness give it a poor score), compactness measures can give inconsistent rankings over the vast majority of districts and can disagree significantly even over which districts to place in the bottom 10 percent of rankings. (See Chapter 2.)

Pildes argues that the role of compactness should be to identify some extreme threshold of bizarreness, where multiple measures of compactness would presumably agree. It has not generally been true that the districts rejected by the Court have met this standard of extreme bizarreness. (See Chapter 3.)

malapportionment for most decadal redistrictings since 1789. What follows in this section is a summary of those findings.

For most of the United State's existence, there were no laws that governed the creation of individual districts. In the early years of the nation's existence, congressional districts were not required at all – states could hold elections for congress at-large, if they wanted. As revealed in contemporary congressional debates, Congress first required districts to benefit political minorities in each state – to counteract the majoritarian bias inherent in at-large elections. This was the primary purpose for districts. Districting criteria were subordinate.

Until 1842, Congress required neither districts nor districting criteria. For most of the period from 1842 until 1919, contiguity was formally required. Before *Wesberry*, however, legal principles guiding the process of congressional redistricting drew little congressional debate,[6] and were never successfully enforced.

Nevertheless, a number of empirical regularities are evident. Historical Congressional districts were, generally, contiguous, composed of entire counties, and

---

[6] In contrast, Congress has more recently repeatedly and vociferously debated formal criteria for apportionment (Young 1988).

moderately malapportioned.[7] By two plausible measures, historical districts were, on average, more compact than modern districts. As is usually the case in history, none of these regularities held absolutely. Even in the early Congresses it is easy to find districts that failed contiguity, split counties, and resembled shoestrings (if not iguanas). Higher levels of malapportionment, splits in county lines, and ill-compactness occurred regularly in postbellum cities, where concentrated populations made it difficult to justify the use of entire counties as building blocks, and redistricters split counties and other political subdivisions. Although some of these bizarre-looking districts corresponded to known historical gerrymanders, many others did not.[8]

The shape of districts changed dramatically following the Court's decision to require strict population equality. Following *Wesberry* and *Reynolds*, the number of districts splitting county lines skyrocketed (See Figure 1-1.), accompanied by decreases in the contiguity and compactness of districts. Minority-controlled districts have been blamed for the decrease in compactness in modern districts (Pildes 1997, 2513), but the

---

[7] Both of the latter two regularities were probably at least partially a consequence of the limits of the census, which in general only made available to redistricters complete data aggregated at the county level.

[8] This contradicts Timothy O'Rourke's contention that modern "bizarre" districts are radical departures from the past, and that bizarre appearance has historically indicated "dysfunctional" districts (O'Rourke 1995, 729, 738).

chronology of changes in district shapes in the U.S. suggests instead that the Court's equal population requirements have been a driving force.

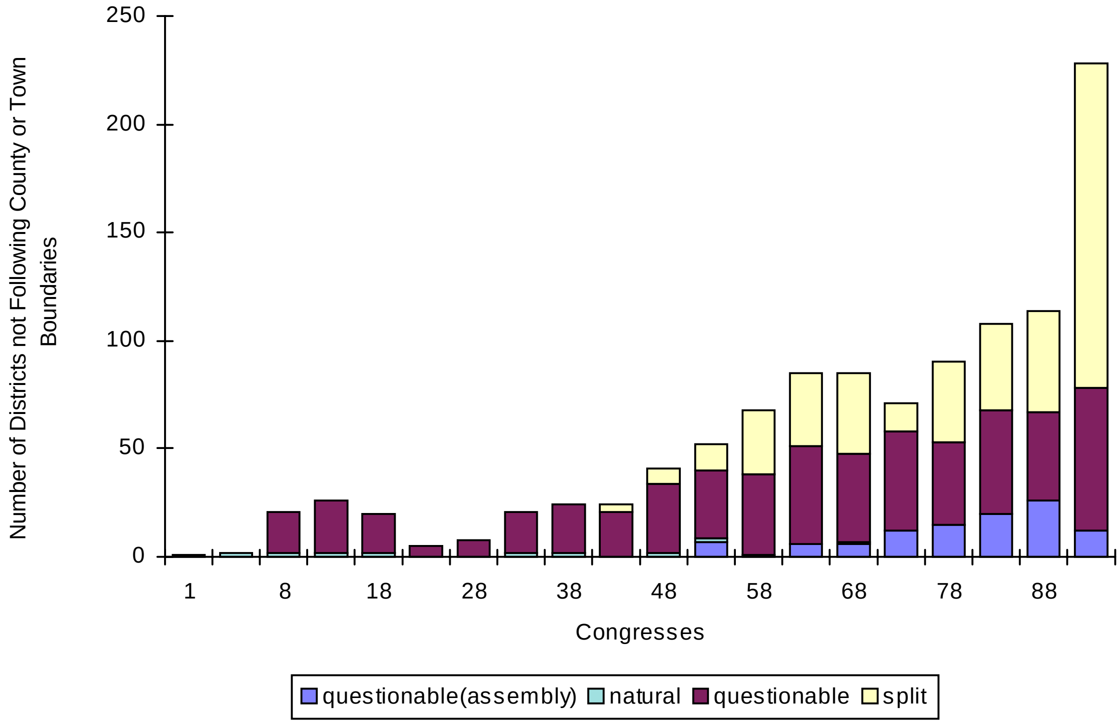


Figure 1-1. Number of districts splitting county or town boundaries by decadal redistricting. 'Questionable' districts split county or town lines to follow ward or assembly ("questionable assembly") boundaries. 'Natural' districts split these boundaries to follow rivers or other large, permanent, natural features. The changes following *Wesberry* are shown in the final column. (See Chapter 3.)

The Court's choice to target minority-majority districts for violating "traditional districting principles," is, in a historic context, doubly ironic. Ironically, one of the culprits responsible for the modern divergence from historical districting principles is the

Court itself -- the most precipitous deviations from historical districting principles were a result of the Court's decisions in *Reynolds* and *Wesberry*. By demanding that apportionment of population be unhistorically equal, the Court weakened the principles of county integrity, compactness, and contiguity. More ironically, the Court has reversed the historical status of representational and geographic principles: Methods of geographic districting were originally adopted so that underrepresented political minorities could have a greater political voice (See Chapter 3.) -- this was the central historical districting principle, and specific geographic principles, such as contiguity, were of lesser importance.

1.3. The Political Effects of Traditional Districting Principles – Much ado about nothing?

Past gerrymanders are the subject of history; current gerrymanders, the subject of politics. Do "traditional districting principles" lead to better electoral outcomes, in and of themselves? Do the "traditional districting principles" limit the damage done by gerrymandering? Do they act as signals warning us of harm?

In this section, I examine the harms with which this Court and past Courts have been concerned in redistricting decisions, and I evaluate the evidence that the use of traditional districting principles can prevent these harms. For more than three decades, the Courts have recognized that redistricting can be used to exclude voters from the political process, or to dilute their vote. In the recent set of redistricting cases, the Court has discovered two additional harms associated with redistricting: expressive harm and racial classification.

1.3.1. Exclusion and Dilution – Direct Harms to Representation

In *Gomillion* v. *Lightfoot* (1960), the Court first acknowledged that the right to vote went beyond the right simply to cast a ballot and have it honestly counted. Black petitioners asserted that an act to redistrict the city of Tuskegee removed substantially all black resident voters and thereby eliminated any meaningful black participation in city elections, and the Court agreed.[9] Shortly thereafter in *Baker* v. *Carr* (1962), *Wesberry* v. *Sanders* (1963), and *Reynolds* v. *Sims* (1964), the Court recognized that constitutional harm did not require effective exclusion, but could result from vote dilution: "And the right of suffrage can be denied by a debasement or dilution of the weight of a citizen's vote just as effectively as by wholly prohibiting the free exercise of the franchise." (Justice Warren, writing for the majority in *Reynolds.*) In all of these cases, the Court recognized that a citizen's right to vote went beyond the right to cast a ballot.

In these cases, the Court recognized both individual and group dimensions of vote dilution. In *Wesberry* and *Reynolds* cases, and many of the succeeding malapportionment cases, the Court treated malapportionment as a harm against the *individual's* right to vote. Even in the early redistricting cases, however, the Court recognized that vote

[9] Recent scholars have attempted to recast *Gomillion* as a racial-classification case. This is incorrect, as I explain below.

dilution could have a group dimension,[10] as shown by Brennan's majority opinion in *Fortson* v. *Dorsey* (1965): "It might well be that, intentionally or otherwise, a multi-member constituency apportionment scheme, under the circumstances of a particular case, would operate *to minimize or cancel out the voting strength of racial or political elements of the voting population*" (85 S.Ct. 501) (emphasis added). The recognition of a group-based dimension of voting dilution becomes clear in *White* v. *Regester* (1973) and in *UJO* v. *Carey* (1977): "But we have entertained (Constitutional) claims that multi-member districts are being used invidiously to cancel out or minimize the voting strength of *racial groups*... The plaintiffs' burden is to produce evidence to support findings that the political processes leading to nomination and election were not equally open to participation by the group in question that its members had *less opportunity than did other residents in the district to participate in the political process and to elect legislators of their choice*" (93 S. Ct. 2339). (Justice White, writing for the majority in *White*, emphasis added.) Far from retreating from this theory of group-based vote dilution over time, the Court expanded the theory of dilution to non-racial groups in *Davis* v. *Bandemer* (1986).

---

[10] Maveety (1991), provides a detailed a history of the expression of group and individual representation rights in early redistricting cases. (See, especially, Chapters 2–4.) Ryden (1997) provides another account. (See, especially Chapters 3 and 7.)

In any particular case, whether a group's vote has been diluted in or excluded from the political process is a matter for detailed empirical inquiry. The literature on statistically assessing minority vote dilution is large and well developed: Grofman, Handley and Niemi (1992) present an excellent summary that need not be repeated here. (Also see King, Bruce and Gelman (1995) for a recently developed sophisticated statistical method of evaluating dilution claims.)

No such literature exists, however, relating traditional (or historical) districting criteria to vote exclusion, vote dilution, or to their prevention. There are many claims in print that traditional districting principles provide a neutral way of limiting gerrymanders, and hence limit the dilution of targeted groups, but little evidence.

The evidence that exists to document the effects of t.d.p.'s shows that the effects of these principles are not as straightforward as such proponents suggest: Preserving county lines, requiring contiguity and requiring geographic compactness clearly can disadvantage dispersed minorities. My statistical analysis of the partisan effects of t.d.p.'s in four decades of congressional redistricting, in Chapter 6, also suggest that these principles have negligible efficacy in preventing partisan gerrymanders.

Furthermore, computer simulations, in Chapter 5, suggest that traditional districting principles, even in ideal circumstances, should not be expected to be neutral. As well as disadvantaging non-compact minorities, compactness standards can disadvantage minority parties that have a geographically concentrated base. Clearly, the Court's enthusiasm for t.d.p.'s cannot be justified by their efficacy against vote dilution.

1.3.2. New, Improved, Theories of Harm?

The *Shaw* line of cases, and their emphasis on t.d.p.'s, cannot be explained using the logic of traditional vote dilution and exclusion cases. To use the Court's own words: "Shaw recognized a claim -analytically distinct- from a vote dilution claim" (from the majority opinion in *Miller*).[11]

This was not, and could not, be a vote dilution case: As in *UJO*, the white population did not suffer vote dilution because "there was no fencing out the white population from participation in the political processes of the county, and the plan did not minimize or unfairly cancel out white voting strength... even if voting in the county occurred strictly according to race, whites would not be underrepresented relative to their share of the population"[12] (97 S.Ct. 1010).

---

[11] Karlan describes the Court as abandoning representation harms, and creating a new harm of "wrongful districting" (Karlan 1996 288,290).

[12] Indeed, it is hard to imagine Justice O'Connor affirming an argument that white voting strength had been *diluted* in *Shaw* after she wrote, in her concurrence in *Bandemer*, arguing that the Court should not recognize vote dilution against "dominant groups." O'Connor emphasized that vote dilution should only be recognized when it affected racial minorities, and then only in the most extreme cases: "As a matter of past history and present reality, there is a direct and immediate relationship between the racial minority's

Two conflicting theories of harm permeate the Supreme Court's recent redistricting cases. In the "expressive-harm" cases, best exemplified by *Vera*, violations of t.d.p.'s are a necessary and integral part of the harm caused by racial gerrymanders. Compliance with traditional districting principles is not merely one piece in a body of circumstantial evidence, but is also a threshold requirement for strict scrutiny. We can avoid strict scrutiny altogether, even if we are motivated by race, if we pay reasonable[13] attention to

---

group voting strength in a particular community and the individual rights of its members to vote and to participate in the political process...Even so, the individual's right is infringed only if the racial minority group can prove that it has essentially been shut out of the political process.'"

It is just as hard to imagine Justices Thomas and Scalia acknowledging that vote dilution had occurred in these cases, when they went so far as to deny its existence in *Holder* (See below in this section.).

[13] The Court does not require that districts be drawn strictly to follow these criteria: "We thus reject, as impossibly stringent, the District Court's view of the narrow tailoring requirement, 'that a district must have the least possible amount of irregularity in shape, making allowances for traditional districting criteria.'" Furthermore, "A § 2 district that is reasonably compact and regular, taking into account traditional districting principles such as maintaining communities of interest and traditional boundaries, may pass strict scrutiny without having to defeat rival compact districts designed by plaintiffs' experts in endless

t.d.p.'s: "For strict scrutiny to apply, the plaintiffs must prove that other, *legitimate districting principles* were -subordinated- to race..." (emphasis added). O'Connor, delivering the judgment of the court, stresses this point repeatedly: "Under our cases, the States retain a flexibility that federal courts enforcing Section 2 lack, both insofar as they *may avoid strict scrutiny altogether* by respecting their own traditional districting principles." Furthermore, she goes on: "We do not hold that any one of these factors is independently sufficient to require strict scrutiny. The Constitution does not mandate regularity of district shape... and *the neglect of traditional districting criteria is merely necessary*, not sufficient" (116 S. Ct. 1953, emphasis added).

For the majority in *Shaw I*, reapportionment was an area where "appearances do matter." In the majority's view, districts that separate people by race, while disregarding political and geographic boundaries, reinforce the perception that members of the same race necessarily share political views. These districts send the "pernicious" message to politicians that they should only represent the majority voting-group in the district. In *Shaw*, violation of "traditional districting principles" is an integral part of the harm perceived by the Court — violation of these principles *actively* causes harm by sending a pernicious message to politicians and to voters.

Although not explicitly named in *Shaw I*, in *Vera*, the court acknowledges that it is relying on a new type of harm: "we also know that the nature of the *expressive harms*

'beauty contests. '"

with which we are dealing." (emphasis added)[14] Pildes & Niemi outline (or define) this theory in "Expressive Harms, 'Bizarre Districts,' and Voting Rights: Evaluating Election-District Appearances After *Shaw,"* (Pildes and Niemi 1993). As Pildes & Niemi write in their much-cited article: "Expressive harms focus on social perceptions, public understandings, and messages; they involve the government's symbolic endorsement of certain values in ways not obviously tied to any discrete, individualized wrongs." In this theory, ugly districts[15] send a symbolic message — without ugliness, there is no such message.

The opinions in *Miller*, *Shaw II*, and *Abrams* parallel those in *Shaw I* and *Vera* in denouncing ugly districts, but differ from them in reasoning. Unlike in *Shaw I* and *Vera*,

---

[14] It is difficult to determine to what extent the Court formulated a theory of expressive harm in *Shaw I* and to what extent it took the theory from Pildes and Niemi's article itself. In any case, Justice O'Connor, in *Vera*, both frequently cites Pildes and Niemi's article on other topics and adopts their expressive harm terminology when she discusses harm; so it is clear that the theory of expressive harm has now become part of this jurisprudence.

[15] Presumably, other violations of traditional districting principles send the same sort of symbolic message. Although this presumption seems unlikely that not having "corner districts" or a black majority district in Atlanta, the Court gives no separate theory of harm for t.d.p.'s other than compactness.

"traditional districting principles" are no longer an integral part of the harm caused by redistricting. Instead, violations of these principles act as circumstantial evidence of racial intent. For example, in *Shaw II*, which was delivered on the same day as *Vera,* Rehnquist disavows the role of t.d.p.'s that O'Connor affirms. Implicitly rejecting the view that traditional districting principles are relevant because bizarreness is a necessary element of the constitutional wrong or a threshold requirement of proof, Rehnquist states bizarre district lines "may constitute persuasive *circumstantial evidence* that race for its own sake, and not other districting principles, was the legislature's dominant and controlling rationale in drawing its district lines."[16]

In contrast to *Shaw I* and *Vera*, the operating principle in *Miller* and in *Shaw II* seems to be not expressive harm but racial classification. Two authors attempt to explain the Court's logic: In "Affirmative Racial Gerrymandering: Fair Representation for Minorities or a Dangerous Recognition of Group Rights?" Katharine Inglis Butler (1995) argues that the court is applying a principle banning racial-classification. James Blumstein makes a similar argument in "Racial Gerrymandering and Vote Dilution:

---

[16] Nor can adherence to these principles necessarily defeat a claim of racial gerrymandering, as the Court states that bizarreness is not necessary to raise issues of equal protection: "Our observation in *Shaw* of the consequences of racial stereotyping was not meant to suggest that a district must be bizarre on its face before there is a Constitutional violation..."

Shaw v. *Reno* in Doctrinal Context"[17] (1995). In essence, they both argue that under the 15th amendment voters have a right not to be subjected to racial classifications. Voters in this case are not harmed by any electoral or legislative outcome, but by the government's act of classification.

Both Blumstein and Butler view this theory of racial classification as, in Butler's words, "established constitutional doctrine." Each has a somewhat different story about how the doctrine was established.[18] Butler claims that this principle comes originally from *Gomillion* and to some extent from *Brown* v. *Board of Education* (1954). Blumstein

---

[17] These articles were written after *Shaw*, but prior to *Miller*. Following Miller, Blumstein continued to hold that: "From an analytical perspective, Shaw and Miller are not voting cases, but suspect classification cases. They are not wrongful districting cases, they are racial gerrymander cases - racial classification cases" (Blumstein 1996, 505). Butler, as well, holds to the racial classification theory in analyzing *Miller*: "Racial classification was the gravamen of the complaint, the Court said. Bizarre shape is merely one means to determine that race was the basis of the districting plan" (Butler 1996, 216).

[18] Note that, although Blumstein and Butler attempt to account for the genesis of the racial classification standard the origin of the "predominant factor" requirement remains a mystery. Isacharoff (1996), Karlan (1996) and Kousser (1998, ch. 8) dissect this requirement, and so I shall not elaborate upon it here.

traces this principle from *Brown* and more recently from *Northern Florida Chapter of the Associated General Contractors of America* v. *City of Jacksonville* (1993).

Under *Miller* and *Shaw II*, redistricting plans are subject to strict scrutiny when race is the predominant motive in their creation. Under *Vera*, subordination of traditional districting criteria is necessary to trigger strict scrutiny.

Under a racial-classification standard of harm, violation of “traditional districting principles” can act, at most, to overcome a defense that a plan is narrowly tailored for a compelling state interest, or as circumstantial evidence of an impermissible racial classification. (See Butler, section H, and Blumstein 4A, respectively.) But both *Shaw I* and *Vera* treat traditional districting principles as more than mere circumstantial evidence, and both refer not to the government’s act of classification, but to the message sent to voters and representatives, as shown by these passages from the majority’s opinion in *Shaw I* and judgment in *Vera* (respectively): “Significant deviations from traditional districting principles, such as the bizarre shape and noncompactness demonstrated by the districts here, cause Constitutional harm *insofar as they convey the message* that political identity is, or should be, predominantly racial” (116 S.Ct. 1962, emphasis added). “For example, the bizarre shaping of Districts 18 and 29, cutting across pre-existing precinct lines and other natural or traditional divisions, *is not merely evidentially significant; it is part of the Constitutional problem”* (116 S.Ct. 1962, emphasis added).

1.3.3. Theories in Search of Evidence

Is harm needed for racial classification cases? Proponents of the “racial classification” theory of harm wish to argue that no evidence of harm is needed in such cases — beyond the evidence of a classification itself. Despite this avowed preference for metaphysical harm, an examination of the precedents which are supposed to support the racial classification theory gives the lie to such arguments — these precedents are based on demonstrated actual harm, not a metaphysical classification.

Proponents of these theories of harm lay claim, of course, to legal precedent. Blumstein and Butler recast *Gomillion* as a case of racial classification. In addition, Blumstein attempts to reinterpret *Brown* v. *Board*.[19]

---

[19] Butler also traces the lineage of *Shaw I* to *City of Jacksonville*, a case hardly older than *Shaw* itself. Even in *City of Jacksonville*, however, in this case the Court does not recognize a harm based upon pure racial classification — instead the Court recognizes a harm to an outcome, the ability to compete: “the injury in fact is the inability to compete on an equal footing.” Evaluating equal opportunity is more difficult in the political arena, because both the action and outcomes of voting have group dimensions — but the principle that each person should have an *equal opportunity* to participate in the electoral process is nonetheless central to the entire line of minority vote dilution cases. In distinguishing itself from these cases, however, *Shaw* distinguishes its theory of harm from *City of Jacksonville*.

Butler correctly stated that while the black voters removed from Tuskegee were deprived of the municipal vote, “so was every other person who resided outside the city boundaries.” She then concluded that the basis of their claim could not have been a right to vote in municipal elections. Finally, she concluded that the harm was one of classification or segregation.

This argument, however, belongs not to the majority opinion, but to Justice Whittaker’s concurrence. Unlike Butler, Justice Whittaker recognized the Courts reasoning, but disagreed with it: “It seems to me that the decision should not be rested on the Fifteenth Amendment... inasmuch as no one has the right to vote in a political division.” His conclusion is the same as hers “‘fencing Negro citizens out of’ Division A and into Division B is an unlawful segregation of citizens by race in violation of the Equal Protection Clause” (81 S.Ct. 131-2). The Tuskegee case, however, was not one of racial sorting, as Butler claims, but a real case of segregation, as the city was nearly all-white.

The reasoning used by Frankfurter, for the majority in 1960, bears little resemblance to the reasoning in *Miller* and other modern racial-classification cases. *Gomillion*’s operating principle was not an abstract racial classification but the denial of the effective right to vote: “such (legislative) power, extensive though it is, is met and overcome by the Fifteenth Amendment to the Constitution of the United States, which forbids a State from passing any law which *deprives a citizen of his vote* because of his race” (81 S.Ct. 129, emphasis added).

Contrary to Whittaker's and Butler's argument, the majority decision was not based on the belief that everyone in Alabama had an unconditional right to vote in Tuskegee. Instead, *Gomillion* recognized the principle that voters could be harmed, not just by removing the ballot, but by reducing its effectiveness. Furthermore, it was firmly grounded in clear and concrete evidence of just such a harm: "The essential inevitable effect of this redefinition of Tuskeegee's boundaries is to remove from the city all save four or five of its 400 Negro voters while not removing a single white voter or residence."

Like *Gomillion*, *Brown* relied not on an appeal to symbolism, but was firmly rooted in evidence of psychological and educational harm.[20] *Brown* did not raise the symbolism of classification to a harm in itself, but instead declared that the *result* of separate schools was inequality: "Segregation of white and colored children has a detrimental effect upon the colored children. The impact is greater when it has the sanction of the law... Separate educational facilities are inherently unequal." Both the plaintiff's arguments in *Brown* and the Court's decision were based, rightly or wrongly, on the Court's view of the *impact* of segregation. In the cases leading up to *Brown*, in *Brown* itself, and in the cases immediately following, plaintiffs presented, and courts weighed, evidence that

---

[20] And as Karlan and Levinson point out, taking race into account in redistricting bears a close resemblance to the race conscious pupil assignment used by the courts to dismantle segregated schools, following *Brown* (Karlan and Levinson 1996).

segregated education was neither materially nor psychologically equal. (Lively , 1992: Ch. 4,5)

Racial classification can have enormous practical consequences, but in this current line of cases the Court has again embraced symbolism over substance. Rather than evaluating the political effects of racial classification, in reality they are expressing a distaste for politics *tainted* by race.[21]

---

[21] This abhorrence for race-taint is most clear in the Court's treatment of race-as-proxy, in *Vera*, where the Court announced that strict scrutiny was to be applied to any use of racial variables in redistricting, even if these variables were used only for partisan purposes: "But to the extent that race is used as a proxy for political characteristics, a racial stereotype requiring strict scrutiny is in operation. Cf. *Powers* v. *Ohio,* 499 U. S. 400, 410 (1991) (-Race cannot be a proxy for determining juror bias or competence)" (116 S. Ct., 1956).

In censuring the use of race-as-a proxy in districting, the Court relied on *Powers* v. *Ohio* (1991), which declared that race cannot be a proxy for determining juror bias or competence. Although the principle stated is the same in both cases, the nature of the classification has changed, because the context of redistricting is fundamentally different from the context of a jury trial.

In *Powers*, the court claims that the use of race to classify to qualify or select jurors harms their dignity and the integrity of the courts. In redistricting, unlike in jury trials, the use of race as a proxy for partisan voting is distanced both from individuals and from the courts, and hence harms neither.

In jury trials, the *Powers* opinion holds, the Court forbids the use of race as a proxy to guard against discriminatory purposes, and especially to guard against subverting earlier neutral procedures in the trial: “Racial discrimination in the selection of jurors in the context of an individual trial violates these same prohibitions. A State “may not draw up its jury lists pursuant to neutral procedures but then resort to discrimination at `other stages in the selection process.' ““ In *Vera*, the Court specifies that the use of race is forbidden only when it is predominant over other purpose, yet the use of race *as a proxy* is absolutely forbidden. If the ban against race as a proxy is meant as prophylaxis, why would it apply when the use of race is not itself forbidden?

Most important, in *Powers*, the Court’s central argument is that race cannot be used as a proxy for determining juror bias or competence because the relationship is a stereotype only; in reality race has nothing to do with juror competence or bias: *“A person’s race simply ‘is unrelated to his fitness as a juror.’***”** (emphasis added). In redistricting, however, the situation is entirely opposite: race is used as a proxy for predicting partisan votes precisely where race and politics are in reality most closely

What, exactly, causes "expressive harms," and how? Advocates of "expressive harm" argue that vote dilution is not the only meaningful harm that can follow from redistricting, and that redistricting can cause harm by altering the perceptions of voters and representatives. If redistricting does cause changes in perceptions, we should be able to find evidence of this in expressed opinions and behavior.

If district appearance sends messages to voters and representatives, how are these messages transmitted, and how can we detect their effects? How do voters receive this message, when most voters are certainly unaware of the shape of their districts? Why must voters be in bizarre-shaped districts to receive this message?[22]

---

related. If the relationship were merely a stereotype, there would be no reason for partisan gerrymanders to use racial variables in their design.

None of the arguments that are used to ban race as a proxy from jury cases succeeds in redistricting cases. As with harms of classification, all requirements that there be likely harmful consequences are dispensed with. Gone are the severe, direct, individuated, fundamental harms that occur in jury cases. Gone are any requirement that plaintiffs show that they have been representationally injured. The only identifiable harm left is the symbolic taint of the classification itself.

[22] The Court's treatment of standing further muddies this issue. The Court has declared the *Shaw* line of cases distinct from vote dilution and exclusion cases; in line with

this distinction, the rules of standing are much simplified. As the Court declared in *Hays* and reaffirmed in *Vera*, individuals inside a racially gerrymandered, majority-minority district have the right to challenge those district lines; those outside do not: "(w)here a plaintiff resides in a racially gerrymandered district, the plaintiff has been denied equal treatment because of the legislature's reliance on racial criteria, and therefore has standing to challenge the legislature's actions" (115 S. Ct. 2436).

Standing and harm are supposed to be connected intimately. In *Lujan* v. *Defenders of Wildlife* (1992), upon which the Court relies in *Hays*, the Court lists three elements of "the irreducible constitutional minimum of standing," the first of which is that "the plaintiff must have suffered an 'injury in fact' -- an invasion of a legally protected interest which is (a) concrete and particularized, and (b) actual or imminent, not conjectural or hypothetical." For an argument against the court's current approach to standing see Sunstein (1993).

Yet the clear and simple rule of standing announced by the Court appears not at all connected with the opaque and complicated theories of harm that it uses in its decisions. On their face, expressive harms are public harms — they are messages sent by the government and perceived by the voting public. Why do only those voters in majority-minority districts have standing? If expressive harm stems from *symbolic* government action, all voters should have standing — the government symbolically represents all voters when it endorses the use of race in the political process, even if such endorsement

The Court claims that bizarre districts lead to the balkanization of the electorate — causing voters to polarize along racial lines. Where is the evidence of this? What else serves as evidence that district shape harms voters? Why would representatives have an especially strong reaction to district shape — when as professional politicians they already know the extent to which redistricting has manipulated electoral outcomes? How can we detect whether representatives have changed their behavior because of bizarre district shape? Suppose that we have evidence both that a redistricting plan causes psychological harm, and prevents the political harm of vote dilution — how do we weigh these two harms?

---

only results in one district where race has been predominant, or in no minority opportunity districts at all. If expressive harm stems from *actual perception* of government action, there is no justification for the assumption that voters and politicians are unable to perceive the racial motivation of a government action simply because they were not placed in a majority-minority district. Others have commented upon the abandonment of doctrines of standing in these cases: see Kousser (1995) at 640-642, Issacharoff and Goldstein (Issacharoff and Goldstein 1996), and Karlan (1994) at 278-9. Pildes & Niemi (1993) themselves recognize a fundamental tension between expressive harm and individualized theories of standing which the Court has still to confront.

Neither these authors nor the Court have formulated an explanation of individuated expressive harms. Pildes (1997, 2568) and Ely (1197, 590) take the opposite tack, and they argue that every voter in the state should have had standing in these cases.

Unlike televised flag burning, for example, redistricting is technical, low profile, and lacks drama. If expressive harms are rooted in symbolism, why is redistricting of particular symbolic importance? Does harm occur whenever voters believe a district was racially motivated -- even, as Pildes and Niemi (Pildes and Niemi 1993, 193)argue, if these beliefs are wrong? If we measure expressive harm by the offense of actual voters, we are led to the perverse conclusion that “ignorance is bliss” for all concerned.

If district shapes are harmful because they have a pernicious effect on legislators, causing them to single-mindedly represent only the majority coalition in their district - what evidence is there of this behavior? Would Justice O’Connor be surprised to learn that much work in political science has long been based on the assumption that *all* representatives act primarily with a view to pleasing the majority of their constituents, and hence getting reelected?[23] (Downs 1957; Mayhew 1971)

If some individuals have been placed in a racially gerrymandered, majority-minority district because of race, many others have been excluded from that district, and hence placed in adjoining districts for the very same reason. Theories of racial classification raise equally troubling questions. Why do majority-minority districts engender harmful “racial classification,” when adjoining white districts do not? Are the subjects of

---

[23] In addition, as Kousser (1995) says, the argument that irregular districts cause representatives to over-represent minority interests contradicts the argument that such interests are illusory stereotypes. The Court cannot have it both ways.

classification individuals or groups? If individuals are subject to classification, how are individuals classified when the state uses only aggregate data? If instead the harm is that a group is subjected to classification, does not this classification occur as soon as racial data is introduced into the redistricting process — whether or not that data is used to draw any minority district? Under the Court's current theory of racial classification, and until the Court requires concrete evidence of harm, there seems to be no reason to exclude anyone from having standing, when a state uses any racial data at any point in the districting process.

Why does the Court's decisions raise so many questions? Here, at least, the answer is straightforward: Lacking a consistent theoretical framework, historical precedent, and any evidence of political harm, the Court decisions are necessarily unanchored -- bound to raise more questions than they answer.

Where is the evidence of harm in these cases? Both the racial classification theory and the expressive harm theory should logically expand the types of evidence relevant to redistricting cases. Before the use of these theories, plaintiffs could not win a case without evidence of vote dilution; now, the court recognizes harms from racial classification and from pernicious messages as well. Logically, we should expect to see plaintiffs present to the court a broad variety of evidence for such harms. Instead, we see that plaintiffs have not even bothered to refute contrary evidence offered by the defendants. Is there any evidence that redistricting causes such harms?

Although the research community has only recently begun to search for connections between redistricting and harms other than vote-dilution, current research suggests that district shape does not have a large effect on the behavior of elected representatives. Both the classic research of political science (Fenno 1979; Mayhew 1971) and current research (Cameron, Epstein and O'Halloran 1996; Lublin 1997) show that the political behavior of representatives is strongly determined by the constituents of the district. This leaves comparatively little room for bizarre shape, or other factors, in general, to have an intrinsic effect on the behavior of representatives.

In these specific cases, plaintiffs provide little evidence to suggest that the redistrictings in question cause psychological damage, send pernicious messages, or have caused any measurable harm due to racial classification; nor does the Court discuss such evidence. In fact, as Kousser points out, the evidence suggests that minority opportunity districts in North Carolina and Texas reduced racial polarization in voting (Kousser 1995, 648).

In Chapter 6, “Do Traditional Districting Principles Matter,” I attempt to directly measure the effects of district shape, as well as other t.d.p.’s on voting behavior. I use district shape, election, and opinion poll data for four separate elections spanning three decades. To my knowledge, this is the first study of its kind.

I find little evidence to support the theory that bizarre districts cause psychological harms: There is little evidence that the compactness of districts affects electoral outcomes. There is no evidence that compactness affects voters' trust in government, or their opinions about their representative.

I do find evidence that compactness may affect turnout: This bears further investigation, but may be simply a result of incumbent gerrymandering. Moreover, even if ill-compactness directly causes turnout to decrease, this effect is hardly so dramatic to trump all claims of vote dilution.

I also find evidence that compactness can affect the voting behavior of the representative. The effect is weak, but is significant for Republican representatives. Some forms of ill-compactness, are, as opponents argue associated with more extreme voting behavior, but other forms of compactness have just the opposite effect. So it is imperative for the court rely on evidence and analysis of effects, rather than simple assumptions that ill-compact districts cause extreme behavior.

It is still imaginable that redistricting does result in some previously undiscovered harm, and empirical investigation into this issue continues, as it should. Unless and until we discover evidence that these harms are pervasive, the Court should, at least, demand evidence of harm from plaintiffs.

1.4. Lessons from Political Philosophy – Nothing Comes from Nothing

Whether we should regard the effects of redistricting as justiciable harms, or simply as the rough and tumble of politics depends, in large part, on our theories of political representation. Similarly, in the history of redistricting cases, it has been difficult to separate explanations of why gerrymanders should be prevented from theories of political philosophy: If the 14th amendment guarantees "fair and effective" representation, how can we protect this guarantee without a theory of representation? If history informs us that districts have displayed regularity in shape and line, how are we to decide whether this regularity is incidental or integral to the purpose of redistricting? If an analysis of politics suggest that rules for redistricting help or hinder groups, how are we to decide if such effects represent illegitimate bias in the process, or simply the spoils of victory? Justice Frankfurter expressed this point eloquently, in his dissent in *Baker* v. *Carr* (369 U. S. 186, 300):

> Talk of "debasement" or "dilution" is circular talk. One cannot speak of "debasement" or "dilution" of the value of a vote until there is first defined a standard of reference as to what a vote should be worth. What is actually asked of the Court is to choose among competing bases of representation — ultimately, really, among competing theories of political philosophy.

In *Colgrove,* preceding *Baker*, Justice Frankfurter had prescribed a cure consistent with the malady: "The remedy for unfairness in districting is to secure State legislatures that will apportion properly, or to invoke the ample powers of Congress."

The majorities in *Shaw*, *Vera* and *Miller* are not willing to leave issues of fairness to Congress and the states, but neither are they prepared to accept what political philosophers have had to say about representation. Instead, they invent their own theories of harm.

In their approach to political theory, Justices Thomas and Scalia are most extreme, and least consistent — denying the relevance of political theory while espousing a point of view based solely upon it. In *Holder* v. *Hall* (1994), declaring "Stare decisis is not an inexorable command," they propose discarding judicial precedent and Congress's choice of political theory, while pretending to hold no political theory at all. They urge the Court to abjure all consideration of vote dilution because, in their view, like Frankfurter's, any decisions in this area are inherently matters of political philosophy, and thus unfit for judicial consideration:

> In short, there are undoubtedly an infinite number of theories of effective suffrage, representation, and the proper apportionment of political power in a representative democracy that could be drawn upon to answer the questions posed in *Allen*. See generally, Pitkin, [*The Meaning of Representation*], supra. I do not pretend to have provided the most sophisticated account of the various possibilities; but such matters of political theory are beyond the ordinary sphere of federal judges. And that is precisely the point. The matters the Court has set out to resolve in vote dilution cases are questions of political philosophy, not questions of law. As such, they are not readily subjected to any judicially manageable standards that can guide courts in attempting to select between competing theories.

Unlike Frankfurter, who urged, consistent with his abjuration of political theory, that the Courts leave the standard of fairness up to the legislature. Thomas and Scalia urge that Congressional will, as embodied in the V.R.A.,[24] be overthrown, and wish to substitute for it their own theory — the theory that voting equality is necessarily defined solely by the act of casting a ballot and having it counted. (Guinier 1994)[25] These modern justices do not abjure theory, but abuse it in two fundamental ways. First, they mistake their own political preferences for agnosticism. Second, they conflate the absence of a single theory of representation with the presence of infinite theories of representation. While Thomas and Scalia misuse political theory, the Court in *Vera* and in *Shaw I* merely ignore it by confusing substantive political representation with symbolic, non-political representation.

Do questions of vote dilution require the Court to choose from among an "infinite" number of theories of political philosophy? Are the Court's other decisions about voting

---

[24] Parker (1990) provides a thorough discussion of the history of the Voting Rights Act, the circumstances that led up to it and the intent of Congress in passing it and its amendments. His conclusions are quite contrary to those of Thomas and Scalia.

[25] Guinier argues correctly, in my view, that Thomas's opinion in *Holder* is theory-laden. She then proceeds to argue that representation is inherently group-based. For the purposes of my argument, however, the question of whether redistricting is inherently or exclusively group-based need not be settled.

rights and redistricting free from such heady philosophical considerations? Does modern or historical political philosophy offer any guidance?

Zagarri, whose book, *The Politics of Size* (Zagarri 1987), is an in-depth analysis of the development of representation in the United States, offers insight into the theories of representation embodied in the Constitution: "The Constitution embodied the principles of both corporate and proportional apportionment, spatial and demographic theories of representation."[26] (149) Although there was a tension between the Anti-Federalists, who believed in representation based on geographic communities, and the Federalists who believed in representation on the basis of population, both groups shared some ideas. Both groups rejected the British, Burkean, idea of virtual representation, the idea that representation was purely a matter of having one's interests looked after (18).

Moreover, both groups agreed that representation required control over the composition and decisions of the legislature. According to Zagarri (39), advocates of corporate representation argued, to use John Adams' words that the legislature should be

[26] Butler's claim, however, that American states should not provide interest-group representation because states retain a geographic system (Butler 1996, 360) turns history on its head. States have a tradition of geographic *because*, historically, interest-groups were geographically based. It is the representation of interest-groups, not geography for its own sake, that is at the heart of the American tradition of geographic representation.

"a miniature and exact portrait of the people at large. It should think, feel, reason and act like them." (Adams 1776)

Noted Federalists have also recorded their view of representation. Hamilton's and Madison's writings in the *Federalist Papers* demonstrate that the idea that representatives would and should be guided by their constituents' opinions and interests was current at the time of the founding of the republic. Hamilton writes in Federalist 35:

> Is it not natural that a man who is a candidate for the favor of the people, and who is dependent on the suffrages of his fellow-citizens for the continuance of his public honors, should take care to inform himself of their dispositions and inclinations, and should be willing to allow them their proper degree of influence upon his conduct? This dependence, and the necessity of being bound himself, and his posterity, by the laws to which he gives his assent, are the true, and they are the strong chords of sympathy between the representative and the constituent.

And in Federalist 78: "It is not otherwise to be supposed, that the Constitution could intend to enable the representatives of the people to substitute their WILL to that of their constituents."

Historically, the Constitution does not embody one single theory of representation — but it does embody the concept that substantive representation is paramount. Voting is not enough. Neither is having one's interests "looked out for." Representation requires that voters be able to shape the legislature and its decisions.

Political philosophy offers us guidance as well. Taking the historical strands of representation theory and weaving them together, Hannah Pitkin, upon whose work Thomas and Scalia claim to rely, demonstrates that there are many different definitions of political representation. Thomas and Scalia abuse this fact when they conclude from it that all definitions are equal, and hence none are valuable. To the contrary, Pitkin identifies and develops, a core theory of representation that encompasses its many meanings. This core is widely recognized today.

Pitkin explains that different theories of representation have different implications for the types of actions that are valuable (and conversely the types of harm that are possible), and vice-versa. Under *authorization* views of representation, like Hobbes's, representation simply means that the representative's decisions are binding upon the represented. (Pitkin 1967, Chapter 3) Under this theory, access to the ballot is irrelevant because as long as the laws that legislators make are binding upon all people, all people are represented. There can be no exclusion from the political process, because such a harm simply does not exist. Other harms to representation, such as malapportionment and minority vote dilution, are only cognizable under certain theories of representation. If, for example, we instead adopt Burke's theory of *virtual representation*, voters can be excluded from the political process, but malapportionment is irrelevant (Pitkin, Chapter 4).[27]

In fact, not all theories of representation deserve equal treatment — there is a range of theories that are supported by political thought at the time of the creation of the Constitution, by modern political scholarship, and by past precedent. As we have seen

---

[27] Cain (1990) extends this point, to *individualist*-formalist definitions of representation. If we adopt such definitions, voters are represented equally if and only if all voters are treated uniformly and each vote bears equal weight in determining the results of the election — under this theory malapportionment does affect equality, but minority vote dilution cannot, by definition, exist. Cain uses the term "formalistic" rather than "individualist-formalist" to refer to these types of theories. I use the latter term for two reasons. First, I use it to distinguish these theories from Pitkin's "formalistic" theories — Cain's "formalism" refers to the use of formal mathematics to describe representation, whereas Pitkin's formalism refers to the substance of the relationship between representative and electorate. Second, while I agree with Cain's point that minority vote dilution is excluded by some mathematical theories of representation, I think it is important to note group theories of representation are also quite compatible with mathematical formalism, as demonstrated by group-based formal measures of representation such as "power indices." (See Ordeshook 1986 , Section 10.6 for an introduction.)

above, Thomas and Scalia throw up their hands at what they claim is an “infinite number” of theories of representation, citing Pitkin’s *The Concept of Representation* as evidence of the diversity of theories. In doing so, they not only ignore the particular theories of representation embodied in the Constitution, but also miss Pitkin’s central argument, which is that the fundamental principles of political representation can be identified, within bounds: (Pitkin 1967, 209)

> (Political) representation here means acting in the interest of the represented, in a manner responsive to them. The representative must act independently; his action must involve discretion and judgment; he must be the one who acts. The represented must also be (conceived as) capable of independent action and judgment, not merely being taken care of. And, despite the resulting potential for conflict between representative and represented about what is to be done, that conflict must not normally take place. The representative must act in such a way that there is no conflict, or if it occurs an explanation is called for. He must not be found persistently at odds with the wishes of the represented without good reasons in terms of their interests, without a good explanation of why their wishes are not in accord with their interests.

Modern democratic theorists have come to a similar conclusion — democratic representation is not satisfied merely by voting, but requires substantive control of leaders by voters. In the conclusion of a comprehensive survey of theories of democracy, David Held summarizes the requirements for a system to be fully democratic. In essence, democracy requires that all competent adult members be able to understand and participate in the collective decision making process, and that they be able to control both the subjects and the substantive outcome of that process. (Held 1987)

The preeminent modern democratic theorist Robert Dahl expressed the same idea succinctly, if less precisely: "democratic theory is concerned with processes by which ordinary citizens exert a relatively high degree of control over leaders;" (Dahl 1956, 3) More recently, Dahl made the point that that representative government is not merely procedural (Dahl 1989, 175), but substantive:

> Nor is the right to self-government a right to a "merely formal process," for the democratic process is neither "merely process" nor "merely formal." The democratic process is not "merely process," because it is also an important kind of distributive justice.

Rather than misreading Pitkin, the Court should perhaps read Dahl; in its rush to eliminate race from the *procedures* of representation, it has neglected the substance of representation.

Thomas and Scalia's complaint of an infinite number of representational theories, is, to borrow Frankfurter's words in *Colgrove*: "A hypothetical claim resting on abstract assumptions... made the basis for affording illusory relief for a particular evil even though it foreshadows deeper and more pervasive difficulties in consequence." Although writers on representation have many differences, there is a core to representation theory. First, political representation requires that constituents be able to exercise control over their representatives — the Court's complaint that representatives might pay *too* much attention to the constituents in their district is misguided. Second, political representation is not primarily concerned with symbols, but with outcomes; in the context of modern and constitutional political theory alike, the things that matter most are who gets elected and what the legislature does.

1.5. Sound Districting Principles– History, Politics and Philosophy

The Court, in *Shaw*, starts in its ill-starred course by transmuting the "sound" principles in *UJO* to the "traditional" principles of *Shaw*. Whether the principles used by the Court are traditional is a matter of history, and whether they are sound is a matter of reason and politics. Neither history nor politics supports the Court's current use of t.d.p.'s.

Historically, when Congress first required each state to draw district lines for the House of Representatives, its purpose was to expand representation by moderating the majoritarian bias that was a result of widespread at-large elections. (See Section 2.3, also Chapter 3.) Sound districting principles stem from the historical and theoretical purpose of redistricting — representation.

There are a large number of candidates for sound redistricting principles (See Lijphart (1989) for a survey.), although few are universally accepted. What is important is that these principles be judged against the benchmark of representational values.

Court cases before *Shaw* recognized the fundamental representational purpose of redistricting, and the role of sound principles. In contrast to the current Court's use of "traditional districting principles," sound redistricting principles were valued insofar as they contributed to political fairness. This principle of fairness is aptly summarized in *Geffen* v. *Cummings* (1973), where the Court allowed deviations in population that contributed to a more balanced partisan districting plan: "The very essence of districting is to produce a different —a more 'politically fair'— result than would be reached with elections at large..." (93 S. Ct. 2321 at 2329). This principle remains, even in much later cases -- in *Davis* v. *Bandemer (1986)*, which explicitly addressed partisan gerrymandering, the Court emphasized this once again: "The very essence of districting is to produce a different — a more 'politically fair' — result than would be reached with elections at large, in which the winning party would take 100% of the legislative seats" (106 S. Ct. 2797, at 2808).

Historically, the court has recognized that legislatures are best suited to manage the political process of redistricting and its representational implications. (Kilgore 1997, 1306-7) The freedom to play political "games" is intertwined with the responsibility to manage this process. The Court's role before *Shaw* has been, and should be, to ensure that minority players are not excluded from playing, and that the athletic field is not too uneven.

In fact, remnants of the purpose of districting remain in *Vera*. In the opinion, O'Connor recognizes that states have legitimate interests in districting in order to change partisan balance, support communities of interest, protect incumbents, create representational proportionality, and remedy discrimination:

- "We have recognized incumbency protection, at least in the limited form of avoiding contests between incumbent(s), as a legitimate state goal" (116 S. Ct. 1954).
- "For example, a finding by a district court that district lines were drawn in part on the basis of evidence (other than racial data) of where communities of interest existed might weaken a plaintiff's claim that race predominated in the drawing of district lines Cf.post, at 6 (Souter, J., dissenting) (recognizing the legitimate role of communities of interest in our system of representative democracy)" (116 S. Ct. 1954)
- " A State's interest in remedying discrimination is compelling when two conditions are satisfied. First, the discrimination that the State seeks to remedy must be specific, identified discrimination; second, the State must have had a 'strong basis in evidence' to conclude that remedial action was necessary, 'before it embarks'" (116 S. Ct. 1962).

Justice O'Connor states that redistricting is not simply an exercise in following mechanistic principles – it is not a " beauty contest."[28] The judgment in *Vera* explicitly states that district planners do not have to maximize compactness, even for the narrow

---

[28] This distinguishes the role of redistricting principles in *Vera* from that advocated by some proponents of these principles (Polsby and Popper 1991; Stern 1974), who have argued that districts should be created by maximizing formal criteria such as compactness, districts.

tailoring requirement of strict scrutiny — there is still room for representational purposes in creating districts: "...Rather, we adhere to our longstanding recognition of the importance in our federal system of each State's sovereign interest in implementing its redistricting plan" (116 S. Ct. 1960).

This deference to the State's sovereign rings hollow, however, when one considers that the Court has put traditional districting principles and non-representational harms above representational values. History, philosophy, and politics provide three lessons on redistricting. The lesson of history is that redistricting is *traditionally* concerned with representation; other historical features of districts were of secondary importance. The lesson of philosophy is that representation is, at heart, about the rights of constituents to choose representatives and legislative policy — within the context of redistricting, "expressive" harms and racial classifications should be important only where they intersect these rights. Finally, the lesson of politics is that, to the best of our knowledge, violations of traditional districting principles are essentially harmless.

Chapter 2. The Consistency and Effectiveness of Mandatory District Compactness Rules

2.1. Gerrymandering And District Appearance

> The regular distribution of power into distinct departments; the introduction of legislative balances and checks; the institution of courts composed of judges holding their offices during good behavior; the representation of the people in the legislature by deputies of their own election:...
>
> They are means, and powerful means, by which the excellences of republican government may be retained and its imperfections lessened or avoided.
>
> \- *Federalist 9*, Hamilton

The rules that we use to choose representatives lie at the heart of government. In the United States, some of the most controversial of these rules govern electoral districting. Some legal scholars have claimed that gerrymandering can be virtually eliminated by requiring that districts be geographically “compact.” In recent cases, the Supreme court has evidently agreed.

Most proponents of compactness standards explicitly offer them as prophylactics against gerrymandering. Mathematical functions that describe the regularity of a district’s geography or population distribution can be as simple as a measure of the length of a district’s borders, or as complicated as a calculation of the population-weighted moment of inertia for the district. Many scholars have proposed ways to measure compactness — the literature contains more than thirty different measures — but few have systematically analyzed these measures. What, exactly, are different compactness criteria measuring? To what extent do different measures of compactness agree?

In this chapter, I use an axiomatic analysis to join formal measures of compactness with common intuitions about how gerrymandering is carried out. I find that, contrary to the claims of some previous researchers, it is impossible for a single index to capture all recognized forms of geographic manipulations. I then develop a methodology that uses small-case analysis to quantify the agreement among different measures of compactness and to quantify how strictly each measure limits geographic manipulation. Last, I reevaluate some of the previous empirical research on district compactness. I find that the compactness measures are often inconsistent when we use them to evaluate real, as well as hypothetical, districts.

2.2. Legal Controversy And Academic Debate

While the court has made its decisions about compactness (Chapter 1), the academic world is still debating the subject. Polsby and Popper, and other strong proponents of compactness, claim in the *Yale Law and Policy Review* that such a standard could virtually eliminate gerrymandering (Stern 1974; Wells 1982), or, at the least, "make the gerrymanderer's life a living hell." (Polsby and Popper 1991: 353)[29] At the same time,

---

[29] In addition, Wells and Stern make claims which are equally, or nearly, as strong (Stern 1974; Wells 1982).

opponents of compactness measures claim that these standards are at best ineffective (Grofman 1985; Musgrove 1977), or at worst often contrary to substantive representative principles.[30] (Cain 1984; Lijphart 1989; Lowenstein and Steinberg 1985; Mayhew 1971)

Those who believe that compactness measures have some effect argue over which (if any) measurement is best. On one hand, Polsby and Popper, while also advocating a particular measure, claim that practically any of the proposed measures will do: "Compactness that constrains gerrymandering is compactness enough." (Polsby and Popper 1991: 340) On the other hand, Young argues that no compactness measure is acceptable — all are fatally inconsistent with each other: "This reliance on formulas has the semblance, but not the substance, of justice." (Young 1988: 113)

2.2.1. Previous Research on Compactness Standards

Unfortunately, current redistricting theory offers no resolution to the debate over compactness standards. While there is a significant literature of varying degrees of mathematical formality, on the theory of redistricting, the vast majority of that literature

[30] Another set of authors argue, more moderately, that particular compactness standards can be used to signal manipulation of district lines (Grofman 1985; Niemi et al. 1991) — ill-compactness is a warning signal that requires justification, or that compactness is a useful, neutral, and objective criterion for limiting gerrymanders. (Morill 1990), but "compactness alone does not make a redistricting plan good." (Niemi, et al. 1991: 1177)

ignores the spatial distribution of voters and institutional constraints on gerrymandering. Even the three papers that model the spatial distributions of voters, (Musgrove 1977), (Snyder 1989), and (Sherstyuk 1993), do not formally evaluate compactness measures within their models.[31]

Most comparisons of compactness measures have been informal: Frolov (1974) comments on a number of compactness measures used by geographers. Young (1988) shows how a number of compactness criteria can produce counterintuitive results.

Much research into compactness has focused on creating compactness standards, mostly *ad hoc*. Table 2-1 shows many of these standards. Empirical research in this area has been limited primarily to the measurement of particular districts or plans. No one, previously, has compared a large sample of standards over a wide range of district plans. Instead, the majority of studies selectively apply chosen measures of compactness to actual and proposed district plans (Hofeller and Grofman 1990; Niemi and Wilkerson 1990; Pildes and Niemi 1993; Reock 1961; Schwartzberg 1966). Two studies compare a

---

[31] Sherstyuk (Sherstyuk 1993) shows that both population equality and substantive contiguity (i.e., excluding "telephone line" style contiguity) limit the opportunity for manipulation. She concludes in general that the addition of any substantive redistricting criteron tends to make gerrymandering more difficult but that the effects and neutrality of such criteria as compactness will depend on population distribution, and may conflict with redistricting goals.

variety of compactness measures using empirical data: Flaherty and Crumplin (Flaherty and Crumplin 1992) use several measures of compactness to measure proposed provincial districts in British Columbia; they recommend two particular area measurements, but do not explicitly compare the consistency of different measures. In one of the few studies to examine the consistency of compactness measures, Niemi et al. (1991) evaluate Congressional district plans for the states of California and Colorado, and state house plans from Indiana, Rhode Island, and New York. They calculate the correlation among selected measures on these plans and find varying levels of consistency among measures, concluding that measurements are most useful when several standards are used simultaneously to compare different plans for the same state.

	Length v. Width	*Earliest Use*
LW_1	W/L: where L is longest diameter and W is the maximum diameter perpendicular to L	(Harris 1964)
LW_2	W/L: from circumscribing rectangle with minimum perimeter	(Niemi, et al. 1991)
LW_3	1/(W/L): rectangle enclosing district and touching it on all four sides for which ratio of length to width is maximum	(Niemi, et al. 1991) modification of (Young 1988)
LW_4	W/L, where L is longest axis and W and L are that of a rectangle enclosing district and touching it on all four sides	(Niemi, et al. 1991)
LW_5	L–W where L and W are measured on north-south and east-west axes, respectively	(Eig and Seitzinger 1981)
LW_6	diameter of inscribed circle/diameter of circumscribed circle	(Frolov 1974)
LW_7	minimum shape diameter/maximum shape diameter	(Flaherty and Crumplin 1992)
	Measurements Based on Area	
A_1	The ratio of the district area to area of minimum circumscribing circle	(Frolov 1974)
A_2	The ratio of district area to the area of the minimum circumscribing hexagon[32]	(Geisler 1985), cited in (Niemi and Wilkerson 1990)
A_3	The ratio of district area to the area of the minimum convex shape that completely contains the district	(Niemi, et al. 1991)
A_4	The ratio of district area to area of the circle with diameter equal to the districts' longest axis	(Gibbs 1961)
A_5	The area of the inscribed circle/area of circumscribed circle	(Flaherty and Crumplin 1992)
A_6	The area of the inscribed circle/area of shape	(Ehrenburg 1892) cited in (Frolov 1974)
A_7	(area of intersection of the shape and circle of equal area)/(area of the union of the shape and the circle of equal area)	(Lee and Sallee 1970)
	Measurements Based on Perimeter/Area Ratios	
PA_1	The ratio of district area to the area of circle with same perimeter	(Cox 1927) cited in (Niemi, et al. 1991)
PA_2	$1\text{-}PA_1^{(1/2)}$	(Attneave and Arnoult 1936) cited in (Niemi, et al. 1991)
PA_3	The ratio of perimeter of the district to the perimeter of a circle with equal area	(Nagel 1835) cited in (Frolov 1974)
PA_4	The perimeter of a district as a percentage of the minimum perimeter enclosing that area ($=100(PA_3)$)	(Pounds 1972)
PA_5	A/0.282P	(Flaherty and Crumplin 1992)

[32] When I analyze this measure, I assume that these hexagons must be regular.

PA_6	$A/(0.282P)^2$	(Flaherty and Crumplin 1992)
	Other Shape Measures	
OS_1	The moment of inertia — the variance of distances from all points in the district to the district's areal center of gravity, normalized. Where A is the area of the shape, r is the distance from the center and D is the set of points in the shape this is $\frac{A}{\sqrt{2\iint_D r^2 dD}}$.	(Boyce and Clark 1964)
OS_2	The average distance from the district's areal center to the point on district perimeter reached by a set of equally spaced lines	(Boyce and Clark 1964)
OS_3	(radius of circle having same area as shape)/(radius of circumscribing circle)	(Flaherty and Crumplin 1992)
OS_4	(N-R)/(N+R) where N,R is # of (non)reflexive interior angles (respectively)	(Taylor 1973)

Table 2-1. Shape based measures of compactness for districts.

Although we now have a large number of compactness standards to choose from, and we know how some districting plans measure up under a few of these, we still do not know what these measurements mean. What are compactness criteria measuring — do they measure gerrymandering? Are these measures consistent with each other — does it matter, really, which one we use? How effective are compactness standards at preventing gerrymandering, which measures should we use, and how should particular minimum compactness levels be set? Beyond preventing manipulation, are compactness standards neutral? What other effects could compactness standards have on politics?

In the next section, I address the first two questions. I use an an axiomatic analysis to test the consistency of existing compactness measures against our intuitions about how gerrymandering is performed in practice. I then develop a methodology to answer the third question by quantifying the strictness of each compactness standard.[33]

2.3. An axiomatic examination of compactness criteria

Recent surveys of compactness criteria list 36 different measurement formulas.[34] While there are a plethora of different measurements, and many assertions as to their effectiveness, only *ad hoc* criteria are used to distinguish between them.[35]

We can learn more about compactness measures by examining their formal properties. In this section I use an axiomatic approach to analyze the consistency of different compactness criteria. Then, in the sections that follow, I will extend this formal analysis with an exhaustive analysis of small districts.

---

[33] This methodology identifies effective standards and leaves open the question of whether compactness standards are politically neutral. In Chapter 4, I show that they are not.

[34] See Niemi, et al., (1991) for the most comprehensive listing. (See also Flaherty and Crumplin 1992; Frolov 1974; Young 1988 for alternative treatments.)

[35] For an isolated exception see Blair & Bliss (1967), a largely overlooked, but more formal approach.

This section proceeds as follows. First, I will use a hypothetical example to introduce the issues surrounding geographical district manipulation. Second, I describe three commonly recognized techniques for manipulating district maps, and offer a set of principles that attempt to capture these different types of manipulation. Third, I show how we can use these axioms to eliminate a majority of the standards found in the literature as inconsistent or nonsensical.

2.3.1. A Hypothetical Example

As an introduction to redistricting with compactness standards, consider a hypothetical square state. This square state is inhabited by two parties with distinct policy preferences, the “Republicans” and the “Democrats.” Members of these factions live in each block.

The political structure of this hypothetical state is as simple as its population. The state is divided into four districts, each of which is composed of some number of indivisible blocks, and from each of which a member of the legislature is elected. When one party outnumbers another in a district, a candidate from that party is elected (Figure 2-1).[36]

---

[36] Here I am assuming that ties are decided by a coin toss, everyone votes, and that everyone votes according to their party identification. While these assumptions simplifies reality, it is reassuring that one can predict nearly 90% of contemporary California elections by using only the partisan registration percentages in each district (Kousser

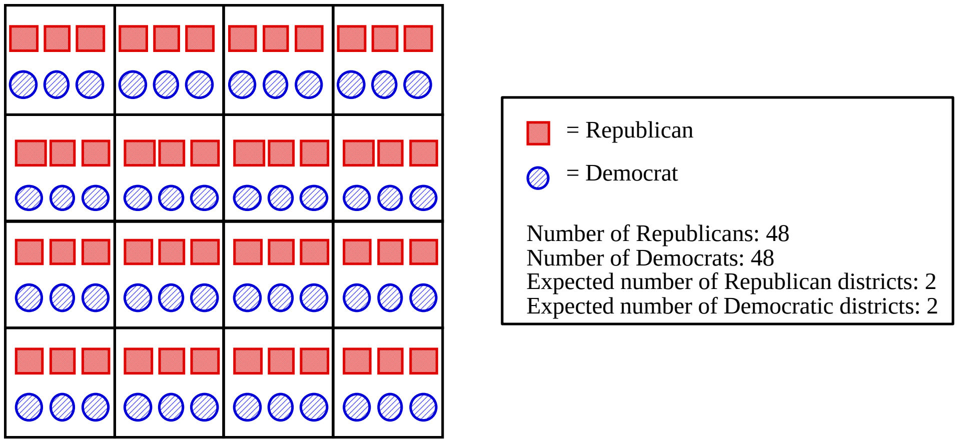


Figure 2-1. Hypothetical state with uniform population distribution.

Consider the situation above, where the population of each group is uniformly distributed across the state. In this most unlikely case, which is illustrated in Figure 2-1, redistricting rules do not matter. No matter which population blocks we use for each district, there will always be, on average, two members from each party in the legislature.

However, if the population distribution is not the same for every block, the situation may be much different: The particular districting plan that the legislature uses and the rules that govern the creation of districting plans in general may strongly influence the composition of the legislature (Figure 2-2).

Rules that constrain a legislature's actions do not necessarily constrain legislative outcomes. For example, if the legislature is required only to draw districts that are contiguous and equal in population, it will still be able to choose between ones that give

---

1995).

an expected majority of seats to either party (Figure 2-2: A,B). Whether these districting rules limit outcomes depends upon how the voting population is geographically distributed.

What happens when compactness is added to the list of district requirements? In fact, the legislature's ability to affect elections depends greatly upon how we measure compactness. Suppose we use Theobald's measure (Section 2.3.2) and define compactness to be the maximum difference between an individual district area and the average area of all districts.[37] In this case, the legislature is not additionally constrained (Figure 2-2: A,B), because every plan that meets the equal population standard will also meet the compactness standard.[38]

But suppose, on the other hand, we use the state of Colorado's definition of compactness, and equate the compactness of a plan with the sum of all the perimeters of its districts. Low numbers are more compact. This compactness measure leads inevitably

---

[37] Formally, if we take a set of districts, number them from 1..*N*, refer to their individual area's as $A_{i,}$, and to the mean of all district areas as $\bar{A}$, then the compactness score of a plan is $\max_i \left| A_i - \bar{A} \right|$. Under this measure a perfectly compact plan has a score of zero.

[38] This is a result of the uniform population density in this state — each population bloc contains both a uniform amount of population and has a uniform area.

to plan C in Figure 2. Plan C is more compact than any other possible plan, even if we discard requirements of contiguity and population equality.[39] In this particular case, a compactness standard gives each party an equal chance of controlling the legislature.

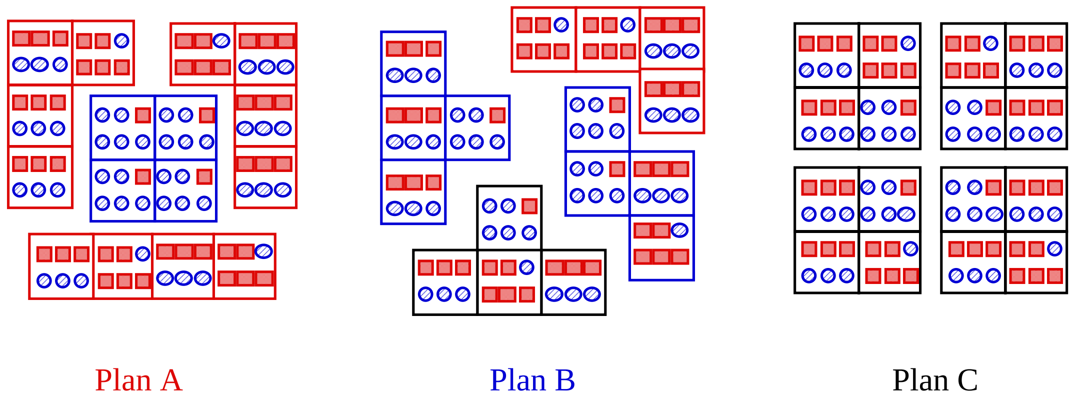


Figure 2-2. Possible redistricting plans under different rules. All three plans meet contiguity, equal population standards, and the first compactness standard defined above. And plans use the same population map. Only Plan C meets the second compactness standard.

---

[39] Many compactness measures find plan C to be uniquely and optimally compact, including comparison to ideal district shape (circle, square, hexagon), length to width ratio, population dispersion and perimeter/area ratios. The intuition behind this is that a square is the most "regular" shape that can be created using these population units, and that only one plan allows all districts to be squares. I have verified that this is, indeed, the optimal plan through an exhaustive analysis.

This example might seem to imply that compactness rules decrease a plan’s partisan bias. Consider, however, the result of the same rules when they are applied to a different distribution of population. In Figure 2-3, our square state contains the same number of Republicans and Democrats as in Figure 2-2, but their locations have changed. In this case, our previously “fair” compactness rule ensures that the Republicans will control the legislature. The compactness rule still limits gerrymandering, in the sense that it makes it impossible to manipulate district lines; however, the rule has a clearly disproportionate effect on different parties. In fact, if someone who knew the population distribution had suggested such a compactness rule, we would have a strong reason to suspect them of partisanship.

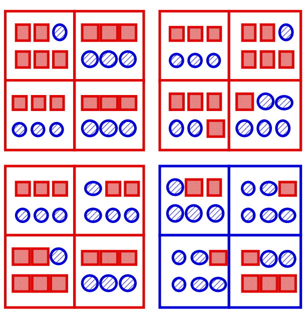


Figure 2-3. Compact plan for another population distribution.

This example illustrates four claims that I will pursue in the rest of this section: First, compactness and other rules governing the shapes of districts may have a powerful effect on the composition of a legislature. Second, rules *can* limit the possibility of manipulation, but some may be stronger than others. Third, the effects that rules have interact with the way in which populations are distributed. Fourth, the effects of a compactness standard may depend very much on how we measure compactness.

It is important to realize that these criteria neither measure nor constrain electoral manipulation directly — they say nothing about the electoral results that can be expected from a particular set of districts. Instead they are proxies that attempt to reflect the ways that gerrymanderers distort district shape to manipulate elections.

2.3.2. Manipulating The Shape Of Individual Districts — Six Axioms

Since most measures of compactness have concentrated on the geographical manipulation individual districts, I will address these measures first. Later, in Section 4.5, I will discuss measures that are based upon the compactness of an entire districting plan and measures that are based upon population dispersion instead of geography.

Most compactness measures claim to describe the *shape* of a district. Therefore we should require that any index of compactness give the same score to two districts that have the same shape.[40] Blair and Bliss (1967) suggest that two objects should be said to have the same shape if we can make them identical through translation, rotation and uniform scaling.[41] I adopt this characterization (Figure 2-4).

---

[40] For district measures that capture population distribution, we would require the measurement to produce identical responses for identical shapes and distributions of population, rather than identical geography (Section 2.3.3).

[41] The definitions of shape that are used by Blair and Bliss (1967) differ slightly, but the axioms are similar. In addition to the three general properties, they claim that a circle

In general, three types of shape distortion and manipulation have been recognized: dispersion, dissection, and indentation (Blair and Bliss 1967; Flaherty and Crumplin 1992; Frolov 1974).[42] While there is no consensus on how these concepts should be precisely measured, it is easy to describe each intuitively. *Dispersion* reflects the symmetry of a shape around its center — a circle is evenly dispersed, whereas a ellipse is less evenly dispersed. *Dissection* reflects discontinuity in the distribution of points across the convex hull of a shape — shapes with holes cut out of them are highly dissected. *Indentation* reflects the smoothness of the perimeter of a shape — most coastlines are examples of indented shapes (Figure 2-4).

---

should be judged to be maximally compact under any reasonable index. I leave this out, as it is an implication of the axiom's I suggest later.

[42] While these authors refer to a shared set of concepts, their terminology sometimes varies. For example, where Flaherty & Crumplin (1992) refer to *compactness,* I refer to *dispersion* in order to distinguish this concept from the more inclusive meaning of the term in general use. Niemi et al. (1991) points out the importance of population.

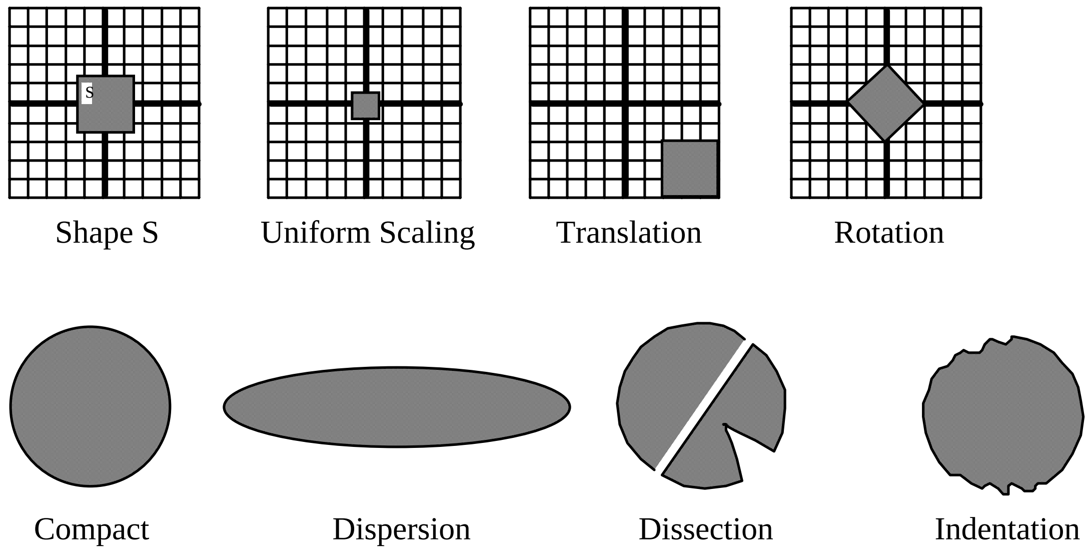


Figure 2-4. Transformation of shapes.

Suppose that we were to place legal limits on the amount of indentation, dispersion and dissection allowed in each district. How would this affect gerrymandering? As we made these limits more stringent, the set of plans from which a gerrymanderer could choose would shrink. In a very simple world, indentation, dispersion, and dissection might closely reflect the ability of a gerrymanderer to pick and choose district plans to her liking.

For example, imagine that the voting patterns within each district are completely predictable, that all voting data is known with certainty, and that people are evenly distributed over each square mile of our hypothetical state. Also imagine that there are only two parties, that they have an equal number of loyal voters, and that these voters are uniformly randomly distributed across space. In this case, the amount of indentation (or dispersion or dissection) that district planners are allowed to use when they create a districting plan directly limits the set of districts from which planners can choose, and will increase the *ex-ante* probability of their being able to choose a winning plan for their party.

If, as the literature indicates, these three types of shape distortion are good proxies for geographical manipulation, then acceptable measures of geographic compactness should capture at least one, if not more, of these principles. In the remainder of this section, I formalize these principles.

First, we will need some definitions:

Let a shape $S = \{s_1, \ldots, s_i\}$ be a finite, nonempty set of simple, continuous, closed, nonoverlapping subsets of the plane where $Area(s_i \cap s_j) = Perimeter(s_i \cap s_j) = 0, \forall i \neq j$.

Let $P: S \rightarrow \Re_+$ be the length of the perimeter of the shape, and let $A: S \rightarrow \Re_+$ be the area of the shape.

Let a compactness measure *C*, be a function $C: S \rightarrow \Re$.

Using these definitions, we can now formally define what it means for a compact measure to capture shape:

1. *Scale independence*: if two shapes differ only in scale, then they should be equally compact.

Formally, $\forall \alpha \in (0,1] \Rightarrow C(\alpha S) = C(S)$.

2. *Rotation independence*: if S_1, S_2 are two shapes which differ only in rotation around the origin, they should be equally compact. Formally, if θ is an angle, and then

$$S' = \{p' | p \in S, p'_x = p_x \cos\theta - p_y \sin\theta, p'_y = p_x \sin\theta - p_y \cos\theta\} \Rightarrow C(S) = C(S')$$

.

3. *Translation independence*: if S_1, S_2 are two shapes which differ only in position, they should be equally compact. Formally, $\theta \in \Re^2 \Rightarrow C(S+\theta) = C(S)$.

A compactness measure must not violate any of these three principles. It would be strange indeed if we could change a district's shape simply by uniformly scaling, rotating, or moving the map upon which it is drawn. If a compactness measure does not meet these basic standards,[43] political actors would be able to manipulate the compactness of their districts simply by manipulating the maps upon which they are drawn.

In the next three principles, I capture the concepts of dispersion, dissection, and indentation. First, let us take dispersion. Compactness measures that claim to capture dispersion are usually based on the ratio of a shape's perimeter to its area. These

---

[43] Remember that these measures are based upon geography alone. Violations of principles 1-3, as stated here, might be quite reasonable if we were measuring population: For example, moving a square district to a different part of the map could completely change the population distribution within that district. Fortunately, we can both preserve these principles and reflect population distributions — see Section 2.3.3.

measures work well for convex shapes, but can confuse indentation and dispersion for nonconvex shapes (Figure 2-5).

Figure 2-5. The Perimeter/Area ratio fails to capture dispersion. The P/A of the figure on the left is less than that of the one on the right.

For example, the shapes in Figure 2-5 have equal area but the perimeter of the "long rectangle" below is much less than the "coastal" square to its right. Measures based on the perimeter/area ratio judge the square to be more dispersed, whereas, intuitively we can see that it is really less dispersed, but more indented.

By measuring the perimeter of the convex hull of the shape, we can avoid this confusion. Intuitively, the convex hull, which we will refer to as "CO," smoothes out the bumps in the shape and allows us to look at its broader outline.

4. *Minimal dispersion*: A compactness measure reflects the principle of dispersion if, for all shapes S_1, S_2, if S_1 and S_2 are of equal area, and the perimeter of the convex hull of S_1 is larger, S_1 is less compact:

Formally, $A(S_1) = A(S_2) \& P(CO(S_1)) > P(CO(S_2)) \Rightarrow C(S_2) > C(S_1)$.

We can also use the convex hull to compare two shapes that have the same general outlines, so as to see which is relatively more dissected or indented:

5. *Minimal dissection*: Let *CO(S)* be the convex hull of shape *S*. If S_1 and S_2 are any two shapes with identical convex hulls, and S_1 has a strictly smaller area, then S_1 should be judged less compact:

Formally, if $CO(S_1) = CO(S_2)$, and $A(S_1) < A(S_2) \Rightarrow C(S_2) > C(S_1)$.

6. *Minimal indentation*: If S_1 and S_2 have identical convex hulls and S_1 has a strictly larger perimeter/area ratio, S_1 should be judged less compact.

Formally, if $CO(S_1) = CO(S_2)$, then $\frac{P(S_1)}{A(S_1)} > \frac{P(S_2)}{A(S_2)} \Rightarrow C(S_1) < C(S_2)$.[44]

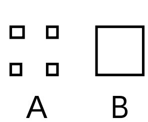
In addition to capturing recognized methods of manipulation, any compactness measure that satisfies any of these axioms will have two other nice properties. Contiguity is usually assumed, *ad hoc*, to limit manipulation;[45] similarly, a circle is often assumed to

---

[44] One alternative to 6 that might be offered is 6': $CO(S_1) = CO(S_2)$, then $P(S_1) > P(S_2) \Rightarrow C(S_1) < C(S_2)$.

Since 6' is analogous to 5, it seems natural, at first glance, but leads to surprising conclusions involving noncontiguous districts.

For example, under 6', shape A is more compact than shape B:



[45] Though see Sherstyuk 1993 for a formal approach to contiguity.

be maximally compact. Under the six axioms above, we need not assume these properties, but can derive them (See the Appendix for proofs):

Result 1. If a compactness measure satisfies axioms 1–4 and either axiom 5 or axiom 6, then it has the following properties:

Contiguity: For any given perimeter or area, the maximally compact shape is contiguous. If C satisfies axiom 5, this is true for any given convex hull as well.

Circle Compactness: A circle is the most compact shape.

Most research on compactness assumes that it can be measured on a single unidimensional scale. Like Flaherty and Crumplin (1992) and Niemi, et al. (1991), I find this assumption to be incorrect. Many measures seem simply to be measuring a different aspect of compactness — there is more than one way to manipulate a shape.

Result 2. It is impossible for a single index of compactness to meet axioms 5 and 6 simultaneously (Figure 2-6). Axiom's five and six can contradict: Under Axiom 5, C(B)>C(A), but under Axiom 6 C(A)>C(B).[46] (See Figure 2-6.)

---

[46] I could reformulate axiom 6 avoid this conflict. For example, let us define axiom 6' as follows:

$A(S_1) = A(S_2)$, then $\frac{P(S_1)}{A(S_1)} > \frac{P(S_2)}{A(S_2)} \Rightarrow C(S_1) < C(S_2)$. But this seems to blur the distinction between indentation and dispersion.

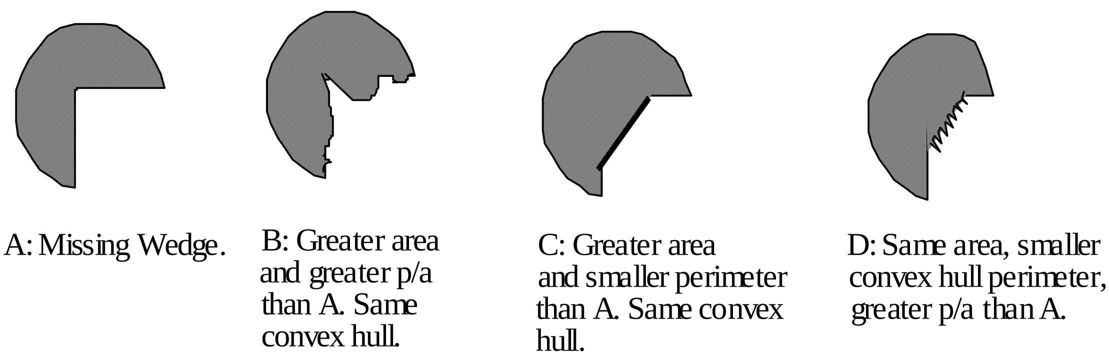


Figure 2-6. Shapes *A* and *B* have identical convex hulls, P(A) > P(B), and A(A) > A(B). Under axiom 5, *A* should be more compact, while *B* should be more compact under axiom 6.

What does Result 2 mean? There is significant debate over whether compactness standards are measuring the same things. This result shows that differences among compactness measures exist and are, to an extent, unavoidable.[47]

---

[47] As I discussed at the beginning of this section, geographical methods of manipulating districts remain proxies for the end goal of gerrymandering — influencing the results of an election. While in some ways more direct, this characterization, less general because it requires assumptions about underlying population distributions, optimization methods for creating compact plans, and electoral goals. First, gerrymanderer's may have different competing goals for electoral results — incumbent protection and partisanship conflict (Owen and Grofman 1988). Second, optimal compact gerrymanders are very difficult to create, and in practice the effect of a compactness

These principles set bounds on a reasonable compactness standard — if a compactness measure contradicts all three of principles of compactness (or any of the shape principles), we should suspect it of measuring something other than geographic compactness. Table 2-2 summarizes the results of applying these axioms. (See this chapter's Appendix for proofs of these results.) In it I list each measure and the axioms that it violates.

---

standard will vary with the particular methods used for creating districts. (See Chapter five.) Third, the extent to which a particular compactness standard constrains a (partisan) gerrymanderer's ability to gain seats depends not only on the compactness measure, but on the spatial distribution of (partisan) voters. (See Chapter 4.)

Measure	*Axiom 1*	*Axiom 2*	*Axiom 3*	*Axiom 4*	*Axiom 5*	*Axiom 6*
LW_1				**V**	**V**	**V**
LW_2				**V**	**V**	**V**
LW_3				**V**	**V**	**V**
LW_4				**V**	**V**	**V**
LW_5		**V**		**V**	**V**	**V**
LW_6				**V**	**V**	**V**
LW_7				**V**	**V**	**V**
A_1				**V**		**V**
A_2				**V**		**V**
A_3				**V**		**V**
A_4				**V**	**V**	**V**
A_5				**V**	**V**	**V**
A_6				**V**	**V**	**V**
A_7				**V**	**V**	**V**
PA_1	**V***			**V**	**V**	
PA_2	**V***			**V**	**V**	
PA_3	**V***			**V**	**V**	
PA_4	**V***			**V**	**V**	
PA_5	**V**			**V**	**V**	
PA_6	**V***			**V**	**V**	
OS_1				**V**	**V**	**V**
OS_2		**V**		**V**	**V**	**V**
OS_3				**V**		**V**
OS_4				**V**	**V**	**V**

Table 2-2. Violations of the measurement axioms by compactness measure are marked by a 'V' in the cell.

* In these cases the measure is sensitive to the scale of the measuring unit used to measure the district boundaries, not to the scale of the map upon which the district boundaries are represented.

The results in this table enable us, in two ways, to trim[48] the set of compactness standards that courts and political scientists should consider adopting: First, although most of the compactness measures meet our first three axioms, eight measures violate, in their standard form, these basic axioms for measuring shape. Three measures, LW_5, OS_2 and PA_5, unequivocally violate these axioms, and so should be rejected. Five others, the remaining PA measures, in their current form, violate axiom 1, but they can be saved we are careful to measure the boundaries of all districts with the same precision.

Second, 13 compactness measures violate all 3 axioms of compactness. In other words, they do not comport with the commonly held intuitions about how gerrymandering is accomplished geographically. They should be rejected in the absence of compelling theoretical or empirical evidence that these measures are, in fact, measuring other aspects of gerrymandering.

---

[48] In addition, the ubiquitous violation of Axiom 4 points to a way in which our measurements of compactness can become more complete: by using at least one measurement that captures indentation. In Appendix 1 I create, for example, one such a measure.

2.3.3. Extension 1: Capturing Population Distribution.

District lines affect elections only when these lines affect how we assign voters to districts: A meandering district line in a dense urban area may indicate political manipulation, but the same line, when found following a river in a sparsely populated rural area, may be devoid of political content. Measures of population compactness attempt to capture this distinction.

If the population density in a state is uniform, the first three of these population measures are equivalent to the geographical measures A3, A1, and OS1, respectively.[49] In general, the difference between these measures and their geographically-based counterparts will depend on how people are distributed in a state (Table 2-3).

---

[49] Population measure four is designed as a computationally simpler approximation of POP_3 (Papayanopoulos 1973). Given that computers are now sufficiently powerful to calculate good approximations of POP_3 directly, this may no longer be necessary.

Also note that POP_1, like its shape counterpart A_3, violates the modified axiom 4 even where population density is uniform.

	Population measures	
POP_1	ratio of district population to the population of the minimum convex shape that completely contains the district	(Niemi, et al. 1991)
POP_2	ratio of the district population to the population in the minimum circumscribing circle	(Niemi, et al. 1991)
POP_3	population moment of inertia, normalized	(Weaver and Hess 1963)
POP_4	sum of all pair-wise distances between centers of subunits of legislative population, weighted by subunit population	(Papayanopoulos 1973)
	Plan Compactness Measures	
PL_1	The sum of the district perimeters	(Adams 1977)
PL_2	The maximum absolute deviation from the average district area	(Theobald 1970) cited in (Niemi, et al. 1991)

Table 2-3. Measures of compactness that evaluate properties other than district shape.

Niemi, et al., (1991) argue reasonably that the compactness of a district should not change when we include unpopulated parcels of land. We can go farther, however, in making measures sensitive to population. It is reasonable to expect that the more people we are allowed to shift from district to district, the larger the potential for political manipulation; therefore, compactness should reflect not only the presence of people, but their numbers.

Fortunately, there is an easy way to transform a compactness measure that evaluates only geography into a compactness measure that it is sensitive to population, as well. We can make this conversion not by changing the measure itself, but by changing the map to which we apply the compactness criteria. By using a map where population and area are made equivalent, we can measure population compactness with any of our geographical compactness measures.

Tobler (1973) shows how to generate this type of map.[50] Compactness measurements made on these maps automatically reflect manipulation of population. For example, if we examine a boundary line that follows an unpopulated river bed, it will seem highly indented on a conventional map, while it will show no indentation at all on the transformed map. On the other hand, if our boundary line is in the middle of a densely populated urban area, any irregularities will become magnified on the population map. If we measure compactness using these maps, we truly look at people, not acres.

2.3.4. Extension 2: Measuring The Compactness Of Plans.

Since districts share borders, the shapes of districts in a plan are interdependent. Most compactness measures, however, examine districts in isolation from the plan in which they are embedded. Even when researchers are forced to define some measure of plan compactness, instead of evaluating the plan as a whole, they often simply measure the compactness of each district and then use the mean or minimum of these district-based scores.

Measures of plan compactness should be sensitive to improvements in districts. At the least, a plan measure should reward making one district more compact, if other districts are not made any worse. Otherwise, the compactness of an entire plan may be

---

[50] Tobler created this transformation as a method for drawing equal population districts, not for creating compact districts per se. He does suggest however, that a hexagon could be used as an appropriate “compact” shape.

determined by a single district, which is unlikely to reflect adequately the degree of electoral manipulation in the whole plan. Researchers have created only two measures that take the entire plan as the basic unit for which compactness is measured (Table 2-3).

Axiom 7 states this criterion more formally. First we need a few definitions. Let plans P_1,P_2 be sets of shapes. We are given a measure *C* of shape compactness, satisfying axioms 1–3, and 4, 5 or 6. Define $CP: P \rightarrow \Re$ to be a measure of plan compactness.

Axiom 7 (weak Pareto comparison): If every district in plan 1 is at least as good as every district in plan 2 and one district is better, then plan 1 is more compact. Formally, let f be a bijection mapping each element of P_1 to a single element of P_2:

$$\exists f, s.t.\ C(S_i) \geq C(f(S_i)) \forall S_i \in P_1, \text{and } \exists S_j \in P_i\ \ s.t.\ C(S_j) \geq C(f(S_j)) \Rightarrow CP(P_1) > CP(P_2)$$

If a district-compactness measure, *C*, satisfies axioms 1–3 and 4, 5, or 6 then the mean district compactness, $PC_{mean}(P) = \frac{\sum_{S_i \in P} C(S_i)}{\# P}$, satisfies axiom 7. Measuring the compactness of a plan based on the minimum district compactness, $PC_{\min}(S) = \min_{\forall S_i \in P}(C(S_i))$, violates axiom 7 (Figure 2-7). In addition, both PL_1 and PL_2 violate scale invariance, although if we wish only to compare two plans that are drawn upon the same map, this is not a serious defect. Even so, PL_2 is unsatisfactory because it violates axioms 4, 5, and 6[51] (Figure 2-7).

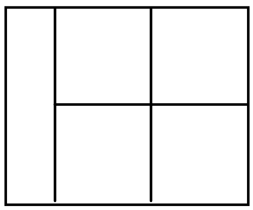
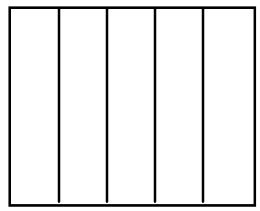
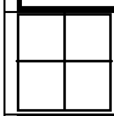
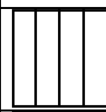
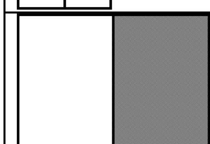
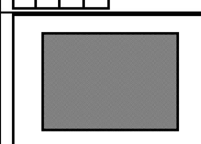
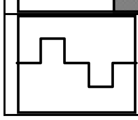
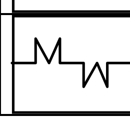
More Compact	*Less Compact*	*Measures Not Classifying These Plans Correctly (Axiom Violated)*
		Any plan measure based on the score of the minimally compact district. (Axiom 7)
		PL_1,PL_2 (Axiom 1)
		PL_2 (Axiom 4)
		PL_2 (Axiom 5)
		PL_2 (Axiom 6)

Figure 2-7. Violations of Axioms 1, 4, 5, 6, and 7 by measures of plan compactness.

[51] Note that on PL_2, when applied to a uniform population map such as we described in the previous section, is equivalent to the equal-population standard.

2.4. Evaluating The Consistency Of Compactness Measures With Small Cases

The analysis in the previous section showed that the worst of the individual compactness measures fail even to measure shape adequately, and that the best of them capture only limited aspects of geographical manipulation. In this section I use an exhaustive analysis of small cases to quantify the amount of agreement among various compactness measures and to quantify how sensitive these measures are to manipulation of district shape.

2.4.1. Generating Districts And Plans

If two compactness measures produce the same rankings over all sets of districts, they are identical for all practical purposes. While few compactness measures are completely identical in this way, we can use several straightforward statistics to analyze the agreement between rankings over a given set of districts.

How do I choose an unbiased set of "test" districts for our comparisons? I use the exhaustive set, the set of all districts that can be created on a given map with population blocks of unit size. The choice to use an exhaustive set allows me to avoid bias in the selection of particular districts; but because this set grows very quickly as the number of census blocks in a district increases, it limits the size of the districts that can be examined.

I start by creating a small artificial district map that consists of a rectangle of population blocks (similar to the examples in Section 2.3.1). I use combinations of these blocks to form individual districts. I then create an exhaustive compactness *ranking* by using one compactness criterion to rank all the possible districts that can be created on that map. Finally, I compare the rankings produced by different compactness rankings to determine the similarity among measures.

Generating district plans is a bit more complicated than generating individual districts. We can characterize redistricting mathematically as a partitioning problem.[52] Imagine that each state in the U.S. is composed of indivisible population units,[53] in this case creating a plan is equivalent to partitioning these units. If we care about what a plan looks like, then we can add a value function to our partitioning that incorporates such criteria as contiguity, compactness, and population equality. I create two sets of plans for

---

[52]A partition divides a set into component groups which are exhaustive and exclusive. More formally:

For any set $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$, a *partition* is defined as
a set of sets $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_k\}$ $s.t.$
(1) $\forall x_i \in \mathbf{x},\ \exists \mathbf{y}_j \in \mathbf{Y}, s.t.\ x_i \in \mathbf{y}_j$
(2) $\forall i, \forall j \neq i, \mathbf{y}_j \cap y_i = \varnothing$

See Stanton (1986) for an overview of algorithms to create exhaustive lists of partitions.

[53] Census blocs for redistricting purposes may often be considered to be practically indivisible.

each map: The first set is exhaustive; it includes all possible redistricting plans. The second set is more selective; it contains all the plans that meet the constraint of population equality. After creating these plans, I then use them to create exhaustive rankings for the plan compactness measures, just as I described using exhaustive sets of districts.

Even when we use maps that contain a small number of population units, we can create a surprisingly large number of distinct district plans. For example, if we want to produce all possible districts from an *n* by *m* rectangle of population blocks, the number of districts, *d*, that we can create is represented by the function $d = 2^{(n)(m)}$. Not only is this large, but it grows exponentially as we increase the number of population blocks in our map. The number of plans in an exhaustive set can grow even faster than the number of districts. If we have *n* by *m* population blocks and want to create *r* districts, we can create $S((nm),r) = \frac{1}{r!}\sum_{i=0}^{r}(-1)^i\left(\frac{r!}{(r-i)!\,i!}\right)(r-i)^{nm}$ plans.[54] If, however, each district in a plan has exactly the same number of blocks, *k*, then the number of plans we need to create is a bit smaller: $\frac{(nm)!}{r!(k!)^r}$.

[54] *S* is known as a “Stirling Number of the Second Kind.” See Even (1973) for an introduction.

2.4.2. Results From The Small Case Analysis

Since the length of our list of exhaustive rankings tends to grow exponentially as we add population units, we can only use this technique on relatively small maps. I examine all rectangular maps that measure 4 by 4 or smaller (2x2,2x3,3x3,2x4,3x4,4x4). Even though these sizes are small, the number of plans we can generate from them is large — up to 90,000 different plans can be generated from the 4 by 4 grid.

The results from this exhaustive analysis reinforce our previous theoretical analysis: many district and plan compactness measures judge districts quite differently. Furthermore, some measures are much more sensitive to the manipulation of district lines.

Measures Of District Compactness Are Inconsistent.

We first turn to measures of district compactness. For this part of the analysis, I selected seven district measures that either satisfied a large number of the six axiomatic criteria or have received particular attention in the literature: measures A_1, OS_1, PA_6, LW_5, PA_3, PA_5, and A_7, as defined previously. I used these seven different compactness measures to rank all the districts that could be created for each map.

Box-plots[55] allow us to compare the distribution of compactness scores when each of these compactness measures is applied to an identical set of districts. These distributions

[55] Box plots are commonly used to compare distributions. In these plots, the top and

of compactness scores have two striking features: First, district scores are concentrated in a narrow range, and, second, there are few extremely compact districts. (Figure 2-8).

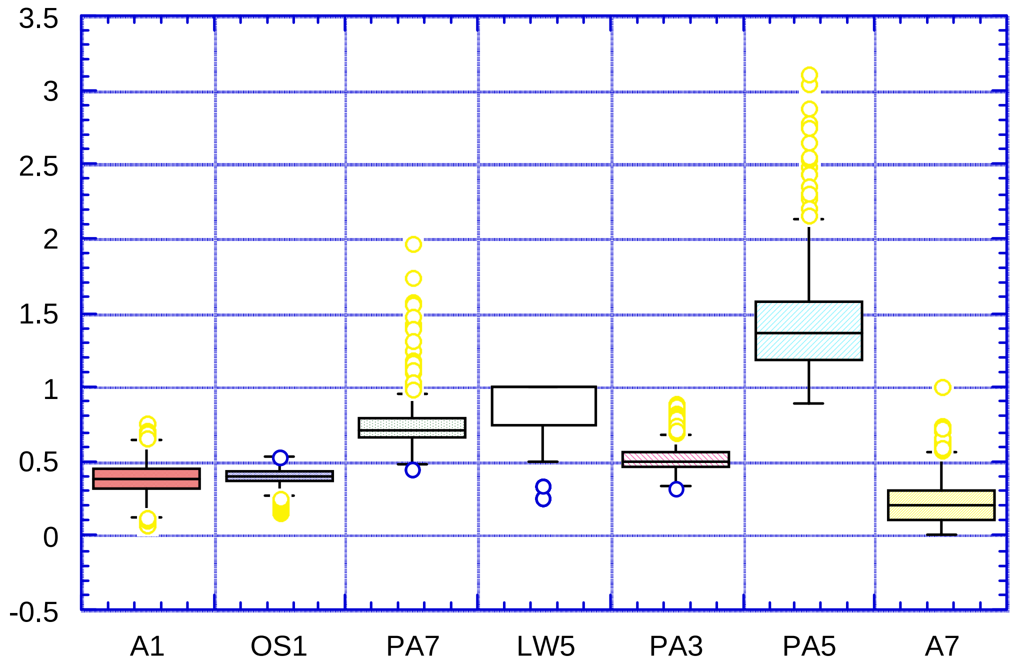


Figure 2-8. Box -plots of district compactness scores for all districts (the exhaustive set) on a 3x4 map.

How can we use these observations to design better redistricting regulations? Some researchers have proposed that we require all districts to meet a specified minimum level of compactness, while others would use compactness scores only as a relative measure to

---

bottom of the box correspond to the 25th and 75th percentiles of the variable, while the whisker lines extend beyond the box by one and one-half times the interquartile range (so that approximately 99 percent of normally distributed data will lie within them.) The median is identified by a horizontal line, and outliers are identified by the small circles.

make comparisons between districts. My results suggest that compactness scores are more useful as relative measures than as absolute measures. Since the distribution of compactness scores is so narrow, it will be very difficult to set a compactness limit that is restrictive without being draconian.

Although the shapes of the distributions of compactness rankings were similar across compactness measures, there were two striking differences among the rankings themselves. First, all compactness measures were not equally strict when judging differences between districts, in fact, some measures ignored all but the grossest differences. One way we can measure this sensitivity is by examining the number of different classes of equivalent scores assigned within each ranking, or *equivalence classes*. The greater the number of equivalence classes, the more sensitive is the measure to district shape.[56]

Table 2-4 shows the number of equivalence classes that each compactness measure produced when it ranked all the districts for a map of given size. In the final column,

---

[56] An alternative way of characterizing the strictness of plans is to compare how difficult it is to generate compact plans under each standard, and to what extent this constrains gerrymandering for given electoral goals. (See Note 119, below.) In Chapter 6, I analyze three compactness standards in this way. As expected, the compactness measure which generated few equivalence classes in this study was also difficult under this alternative characterization of strictness.

notice the extreme contrast between LW_5 and OS_1. LW_5, which is one of the few compactness measures that has been put into law, is extremely insensitive — all 65,000 are assigned one of only 6 distinct scores. Other things equal, we should prefer measures that are more sensitive to district shape to those that are less sensitive, because compactness criteria that capture only the most dramatic differences in district shape are unlikely to strongly restrict gerrymandering.

	2x2 (11)	*3x2* (57)	*2x4* (247)	*3x3* (502)	*3x4* (4083)	*4x4* (65519)
A1	4	11	19	19	39	55
OS1	4	17	45	57	299	953
PA6	2	8	16	17	41	56
LW5	2	4	5	4	6	6
PA3	4	11	20	24	48	88
PA5	4	8	14	17	36	63
A7	3	2	11	15	25	47

Table 2-4. Equivalence classes produced by different measures of compactness. Grid size (number of shapes) is shown in columns.

A second striking difference among compactness measures is the order in which they rank particular districts — many measures do not seem to be measuring the same thing. Table 2-5 shows the concordance[57] among cardinal compactness scores of all districts on

[57] Compactness is for most purposes a relative measure, not an absolute measure. When the courts compare two districts, they will not ask “How do these districts score?” but “Which district is more compact?” To evaluate the similarity of relative compactness

the 3 by 4 map. If these different measures ranked districts in the same way, then all of these concordances should be close to one; many measures seem to have little relationship to each other.

	A1	*OS1*	*PA6*	*LW5*	*PA3*	*PA5*
OS1	0.67 (0.83)					
PA6	-0.25 (-0.39)	-0.19 (-0.39)				
LW5	0.14 (0.24)	0.09 (0.11)	-0.20 (-0.20)			
PA3	0.29 (0.48)	0.25 (0.36)	0.51 (0.60)	-0.38 (0)		
PA5	0.58 (0.73)	0.74 (0.86)	0.07 (0)	0.34 (0)	0.47 (0.71)	
A7	0.43 (0.56)	0.58 (0.74)	0.02 (0)	0.33 (0)	0.34 (0.48)	0.56 (0.72)

Table 2-5. Degree of agreement (Kendall's τ) in pair-wise comparisons for the 3x4 case. (Correlations are reported in parentheses.)

Although compactness measures disagree over how districts should be ranked, if compactness measured agreed about which districts were "best" — at the top of the rankings — other disagreements might be less important. I investigated the possibility of this sort of agreement in two ways: First I recomputed Table 2-5 using only the top 10

---

judgments between each pair of measures, I use Kendall's τ_β. For each pair of measures, I count the number of times where both measures agree that one district is more compact than the other, C ("concurrences"), the number of strict disagreements, *D*, and the number of unilateral ties T_X & T_Y. I then compute Kendall's τ_β using the following formula[57]:

$$\tau_\beta = \frac{C-D}{\sqrt{(C+D+T_x)(C+D+T_y)}}.$$

percent of district rankings, but compactness measures continued to disagree.[58] Second, I calculated the similarity between the ten districts chosen as *most* compact by each measure (Table 2-6).[59] Still, the differences between measures remain: compactness measures disagree over good districts as much as they disagree over bad districts.

	A1	*OS1*	*PA6*	*LW5*	*PA3*	*PA5*
OS1	0.59					
PA6	0.10	0.15				
LW5	0.20	0.23	0.14			
PA3	0.45	0.50	0.10	0.18		
PA5	0.59	0.82	0.15	0.23	0.81	
A7	0.44	0.45	0.10	0.16	0.46	0.40

Table 2-6. Similarity between top ranked shapes.

Population Measures And Consistency.

Each of the seven measures that I tested in the previous sections looks only at geography. Would compactness measures that are based on population be more consistent? Only two population-based measures meet our previously discussed

---

[58] To limit comparisons in this way, one must choose a particular measure to select the top 10 percent of districts, or repeat the process for each measure. I chose the latter approach as more thorough, but omit the seven resulting tables to save space.

[59] Similarities between districts are computed with the following formula $\frac{\#blocks(S_1 \cap S_2)}{\#blocks(S_1 \cup S_2)}$. This is based on Lee and Sallee (1970) although I have adapted it to the discrete case.

requirements, and I examined both of these. So as to see how these measures would be affected by differences in population distribution, I assigned a random population weight (a discrete uniform distribution) to each population block in the model. I then examined these measures using the same techniques that I used for the previous analyses.

Range of Population Distribution	*τ(AC,PAC)*	*τ(MI,PMI)*	*τ(PMI,PAC)*	*Equivalence Classes (PAC)*	*Equivalence Classes (PMI)*
(0,1)	0.59	0.50	0.74	19	121
(0,2)	0.56	0.58	0.77	23	192
(0,3)	0.56	0.42	0.73	69	290
(0,4)	0.59	0.58	0.77	113	340
(0,5)	0.81	0.81	0.54	123	340

Table 2-7. Degree of agreement between rankings, and equivalence classes, for population based measures of district compactness. Each row records results using different parameters for the random distribution of population. The first three columns show the index of concordance between pairs of measures (3 by 4 map).

In Table 2-7, I compare rankings between the population-based measures and between these measures and their geographical counterparts. Because the model assigned random weights to population blocks in each run, the results varied somewhat from run to run, but the patterns in the data remained consistent: Population based measures are no more consistent with each other than are their geographical counterparts.[60]

---

[60] Population measures are, however, unsurprisingly, more closely related to other population measures than to other geographical measures.

How Effective Is Mandatory Plan Compactness?

When we use compactness measures to rank entire district plans rather than single districts, we continue to see inconsistencies among different measures. I used eight different measures of plan compactness: the two measures specifically designed for plans (PL1,PL2), and the six measures based upon the average and the minimum of individual district scores (A1,OS1, and PA6). I used these measures to evaluate exhaustive sets of both *balanced* and *unbalanced* plans: Plans are *balanced* when each district has an equal area, and they are *unbalanced* when they may contain districts with unequal (but nonzero) area.[61] The distributions of compactness scores for balanced and unbalanced plans in a typical case are shown below: (Figure 2-9)

[61] Note that PL2 is zero for all balanced plans.

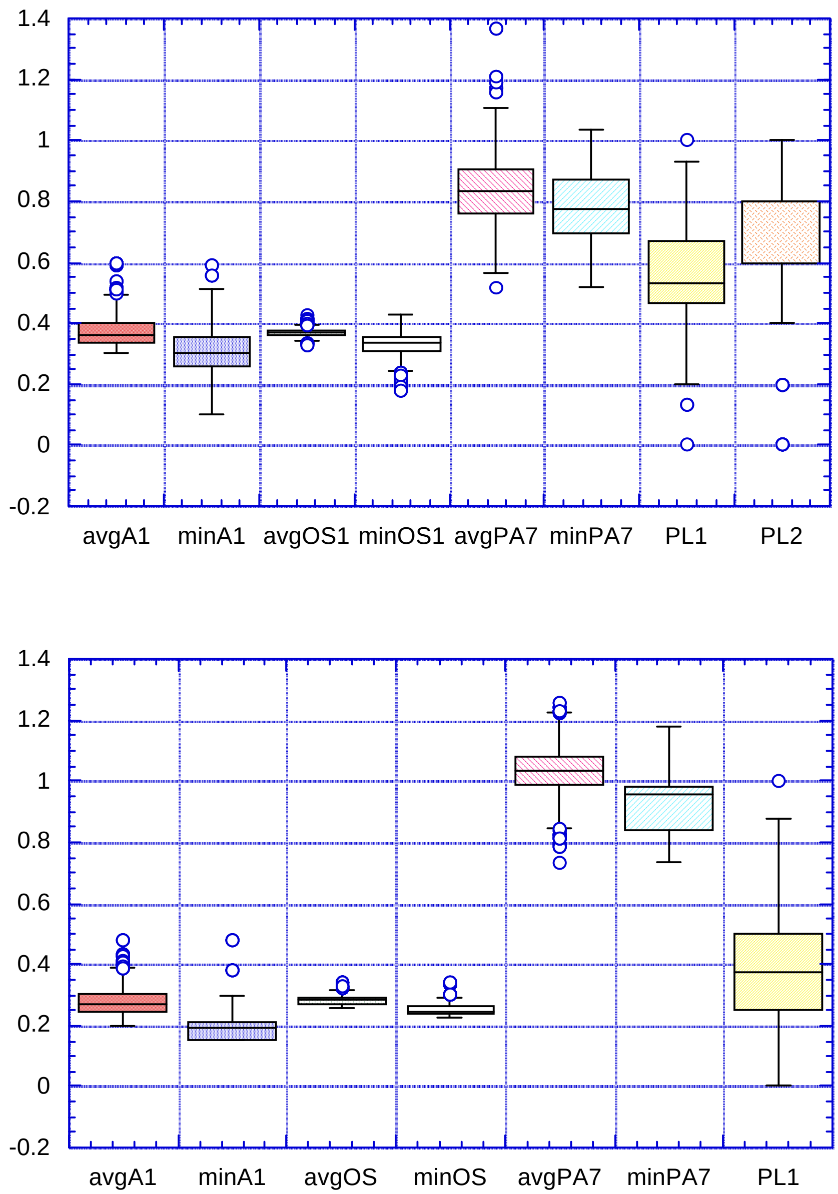


Figure 2-9. Box plots of plan compactness for a 3x4 grid, partitioned into two districts. “Avg” indicates the average district score using a given measure, while “min” indicates the score of the minimally compact district under that measure.

Measures PL1 and PL2 are normalized to (0,1). Top: distribution of balanced plans. Bottom: distributions of unbalanced plans.

As in our examination of individual districts, similarities among the distributions of compactness scores belie differences in the ways that each compactness measure ranked plans. If anything, plan measures disagreed more frequently over how to evaluate plans than district measures differed on the rankings of individual districts (Table 2-8):

	Avg. A1	*Min. A1*	*Avg. OS1*	*Min. OS1*	*Avg. PA6*	*Min. PA6*	*PL1*
Min. A1	0.47						
Avg. OS1	0.40	0.45					
Min. OS1	0.19	0.54	0.49				
Avg. PA6	-0.17	-0.28	0.05	-0.17			
Min. PA6	-0.13	-0.20	0.14	-0.07	0.75		
PL1	-0.13	0.09	-0.22	0.12	-0.76	-0.68	
PL2	0.01	-0.21	-0.22	-0.56	0.25	0.06	-0.30

Table 2-8. Similarities in plan compactness rankings (Kendall's τ_β) or the 3x4 case with 2 districts.

Table 2-9 shows the number of equivalence classes created by each plan compactness measure for a selected map. The number of classes varies between plans measures, supporting our previous conclusions that some measures are much more sensitive than others.

Grid Size (# of plans)	*districts per plan*	*Avg. A1*	*Min. A1*	*Avg. OS1*	*Min. OS1*	*Avg. PA6*	*Min. PA6*	*PL1*	*PL2a*
2x4 (35)	2	7	4	13	13	12	9	7	n/a
2x4 (126)	2	16	15	33	26	28	11	8	4
3x3 (280)	3	14	3	28	7	28	5	7	n/a
3x3 (3024)	3	93	16	309	13	162	12	9	5
3x4 (462)	2	12	4	66	48	61	20	12	n/a
3x4 (2046)	2	61	29	296	142	157	25	15	6
3x4 (5775)	3	60	5	570	30	438	17	12	n/a
3x4 (88534)	3	689	31	7960	74	2886	32	14	7
3x4 (15400)	4	152	7	571	12	715	9	9	n/a

Table 2-9. Equivalence classes for plan based measures of compactness. Balanced plans are indicated by *italics*.

2.5. Discussion

In this chapter, I have answered the question “Are compactness measures consistent?” and started to answer the question of “Are compactness measures effective?” Many advocates of compactness assume that the choice of a particular compactness measure is relatively unimportant. My research shows this assumption to be false: The worst compactness measures, such as raw ratios of perimeter to area and length to width, fail to capture any of the common intuitions about how geographical gerrymandering works. The best compactness measures can capture only limited aspects of geographical manipulation — gerrymandering is multifaceted, and no single one-dimensional index suffices to capture all aspects of it.

Compactness standards are only proxies for electoral manipulation — no author has based their compactness measure on an explicit theory of the electoral effects of district lines. Most measures claim to flag suspect district shapes, shapes that may indicate undue manipulation of district lines. Previous examinations of compactness have been hindered by the absence of a set of reasonable minimal criteria for compactness measures, resulting in a multiplication of measures of questionable value. Although no single perfect measure of compactness exists, by developing a set of minimal standards for compactness measures, I have been able to eliminate many measures that fail to comport with common intuitions about gerrymandering and common understandings about measuring ìshapeî. And I have also been able to develop corrections for commonly used, but flawed, measures of compactness.

Appendix: Proofs

We can simplify the analysis by recognizing that some measures[62] will produce identical rankings over all shapes, and will be indistinguishable under axioms 1–6.[63]

---

[62] In particular, the following sets are clearly identical: (LW_3, LW_4), (LW_6,A_5), (A_1,OS_3), (PA_1,PA_2,PA_6), and (PA_3,PA_4).

[63] While one measure might be simpler than another, in practice, to compute, I ignore this distinction.

Measures OS_3 and A_1,A_2 and A_3 clearly satisfy axioms 1–3 and 5 thus meeting the axiomatic criteria. However, most of the other measurements violate at least one of the first three, or all of the latter three axioms, raising doubts as to their consistency.

Many of these indices violate at least one of the first three “shape” axioms:

- Measure PA_5 violate axiom 1. Convex districts of exactly the same shape, but different sizes may be assigned different values.[64] All of the perimeter/area measures, PA_1—PA_6, are subject to a more subtle violation of scale invariance in practice, which has not been previously recognized. If districts have natural boundaries, these measures can be affected by precisely how we measure district lines, for districts will seem to be less compact when seen on a map which has a fine scale than on a map with a larger scale.[65] For comparisons to be consistent, we must use the same precision to measure all district lines.

---

[64] Flaherty and Crumplin (1992) note that, in general, that perimeter/area measures are not scale invariant.

[65] Suppose you were trying to measure the length of a section of California shoreline, perhaps the section between San Francisco and Los Angeles. If you used a coarse approximation, perhaps by measuring the length of Route 1, which runs along the shore nearby, you would guess that the shoreline is several hundred miles long. If you tried to make more precise measurements by walking along the beach, your path might expand to several thousands of miles. Finer measurements will reveal the shore to be of ever-

- For any finite number of sample points, chosen at fixed positions along the edge of the shape, OS_2 violates axiom 2, because rotating a shape may change the choice of sample points, and hence the compactness measurement (Young 1988).[66] Measure LW_5 also, by its definition, fails axiom 2.

Most compactness indices reflect at least one principle of shape manipulation, but not others. In most cases, these measures obviously satisfy one shape axiom, but violate others. I demonstrate these violations by producing shapes that are misclassified by particular measures.

- Measures A_1, A_2, A_3, and OS_3 clearly satisfy axiom 5,[67] although they violate axioms 4 and 6 (Figure 2-10, Figure 2-11).[68]

---

increasing length.

[66] Young (1988) does not use an axiomatic characterization, but his example can easily be applied to show the violation. He gives an example where changing the particular sample points used to measure a shape changes its compactness. If, instead, we consider the sample points to be fixed in orientation (e.g., one point sampled at "12-0'clock," "1-O'clock," etc.), and instead rotate the shape, Young's example shows that OS_2 is not rotation invariant. We can fix this rotation invariance by using a rotation-invariant reference point for our samples, but this would simply force the measure to be fail axiom 3.

[67] If two shapes have the same convex hull, the radius and area of the circumscribing

- Measure OS_1 violates axioms 4, 5, 6 (Figure 2-10, Figure 2-12, Figure 2-11).[69]

- While the perimeter-area measures violate scale invariance, we can normalize them to correct this defect, and, at the same time, satisfy axiom 7. One way to correct these violations is to normalize the shape being measured by the area of its convex hull. I define a new measure PA_7 to be: $\dfrac{P\left(\frac{1}{\alpha}S\right)}{A\left(\frac{1}{\alpha}S\right)}$, where $A(Co(S))=\alpha$/ PA_7 satisfies axioms 1–3, and 6. Similar transformations could be used for other measures. All of the perimeter-area measures can violate axiom 5 (Figure 2-10).

---

circle, hexagon or convex figure will be the same. Hence if shape A has the same convex hull as B yet B has a greater area, B will be ranked higher under these measures — satisfying axiom 5.

[68] Under the A_3 measure any convex shape is perfectly compact - contradicting axiom 4. Measure A_2 is created under the assumption than the most compact figure is a hexagon, which leads to a similar violation of axiom 4. While A_1 does agree with axiom 4's implication that the most compact shape is a circle, it can violate the axiom in less obvious ways, as illustrated below.

[69] Blair and Bliss (1967) show that OS_1 satisfies axioms 1–3. They also show that under OS_1 the most compact shape is a circle. While OS_1 fails axioms 1–3, it does seem to be capturing legitimate aspects of shape manipulation; rather than focussing on dissection or dispersion alone, it may be capturing a combination of both.

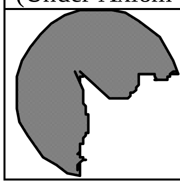
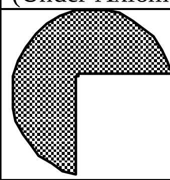
More Compact (Under Axiom 5)	Less Compact (Under Axiom5)	Measures Not Classifying These Correctly
		$PA_1 – PA_6$

Figure 2-10. Violations of axiom 5. Shapes on the left have the same convex hull, but greater area hulls than those on the right.

A number of compactness indices violate all three principles:

- All measures listed violate axiom 4 (Figure 2-11). We can, however, create a measure that satisfies axiom 4, as well as axioms 1-3, $OS_5 = \frac{P(CO(s_{norm}))}{A(S_{norm})}$.[70]
- All with the exceptions of PA_1–PA_6 and OS_2 violate axiom 6, because changes in perimeter that do not affect convex hull, area and shape diameters are ignored.
- Measures LW_1–LW_7, A_4–A_7 and OS_4 violate all three compactness axioms (Figure 2-12, Figure 2-11).

[70] To avoid violating axiom 1, we uniformly scale the shape so that it has unit area, producing S_{norm}.

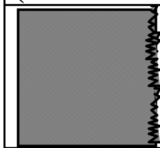
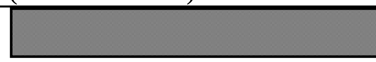
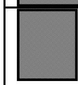
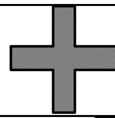
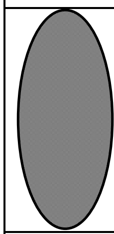
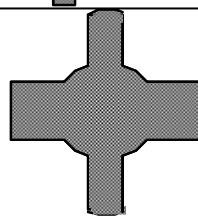
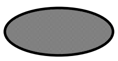
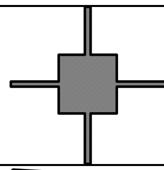
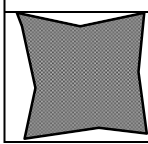
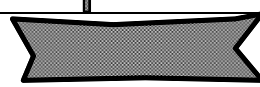
More Compact (Under Axiom 4)	Less Compact (Under Axiom 4)	Measures Not Classifying These Correctly
		PA_1–PA_6
		LW_1–LW_5
		LW_6,LW_7, A_1,A_2, OS_3,A_4–A_7
		OS_1,OS_2[71]
		OS_4,A_3

Figure 2-11. Violations of axiom 4. Shapes on the left have the same area, but smaller convex hulls than those on the right.

---

[71] Remember that this measure uses a sample of points on the perimeter: This particular example works if we take our sample points at the compass points; it is easy to create other examples for other specified sampling methods.

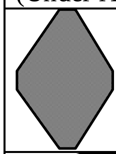
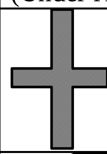
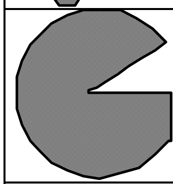
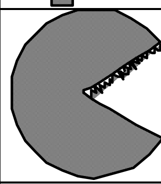
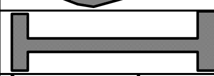
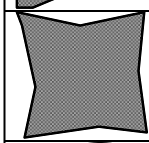
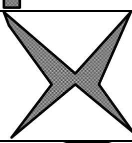
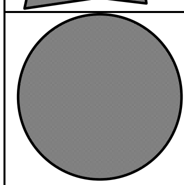
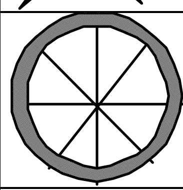
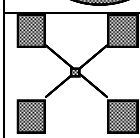
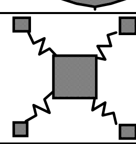
More Compact (Under Axioms 5,6)	Less Compact (Under Axioms 5,6)	Measures Not Classifying These Correctly
		LW_1–LW_5
		LW_6–LW_7, A_4–A_6
		A_7
		OS_4
		OS_2[72]
		OS_1

Figure 2-12. Violations of axioms 5 and 6. Shapes on the left have the same convex hulls, greater area and a smaller perimeter/area ratio than those on the right.

For any compactness measure C, satisfying axioms 1–4 and either 5 or 6, the two properties hold.

---

[72] This example is based on the arguments in Young (1988).

Property 1 (Contiguity): For any given perimeter or area, the maximally compact shape is contiguous. If C satisfies axiom 6, this is true for any given convex hull as well.

Property 2: A circle is the most compact district.

The shape that uniquely minimizes the perimeter for any given area is a circle.[73] Thus any compactness measure satisfying axiom 4 will judge a circle to be most compact, and property 1 is shown.

This fact can also be used to show part of property 1. Because a circle is contiguous, and a circle is the most compact shape, the most compact shape for any given perimeter or area is contiguous.

This leaves me to show that under axiom 6 the most compact shape with a given convex hull is contiguous. I do this by showing that for any given convex hull *X*, the shape s.t. $S^*=CO(S^*)=X$ is most compact. Because this shape is contiguous, property 2 follows.

C satisfies axiom 6.

---

[73] This is a well known isoperimetric inequality. For a compendium with original sources, see (Mitrinovic, Pecaric and Volenec 1989).

Define: A shape *S* is *discontiguous* if $\exists s_i \in S, s.t.\ s_i \cap \left(\bigcup_{s_{j \neq i} \in S} s_i \right) = \varnothing$. Let *S*- $= S \cap CO(S)$. A shape *S* is *measurably* discontigous if it discontiguous and *A(S-)•0,P(S-)•0.*

If *CO(S)=X*, *S•S**, *S* is measurably discontiguous, then A(S)<A(S*). Hence C(S*)>C(S) by axiom 6. Q.E.D. I conjecture without proof that *S** minimizes the perimeter/area ratio for all shapes with convex hull *X*.

Chapter 3. **Traditional Districting Principles: Judicial Myths Vs. Reality**

3.1. Redistricting Principles In The Courts

One person, one vote. With this principle, the Court permanently changed representation in the United States. Equal population requirements changed the face of legislative redistricting in the 1960's when the Supreme Court applied it to congressional districts in *Wesberry* v. *Sanders* (1964) and to state legislatures in *Reynolds* v. *Sims* (1964). Equality in district population was valued not only as instrumental to other goals, but for itself, as Justice Black in *Wesberry* explained: "as nearly as practicable one man's vote in a congressional election is to be worth as much as another's... To say that a vote is worth more in one district than another would... run counter to our fundamental ideas of democratic government" (emphasis added).[74]

As Justice Brennan made clear, the court based its decision in large part on a particular understanding of the historical meaning of the 14th amendment and of article 1, ß2 of the constitution. And as widely accepted as this principle has come to be, it has been subject to severe historical criticism, criticism that has never been resolved. For example, Berger [# 1977] claims that malapportionment was historically present and accepted before and during the creation of the 14th amendment, and hence that the equal protection clause could not have implied the equal population principle (from Chapter 5): "Certainly there was no disclosure that such intrusion (on apportionment) was

[74] This is argued in detail in Lowenstein (1990).

contemplated; there is in fact striking evidence that malapportionment was an accepted practice."

This claim has never been thoroughly examined. Although previous authors have studied the history of apportionment between states (Balinski and Young 1982; McKay 1965; Schmeckebier 1941), studies of the history of apportionment *within* states is limited to isolated states and periods. (See Dixon 1968 for a survey.) And our knowledge of the apportionment of *congressional* districts has been particularly limited (Dixon 1968; Pildes and Niemi 1993; Schmeckebier 1941). In Section 3.4.2 I fill this longstanding gap in the literature, and I address Berger's claim.[75]

In the courts, many types of districts have been under attack, but congressional districts have undergone particularly close recent scrutiny by the Supreme Court. In particular, all of the cases in the last five years in which the Court has particularly empasized compactness and other "traditional" have been cases involving congressional districts. While this does not imply that the Court's statements about such principles

---

[75] Probably the most extensive empirical study of compactness in U.S. Congressional districts is, Pildes & Niemi (1993), should be noted for examining the compactness of districts in the 1980's and 1990's and for proposing a novel legal theory to explain the court's actions in *Shaw*. Neither this study, nor any other I am aware of, systematically studies the compactness of historical districts, the contiguity of districts, or the extent to which districts have followed "traditional boundaries."

exclude other kinds of districts, there are a number of reasons why the Court might treat Congressional districts differently than legislative districts: Legislative and congressional districts have somewhat different legal, historical, and even philosophical traditions. The laws that govern legislative districts have varied over place and time. Many states have required that legislative districts be contiguous, compact, or that they follow county boundaries and other states have required that each county have its own legislative districts. At the same time, Congressional districts have not, for the most part, been subject to such requirements — and even when these requirements were on the books, many questioned whether Congress had the property authority to make them and whether they were enforceable. (See Section 3.4.1.)

These myriad differences often stem from a more fundamental difference between congressional and (some) legislative districts: Congressional districts were written into the Constitution explicitly to provide representation on the basis of population. In contrast, many states' constitutions provided that legislative representation be based upon other non-population principles such as the representation of counties, cities, or other geographical and political units. I have followed the Court's path, and chosen in this chapter to discuss Congressional districts. As a practical matter, as well, records of congressional districts and representation are more complete and accessible than records of legislative districts.

As population equality changed the face of legislative redistricting in the sixties, a new set of principles has the potential to change redistricting in the 1990s. In *Shaw* v. *Reno* (1993) the Court labeled several principles "traditional" and "objective" factors, and indicated that they could serve to defeat racial gerrymandering. These principles were reemphasized in *Miller* v. *Johnson* (1995), in which the Court listed many of these criteria[76]:

> The plaintiff's burden is to show, either through circumstantial evidence of a district's shape and demographics or more direct evidence going to legislative purpose, that race was the predominant factor motivating the legislature's decision to place a significant number of voters within or without a particular district.
>
> To make this showing, a plaintiff must prove that the legislature subordinated traditional race-neutral districting principles, including but not limited to *compactness, contiguity, respect for political subdivisions or communities defined by actual shared interests*, to racial considerations. Where these or other race-neutral considerations are the basis for redistricting legislation, and are not subordinated to race, a state can defeat a claim that a district has been gerrymandered on racial lines.

---

[76] In *Abrams* v. *Johnson* (1997), the court again stressed the importance of traditional districting principles in upholding the districts drawn to replace those invalidated by *Miller*.

More recently in *Bush* v. *Vera* (1996) the court extended and clarified the role of these criteria. Writing the plurality opinion for *Vera*, Justice O'Connor made compactness and regularity[77] particularly important criteria to follow for those who wish to pass strict scrutiny and to avoid plaintiffs' substitute redistricting plans: "A district that is reasonably *compact and regular*, taking into account traditional districting principles such as maintaining *communities of interest and traditional boundaries*, may pass strict scrutiny without having to defeat rival compact districts designed by plaintiffs' experts in endless beauty contests."

How, exactly, are we to evaluate districts by these principles? From which traditions did these principles spring? These opinions offer little guidance. The court neither supports its implicit claim that these particular principles deserve special status, nor provides us with a foundation for deciding in general what principles merit such an appellation.[78] I answer these questions by analyzing historical congressional districts. Before presenting this analysis, I briefly turn to data sources and measures.

---

[77] In this line of opinions, and especially in *Vera*, the court uses "noncompact" to refer to the overall shape of the district, and "regular" to refer to the meanderings of a district's boundary. This differs from the standard terminology in political science, where "compactness" has been used to refer to both properties.

[78] Nor is such a foundation to be discovered in *Gaffney* v. *Cummings* (1973) or *Karcher* v. *Daggett* (1983), the cases to which, on this issue, *Shaw* and *Miller* refer.

3.2. Data Sources

No single source contains geographical and population data for U.S. congressional districts over the entire period from 1789 through the present; for different periods I turned to several different data sources. Data on the geography of election districts is available from several overlapping data series: For election districts used between 1789 through 1912, I extracted geographical data from the *United States Congressional Districts and Data* series (Parsons, Beach and Hermann 1978; Parsons, Dubin and Parsons 1990; Parsons, Beach and Dubin 1986). This data source leaves out maps of some urban districts, so I extracted geographic data from *The Historical Atlas of United States Congressional Districts, 1789-1983* (Martis and Rowles 1982). This atlas contains district maps and lists the political units, typically the counties, cities, and wards, that the district comprises, but does not contain population data.[79]

Parsons' data series ends in 1912, and with it detailed data at the district level until the creation of congressional district data books with the census in 1960. For decadal district population data from 1913 through 1953, I used the figures in the *Congressional Directory* (Joint Committee On Printing 1913; Joint Committee On Printing 1923; U.S. Government Printing Office 1933; U.S. Government Printing Office 1943; U.S. Government Printing Office 1953). For the period 1963 through 1994, I extracted geographic and population data from the Congressional Quarterly's publications:

---

[79] The district maps are somewhat less detailed than the maps in Parsons, as well.

Congressional Districts of the United States (Congressional Quarterly 1964), *Congressional Districts in the 1970's* (Moxley, Walker and Healy 1974), *Congressional Districts in the 1980's* (Congressional Quarterly 1983), *Congressional Districts in the 1990's* (Congressional Quarterly 1993).[80]

I extracted the tabular data using an optical character recognition system, in addition to entering data manually.[81] Extracting compactness data from the district maps was more complicated: First, I digitized each district map using an optical scanner.[82] Second, I used image-processing software[83] to identify the boundaries of each district and to

---

[80] These books are based largely on data in the United States Census's *Congressional District Data Book* and *Congressional District Atlas* for the relevant period. This series leaves out maps for the 1963 districts, so I used the maps in (Martis and Rowles 1982).

[81] I used the commercial character recognition package Omnipage 3.0. All numerical data was independently double-checked to ensure correct entry.

[82] I used a HP-Scanjet III optical scanner operating at various resolutions ranging from 150–600 d.p.i. I used the higher level resolution when maps were particularly small and finely detailed.

[83] I used the software package NIH-Image (version 1.6), a program in the public domain developed by the National Institute of Health especially for mathematical analysis of two-dimensional digital images. This program has built-in routines that remove noise

estimate its geographical properties. Third, I used image analysis software to apply standard mathematical formulas (described in Section 3.1) that calculate compactness scores.

3.3. Evaluating Districts: Compactness And Population Measures Compared

3.3.1. Quantitative Measures of Malapportionment and Compactness

There are a number of different methods to measure malapportionment used in the scholarly literature and by the courts. The most popular early measures of population variation were the difference between the largest and smallest districts (divided by the mean), the population variance ratio, which is the ratio of the largest to the smallest district, the maximum (or average) percent deviation from the mean, and the electoral percentage, which is the minimum percentage of the population represented by a bare majority of seats.[84] Later court cases have tended to stress the difference between the

---

from images, that automatically identify the outlines of selected shapes (districts in this case), and that measure perimeter, area (etc.) of a selected shape. It was necessary, however, to guide the program in its selection of districts, and to correct defects in the district maps, such as boundary lines that were obscured by text markers or map symbols, boundary lines that overlapped solely because of line-weight, and the like.

[84] See Dixon (Dixon 1968) Chapter 17, Section 4 and Chapter 18, Section 2.

largest and smallest districts, and this measure was emphasized in *Karcher* v. *Daggett* (1983). These measures have a number of theoretical faults, and Foster (1985) argues cogently that such measures as the Gini index, Theil's measure of entropy, and the coefficient of variation have more desirable theoretical properties (Foster 1985).

For this data, all of these methods give very similar evaluations of districts. For the majority of this chapter, I use a common and easily understandable measure, the population coefficient of variation.[85]

Contiguity is the most often mentioned geographic principle. A simple idea in theory, it is less so in practice. In the context of redistricting, contiguity is meant to be a signal of the political manipulation of districts, not just a formal and accidental property of district shapes. If we are to use contiguity in this fashion, two hurdles[86] must be overcome. First, mathematical contiguity does not reflect a constraint on electoral manipulation, as any given noncontiguous district (or set of districts) can be made

---

[85] This is the standard deviation divided by the mean. Think of it as a measure of the average deviation.

[86] Another approach to making contiguity practical is to examine the costs of traveling and communicating in the district. Under this approach, for example, a district would be considered non-contiguous if it were divided by an impassible mountain. While this approach has merit, the historical data is not rich enough to consistently apply it.

contiguous by adding arbitrarily thin connecting lines, without materially changing the results of an election held in that district (Sherstyuk 1993). Second, breaches of contiguity may be difficult or impossible to avoid because of geographic obstacles, such as large bodies of water, and such non-contiguities occur in the absence of any political manipulation.

To overcome these hurdles, I divided districts into three categories in order of divergence from real-world contiguity: practically contiguous, questionably contiguous, and non-contiguous. All districts that are formally contiguous, or that only deviate from contiguity because of islands off the coast of the district, I put in the first category. Into the second category I put districts that were otherwise contiguous but that contained islands that were not directly off the coast of the district, districts that were non-contiguous but could be connected by straight bridges, and districts that were connected only by "points."[87] In the non-contiguous category, I put all other violations of formal contiguity (Figure 3-1).

---

[87] More formally, I classified a district as questionably contiguous when more than ten percent of the district's area was connected to the rest of the district by a passageway no longer than one percent of the district's length.

Figure 3-1. Three Odd District in Early New York Congressional Districting Plans. Part A shows the plainly non-contiguous fifth district in New York's first (1788) congressional districting plan. Part B shows the seventeenth district in the thirteenth (1812) Congressional district plan; this plan is of questionable contiguity because it is connected only at a single point. Part C shows district two in the twenty-third (1832) Congressional districting plan. The light shading shows areas covered with water. This district is of questionable contiguity because the island portions of the district are not joined to the nearest mainland district.

Breaches of "traditional boundaries" are even less often subjected to formal measurement. City and county boundaries, although often referred to by the courts as "traditional boundaries" are, at least at times, political boundaries, subject to change through the political process. Yet one suspects that there is some truth to the courts' distinctions in this case, that city and county are less manipulable than districts, or that, at least, because of the extent to which these boundaries affect local government, these lines are manipulated for different purposes than are election districts. On this intuition, districts were placed in several categories. Districts “followed” traditional boundaries if they were composed of lasting independent political units: whole counties with the addition or subtraction of whole towns, cities, parishes, boroughs, or townships. Districts “split” traditional boundaries if they were defined explicitly in non-political terms, namely roads, streets, and (after the 92nd congress) census blocks and tracts. With a handful of exceptions, the remaining districts were classified as “questionable,” with a subcategory for those districts splitting county and town boundaries only to follow assembly district lines.[88] This categorization is admittedly rough, but the overall patterns

---

[88] And, from the 93rd congress, “split” districts used census blocs and tracts. Also, in practice, districts that were “split” only along (explicitly defined) natural boundaries were quite rare: New York’s fifth district in the first congress mentions the Hudson (but the county may not have been well defined), the third and fourth districts in Maryland were divided by the Monocacy River from the 3rd through 22nd congresses, and two districts around Pittsburgh, PA, were also split along several rivers during most of the Congresses from the 33rd through the 67th. These districts were put in the “followed traditional boundaries” category, while two districts in South Carolina’s fifty-third congress that split one town along railroad tracks was put in the “questionable” category.

in the data are clear enough that changes in the categorization would not change any of the conclusions in this chapter.

The literature contains many more ways of measuring the compactness of a set of districts than it does for measuring the malapportionment in those district. As I showed in Chapter 2, geographically-based compactness measures fall into three rough categories: measures that compare the perimeter of a district to its area, measures that compare the length of a district to its width, and measures that compare the area of a district to the area of an idealized shape that encloses the district. I use three populaAs we measures that could be reasonably computed from the available historical data, selected from among these categories (Table 3-1).[89]

[89] Compactness measures can also be computed based upon population distribution instead of geographical distribution (See Chapter 2, and Niemi, et al. 1991 for a survey). State laws, constitutional provisions, and court cases, however, stress the geographical measures almost exclusively.

Name	Measurement
Normalized Area/Perimeter (Norm)	$A/(0.282P)^2$
Area of Circle (AC)	The ratio of the district area to area of minimum circumscribing circle (Normalized to the [0,1] interval.)
Length/Width (LW)	The length of the minor axes/major axes for the best fitting ellipse.[90]

Table 3-1. Selected compactness measures.

3.4. Historical Patterns

3.4.1. Congressional and State and Regulation of Congressional Districts

For many years, judges consistently refused to enforce state provisions designed to control redistricting. Although willing to hear an initial flurry of challenges in the 1890s, state courts universally failed to provide any positive remedy. Later attempts to enforce

[90] Measures that compare length and width are common (See Niemi, et al. 1990.) but these measures tend to be overly influenced by outlying points and are not necessarily orientation independent (Young 1988). By fitting an ellipse to the shape and measuring the axes of the ellipse both of these problems are reduced:

I calculate the best fitting ellipse by using the 'ellipse of concentration' (See Cramer (1946)) which equates the second order central moments of the ellipse to those of the distribution of points in the district, and then adjust the resulting ellipse slightly so that it has the same area as the district being measured.

redistricting rules in the courts also met with failure, culminating in *Colgrove* v. *Green*, in which the Supreme Court declared districting to be a non-justiciable "political question." (See Chapter 1 in Cortner 1970.) At the time these rules were created, usually in the 19th century, however, the courts' future direction had not been foreseen. And as we shall see in this section and in Section 3.4.3, some regulations on districts were effective, if only for a short time. If population equality, contiguity, or compactness are traditional districting principles, we should expect to see them in the laws and or congressional debates of the time.

Before 1842, there were no laws governing the construction of congressional districts.[91] In 1842, Congress passed the first such law, which required that all states use contiguous single-member districts (Table 3-2). State regulation of congressional districts followed soon after, but was limited to two states in this period (Table 3-3).

Starting Year	*Congressional District Requirements*
1842	Single Member Districts, Contiguity
1850	No provisions
1862	Repeated 1842 Provisions
1872	Added "practicable" Population Equality
1881	Repeated
1891	Repeated
1901	Added Compactness
1911	Repeated

Table 3-2. Congressional Redistricting Laws, 1789-1913.

Source: Schmeckbeier (1941), ch. 9.

[91] Between 1816 and 1826 there were a number of attempts to pass an amendment requiring congressional districts, see Schmeckbier 1941, pgs. 131-131.

Starting Year	State	Requirement
1849	California	Multi-County Districts Must Be Constructed of Contiguous Counties
1872	West Virginia	Contiguity and Compactness

Table 3-3. State Constitutional Governing Congressional Redistricting, 1789-1913. Based on data from: McKay (1965), State Summaries Appendix.

Although congressional districts were substantially unregulated, state *legislative* districts were often subject to a number of rules. As Table 3-4 shows, many states were apportioned on a county basis, or had provisions against splitting counties, and many others had contiguity requirements (Table 3-4). Perhaps these requirements for state legislative districts indirectly affected congressional districts, or perhaps they reflected the norms of the time, since despite the absence of official regulation, most congressional districts were contiguous (as we shall see in Section 3.4.3); and, with the exception of districts in large urban areas (See the appendix to this chapter.), most congressional districts during this period were composed of whole counties (Table 3-4).

Method of Apportionment of Legislative Districts	*States*[92] (Start of Provision)	*Restrictions for at least one house* (Date, if different from column 2)
Entire Legislature Apportioned by Counties	Delaware (1787), Georgia (1788), New Jersey (1787), North Carolina (1789), Wyoming (1890)	*Contiguity:* *Compactness:*
At Least One House of Legislature Apportioned Primarily by Counties	Maryland (1788), Montana (1889), New Jersey (1844), North Carolina (1835), South Carolina (1865), Virginia (1788)	*Contiguity:* North Carolina (1868) *Compactness:*
Apportionment Based Primarily on Cities or Other Geographical or Political Units	Connecticut (1788), Rhode Island (1790), South Carolina (1788), Virginia (1830), Vermont (1788),	*Contiguity:* Virginia (1902) *Compactness:* Rhode Island (1842)

[92] Connecticut apportioned its lower house by cities and town, but its upper house was elected at-large. Before 1840, Massachusetts apportioned its upper house on the basis of taxes. New Hampshire based its upper house on taxes and lower house on the number of “ratable polls.” In 1788, New York added provisions to apportion by population and area in the upper house and to guaranty representation for some counties in the lower house. Prior to 1790, Pennsylvania elected a unicameral legislature on the basis of cities and counties. In 1874, it adopted compactness and contiguity for cities that were split into at least four districts.

Population Exclusively	California (1850), Colorado (1876), Illinois (1818), Indiana (1816), Iowa (1846), Minnesota (1858), Nebraska (1867), Nevada (1864), Ohio (1803), Oregon (1859), South Dakota (1889), Texas (1845), Washington (1889), Wisconsin (1848)	*Protected County Lines:* California, Colorado, Indiana (1851), Minnesota, Texas (1876), Wisconsin *Contiguity:* Nebraska, Texas (1876), Washington, Wisconsin *Compactness:* Nebraska, Wisconsin
Primarily Population	Alabama (1891), Arkansas (1836), Florida (1845), Idaho (1890), Kansas (1861), Kentucky (1792), Louisiana (1812), Maine (1820), Massachusetts (1788), Michigan (1837), Mississippi (1817), Missouri (1821), New Hampshire (1788), New York (1788), North Dakota (1889), Pennsylvania (1790), Tennessee (1796), Utah (1896), West Virginia (1863)	*Protected County Lines:* Arkansas (1874), Idaho, Kentucky (1799), Mississippi (1831), New York (1894), North Dakota, Tennessee, Utah, West Virginia (1873) *Contiguity:* Arkansas (1868), Louisiana (1868), Massachusetts (1857), Mississippi (1831), Missouri, New York (1846), Utah, West Virginia (1873) *Compactness:* Arkansas (1868), Missouri, New York (1894), West Virginia (1873)

Table 3-4. Historical Provisions For State Legislative Districts. Start

Source: (1965), State Summaries Appendix.

Congress first passed regulations governing congressional districts as part of the apportionment law of 1842. This law, in addition to assigning seats in congress to each state, specified that all members of congress were to be elected from single-member districts, in effect banning at-large elections.

Three topics occupied the bulk of the debate over this law on the floor as recorded in the *Congressional Globe* (Congressional Globe 1842, pgs. 435–7, 445–7, 452–4, 526–32, 583–5, 601). First, unsurprisingly, was the question of how many seats each state should receive.[93] Second was whether Congress had the authority to mandate single-member districts. Last followed a debate over whether at-large elections unfairly increased the influence of large states in congress by providing the majority parties in those states with a large electoral bonus. In the midst of these debates, the contiguity provision seems to have been generally accepted without mention, and there was little concern expressed over the subject of gerrymandering.[94]

The effect of the congressional mandate for contiguous single-member districts was swift. In the 23rd congress, prior to the districting legislation, 20 percent of representatives were not elected from single-member districts; whereas immediately after

---

[93] Debates over the size of the house, the method of fractions to be used to distribute seats, and the number of seats given to individual states recurred regularly in every apportionment debate that I examined, from 1842 to 1911. Balinski and Young (1982) cover the history and principles of apportionment between states quite thoroughly, and I shall not pursue it here.

[94] The one mention of a gerrymander in the records of the floor debates is a hypothetical and hyperbolic rhetorical question asking whether if congress could mandate single member districts, and why could it not then mandate particular gerrymanders.

the legislation, non-district representatives dropped to nine percent of the total, and dropped further to 1 percent by the 31st congress (Table 3-5). This legislation was not entirely effective, however. Despite requirements in later apportionment acts for single member districts through the 67th congress, at-large elections were used in some districts, though in reduced percentages (Table 3-5).

Congress	*States Deviating from Single Member Districts*	*Percentage of Representatives Not From Single Member Districts*
1	4 / 13 = 31%	20 / 65 = 31%
3	7 / 15 = 47%	46 / 105 = 44%
8	8 / 17 = 47%	40 / 142 = 28%
13	10 / 18 = 56%	63 / 182 = 35%
18	9 / 24 = 38%	55 / 213 = 26%
23	9 / 24 = 38%	49 / 240 = 20%
28	4 / 26 = 15%	21 / 231 = 9%
33	2 / 31 = 6%	3 / 234 = 1%
38	2 / 23 = 9%	4 / 184 = 2%
43	11 / 37 = 30%	19 / 292 = 7%
48	9 / 38 = 24%	16 / 325 = 5%
53	5 / 44 = 11%	9 / 356 = 3%
58	6 / 45 = 13%	10 / 386 = 3%
63	5 / 48 = 10%	8 / 435 = 2%
68	1 / 48 = 2%	2 / 435 = 0%
73	12 / 48 = 25%	54 / 435 = 12%
78	9 / 48 = 19%	13 / 435 = 3%
83	5 / 48 = 10%	7 / 435 = 2%
88	8 / 50 = 16%	17 / 435 = 4%

Table 3-5. Percentage of Districts and States Deviating From Single-Member Districting

Although the Court relied partially on the 14th amendment for authority to regulate malapportionment, there is little reference to malapportionment in the debates that surrounded the amendment (Avins 1966).[95] Close analyses of these debates conclude that the 14th amendment was meant generally to be “open textured,” and nothing in the record precludes its application to malapportionment, nor is there explicit evidence that it was meant to encompass malapportionment (Kelly 1965; Van Alstyne 1965).

When, in the apportionment following the passage of the 14th amendment, “practicable” population equality was added to the requirements for districts there was much debate over what this amendment required of apportionment. Although the record of floor debate touches neither on gerrymanders nor on the new population equality requirement for districts, it shows a vociferous argument over whether explicit provisions should be added to the apportionment law to enforce it by determining the number of qualified voters that were denied suffrage in each state and reducing suffrage accordingly.[96] (Congressional Globe 1871, pgs. 64–66, 78–84, 105–112, 608) Member

[95] The Court relied upon the 14th amendment to regulate state legislative districts, in *Reynold* and upon Article 1, ß 2 of the Constitution in deciding *Wesberry*, which prohibited malapportionment in congressional districts. The appellants in *Wesberry*, however, in their briefs before the court, argued much of their case on 14th amendment grounds — particularly on the grounds that widespread and extreme malapportionment violated the equal protection clause.

[96] Section ß2 of the 14th Amendment reads in part: “But when the right to vote at any election... is denied... or in any way abridged... the basis of representation therein shall be

of Congress Ullyses Mercur, a Pennsylvania Republican, expressed the sentiment of those in favor of such a provision in this statement: “This 14th amendment, like many other parts of the Constitution, does not enforce itself. It required legislation in order to give practical effect to its terms. Now, I take it upon myself to say that Congress has passed no law calculated to give effects to the terms and restrictions of this 14th amendment” (page 78).

In the apportionment legislation of 1901, congress added “compactness” to the list of requirements for districts, and compactness seems to have entered for the first time into the record of congressional floor debates.[97] Although mentioned in state constitutions as early as 1821, compactness was never formally defined, either in state constitutions or in the 1901 and succeeding apportionment bills. There was a short debate on the house floor over the compactness clause (Congressional Record 1901, pgs. 605–6, 647–9), which centered around whether it could be measured. An excerpt from this debate shows the purpose of the compactness clause and its limitations (pg. 605):

---

reduced in the proportion which the number of such male citizens shall bear to the whole number of male citizens twenty-one years of age in such state.”

[97] In the floor debates of 1882, Representative Beltzhoover complained of “dumbbell” shaped districts, and claims that contiguity and population equality are not sufficient to prevent this abuse. But he does not call for compactness by name (Congressional Record 1882, pg. 1603).

Mr. Rixey (Maine): I want to ask a question in regard to the phraseology of the bill... The bill... provides that the district "shall be composed of contiguous and compact territory." The words "and compact" seem to be added in this apportionment bill for the first time in the history of apportionment... Now, what I want to know is, who is to be the judge as to when districts are sufficiently compact?

Mr. Kluttz: (North Carolina) I admit the force of the gentleman's question, and that it has never been in an apportionment before, so far as I know.

Mr. Rixey: But what I want to ask is, who is to be the judge as to when a district is sufficiently compact?

Mr. Kluttz: The language has heretofore been "contiguous." When the committee discussed it, I will say that the word "compact" was added at the suggestion of one of the members, the object of it being to prevent shoe-string[98] districts;...

---

[98] "Shoe-string" districts probably referred to the anti-black Mississippi Congressional districts of 1883 and 1893. See Kousser (1992) .

Representative Claude Kitchin, North Carolina, Democrat, objected, "This committee amendment proposes to put in the words 'and compact,' which, I submit, is unwise as well as unauthorized by the Constitution, because 'compact' may be liable to various constructions and become the cause of great confusion hereafter. Disappointed and defeated candidates, ever ready to complain, may base contests upon the shape of their districts and give the House an opportunity to unseat the successful candidate, and opportunity is often deemed duty."[99]

Further debates of the time left the question of how to measure compactness unresolved. Despite this, the provision was adopted in the 1901 apportionment act. From 1901 to 1929, although congress passed one reapportionment law and held five hearings[100] on apportionment, neither compactness nor malapportionment received

---

[99] In fact, three recent supreme court redistricting cases, *Miller*, *Vera* and *United States* v. *Hays* (1995), included defeated candidates for congress among their plaintiffs.

[100] Many early committee hearings were not recorded and there is no official written record of congressional hearings on apportionment prior to 1915 (Congressional Information Service 1980).

significant attention.[101] When, after a decades delay, congress finally passed an apportionment act in 1929, district criteria were dropped without discussion.

In summary, legislative history of both the 14th amendment and of subsequent apportionment legislation is agnostic on the subject “traditional” districting criteria. The lack of debate can equally be interpreted as consensus or unconcern. It is possible that such criteria were commonly accepted as implicit in fair representation under the 14th amendment, and were included in the apportionment legislation because most recognized that this amendment was not self-executing. But it is also possible that such criteria were regarded as minor, unnecessary, expeditions into the control of districts. Was there a general principle of equal apportionment and compactness operating at this time? As the legislative record is not definitive, we must look for these principles by examining the districts of that time. In the next two sections we will examine historical patterns of malapportionment, contiguity and compactness.

3.4.2. Regional Patterns of Malapportionment

Figure 3-2 presents a graph of state congressional malapportionment, weighted by the number of congressional districts, for the period 1789-1963. To my knowledge, this is the first time that these figures have appeared in print. Malapportionment improves dramatically with the second apportionment (3rd congress), decreasing and converging

---

[101] Although in 1915 there was a hearing before the committee on elections, in the House, on a bill in favor of proportional representation.

around the time of the Civil War and reaching an overall low-point at the time of the 43rd congress, at the time of the first redistricting after the passage of the 14th amendment. After that war, regional malapportionment remained stable and then gradually got worse from after 1903 through 1943 (Figure 3-2).

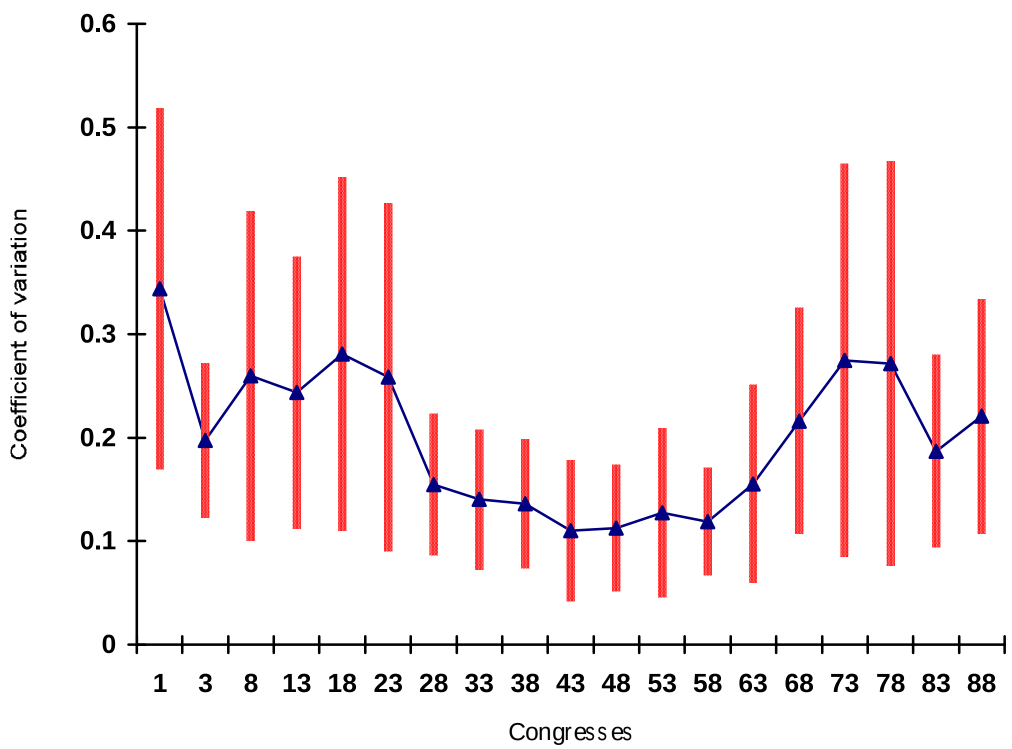


Figure 3-2. Malapportionment Across Time: The horizontal line shows average state malapportionment, weighted by the number of districts in the state. Vertical lines extend for one standard deviation from each of the sample averages.The coefficient of variation is used to measure malapportionment. Malapportionment of the white population is shown prior to the forty-third

congress (1873), and malapportionment of the total population is shown from that congress forward. These malapportionment measures are computed by state and averaged over ICPSR standard regions.

Figure 3-3 and Figure 3-4a–h show malapportionment by region[102] for this same period. Clearly, malapportionment varied significantly across regions and over time.

---

[102] I use the ICPSR standard region categories: The *New England* region comprises Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island and Vermont. *Middle Atlantic* comprises Delaware, New Jersey, New York, and Pennsylvania. *East North Central* comprises Illinois, Indiana, Michigan, Ohio, and Wisconsin. *West North Central* comprises Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, and South Dakota. *Solid South* comprise Alabama, Arkansa, Florida, Georgia, Louisiana, Mississippi, North Carolina, South Carolina, Texas, and Virginia. *Border States* comprises Kentucky, Maryland, Oklahoma, Tennessee, and West Virginia. *Mountain States* comprises Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah and Wyoming. *Pacific States* comprises California, Oregon and Washington. *External States* comprises Alaska and Hawaii.

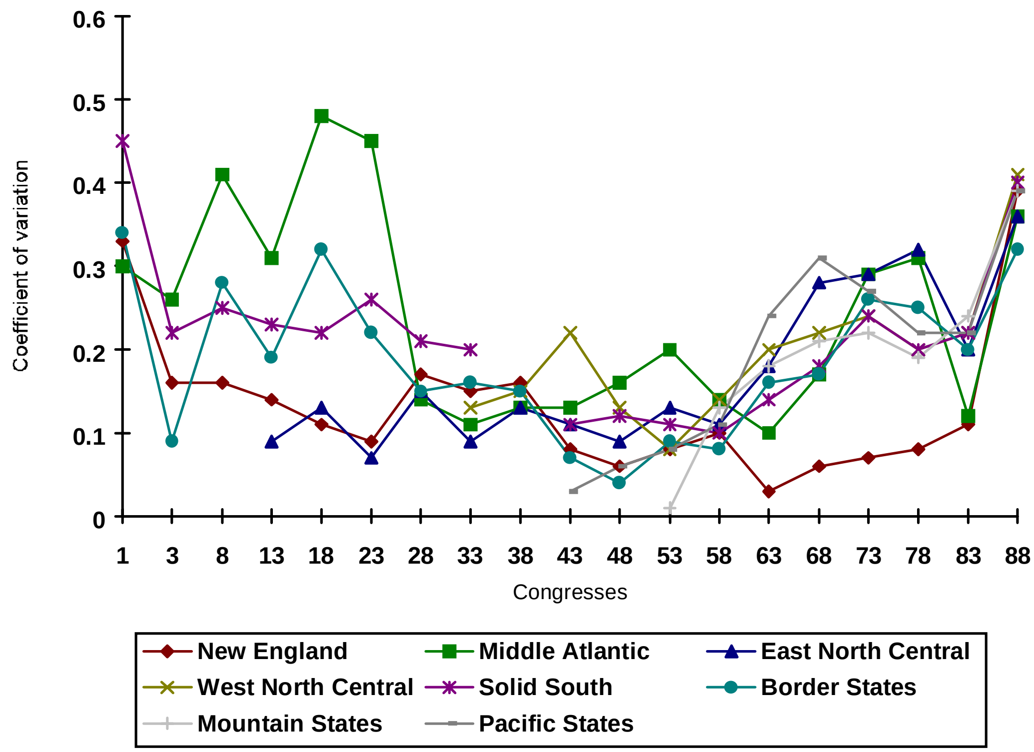


Figure 3-3. Malapportionment Across Time and Region: The coefficient of variation is used to measure malapportionment. Malapportionment of the white population is shown prior to the forty-third congress (1873), and malapportionment of the total population is shown from that congress forward. These malapportionment measures are unweighted averages over ICPSR standard regions.

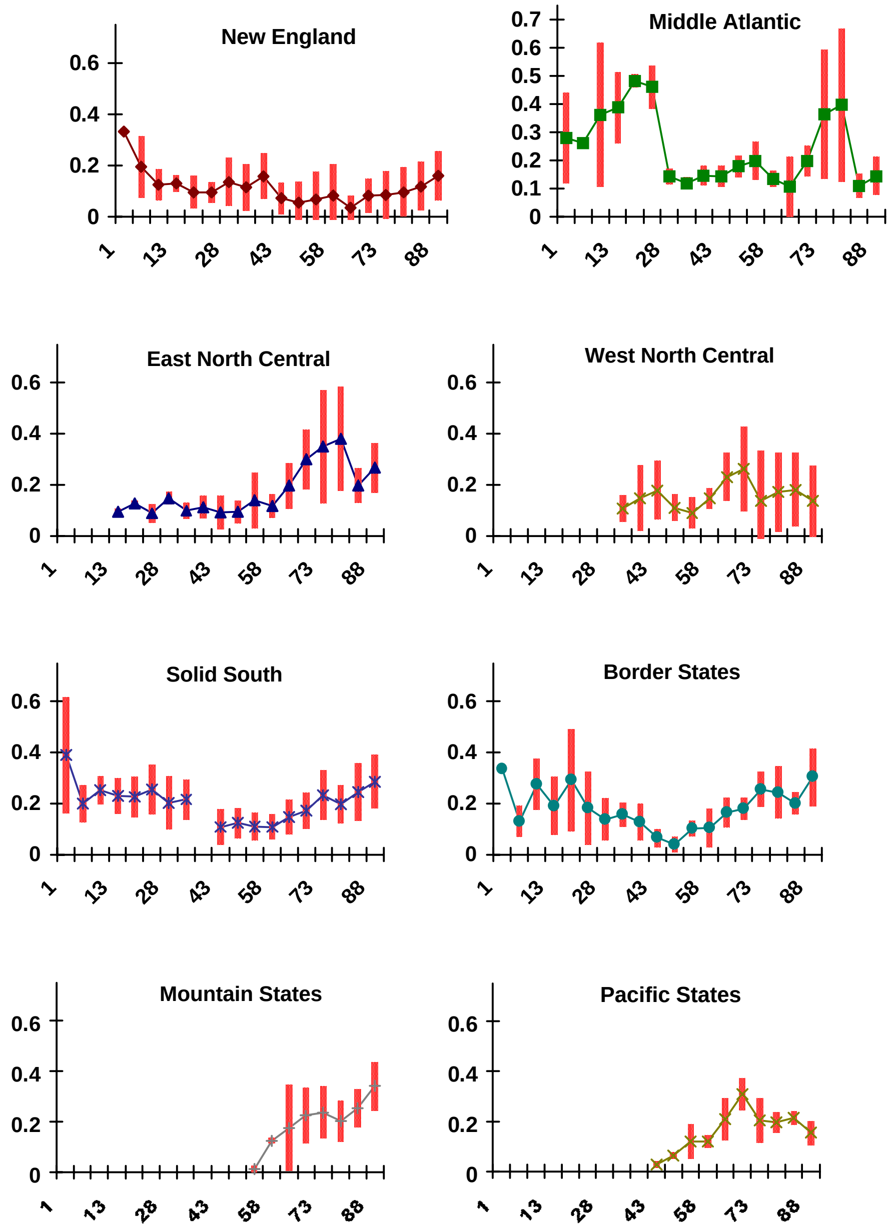


Figure 3-4. Regional Variation in Malapportionment: Malapportionment is shown as in Figure 3 except that regional lines are *weighted* averages over ICPSR standard regions. Vertical bars show plus/minus one standard deviation.

Regional malapportionment, with the brief exception of the West North Central region,[103] seems to follow the same trends in time as did the country as a whole, decreasing dramatically with the second apportionment (3rd congress), decreasing and converging around the time of the Civil War, and remaining relatively stable until after the 58th congress (1903). Note that the congressional requirement for population equality between congressional districts came after malapportionment had already dropped considerably. Malapportionment then became increasingly worse and increasingly divergent through the 78th congress (1943), but in most regions malapportionment improved greatly in 1953[104] and then worsened only somewhat afterwards.

---

[103] Malapportionment in the "West North Central" region does rises in this period. But this is due primarily to the introduction of Kansas into the Union only two years earlier. The regional mean, excluding Kansas, is approximately 0.14, bringing it down amongst the other regions.

[104] Much of this improvement came from improvements in New York, Illinois, and Ohio — all large states that underwent a dramatic improvement in equality of apportionment between the 78th and 53rd congress. These states accounted for more than 20% of congressional districts, and congressional district lines in these states remained essentially unchanged for several decades prior to and including 1943, and were radically redrawn afterwards. In New York, Republicans won the Governor's seat, and took advantage of completing their control of the state government by passing a Republican

In both cases of major legal sanctions for equal population apportionment, the 14th amendment in 1868 and *Wesberry* v. *Sanders* in 1964, then, legalization followed and exacerbated changes already begun in practice. This strongly suggests that equal population requirements, whether effected by constitutional action, or judicial decision, was in accord with political changes in the society, which, no doubt, made the reception of these standards much easier. As we shall see, this was markedly not the case with the compactness language in *Shaw* v. *Reno* and its progeny, which may well portend difficulties in implementing these judicial mandates.

shows that malapportionment differed significantly by region, especially prior to the Civil War and following the 78th congress. The worst malapportionment occurred in the Middle Atlantic states[105], although the South takes this title for several decades centering around the time of the Civil War. The Northwest Territories ("East North Figure

---

gerrymander (Tyler and Wells 1962). Prolonged periods of divided government, exacerbated by urban-rural splits, may explain redistricting in Ohio and Illinois, as well (Jewell 1962).

[105] The figures in this region are driven by New York, to a large extent. New York had many districts, so it weighs heavily in the region, and had some of the worst extremes of malapportionment in the U.S. The worst of the worst malapportionment came from New York, including the plans of the 18th and 78th Congresses that are shown in Figure 3-6.

3-3Central" in Figure 3-2) showed the least malapportionment for the antebellum period, while New England showed low malapportionment overall. This figure also sheds light on the effect of the Northwest Ordinance (1787) on representation (See Section IV.2.a in Dixon, 1968). Article 1 of the ordinance stated that "The inhabitants of the said territory shall always be entitled to... proportionate representation of the people in the legislature." This article seems to have had an effect, as congressional representation in those states is generally divided more equally than anywhere else in the country.

Figure 4 shows average levels of malapportionment. What about the most extreme cases? Figure 5 shows the ratio of the largest to smallest districts (population variance ratio) from the most extremely malapportioned plans in each decadal reapportionment. In effect, it shows the worst of the worst (Figure 3-5).

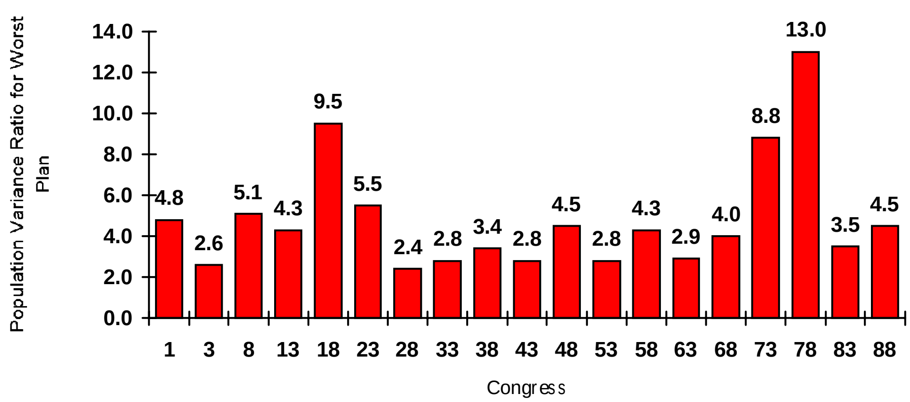


Figure 3-5. Worst plans, by the population variance ratio.

Figure 3-5 supports our conclusion that malapportionment generally decreased over time, and reached a low point around the time of the 43rd congress. In addition, we can see that even the extremes in congressional malapportionment were relatively mild when compared to malapportionment at the state level. For example, while the worst offender of population equality in the 43rd congress had a population variation ratio (p.v.r.) of 2.8, the Florida state senate for that time had a p.v.r. of 73.7[106] (Dixon 1968), which was more than seven times worse than any congressional plan, in any state, during the entire period of 1789-1963.

Critics of the decisions in *Wesberry* and *Reynolds* point to the fact that political districts were malapportioned both at the creation of the constitution, and at the time of the 14th amendment. This fact, while true, ignored the degree of malapportionment during these periods.

The weighted average of state malapportionment for the 88th congress was 0.22, exceeding that of the third congress (0.20), and exceeding by far the average malapportionment of 0.11 during the 43rd congress. At the level of individual states' plans, 24 out of the 31 states were worse in the 88th congress than during the 43rd, but only 3 out of 8 states were worse during the 3rd congress than in the 88th.

---

[106] Even Florida's malapportionment is dwarfed by California's population variance ratio of 422.5 in 1962, before *Reynolds* (Dixon 1968).

Table 3-6 compares malapportionment by region at the time of the 14th amendment to malapportionment just prior to *Wesberry* and *Reynolds.* By most measures, malapportionment in state congressional delegations was much greater at the time of the 88th congress than at the time of the 43rd congress. Comparing Table 3-6 ,Figure 3-4 and Figure 3-5 shows that malapportionment during the 88th congress was worse than that directly following the creation of the constitution. Also telling are the extremes of malapportionment. Columns 3 and 4 in Table 3-6 show the average ratio of largest to smallest districts. Figure 3-5 shows the extremes for each period. The extremes of congressional malapportionment were still considerably below the unprecedented levels of two decades earlier, but were greater than they had been at the time of the 14th amendment, and on the rise from the previous decade (Table 3-6).

Region	(max-min)/mean		P.V.R.		Gini		Coeff. of Variation	
	43rd	88th	43rd	88th	43rd	88th	43rd	88th
New England	0.19 (0.11)	0.40 (0.14)	1.22 (0.14)	1.59 (0.24)	0.03 (0.02)	0.05 (0.01)	0.07 (0.049)	0.12 (0.01)
Middle Atlantic	0.56 (0.13)	0.57 (0.13)	1.76 (0.19)	1.83 (0.23)	0.08 (0.03)	0.08 (0.03)	0.14 (0.027)	0.14 (0.06)
East North Central	0.30 (0.15)	0.88 (0.24)	1.39 (0.25)	2.47 (0.51)	0.05 (0.03)	0.12 (0.03)	0.09 (0.05)	0.23 (0.07)
West North Central	0.46 (0.18)	0.32 (0.17)	1.67 (0.42)	1.42 (0.36)	0.09 (0.04)	0.07 (0.044)	0.18 (0.10)	0.15 (0.14)
Solid South	0.28 (0.13)	0.96 (0.40)	1.34 (0.19)	2.73 (0.94)	0.05 (0.03)	0.15 (0.06)	0.11 (0.06)	0.29 (0.10)
Border States	0.19 (0.06)	0.81 (0.26)	1.22 (0.09)	2.34 (0.58)	0.04 (0.01)	0.15 (0.05)	0.07 (0.02)	0.30 (0.10)
Mountain States		0.63 (0.27)		2.12 (0.82)		0.14 (0.05)		0.34 (0.08)
Pacific States	0.05 (0.0)	0.68 (0.03)	1.05 (0.0)	1.95 (0.01)	0.01 (0.0)	0.08 (0.01)	0.03 (0.0)	0.16 (0.04)

Table 3-6. Malapportionment at Time of *Reynolds* compared to Malapportionment at time of 14th Amendment (weighted by number of districts)

3.4.3. Geographical Criteria

While it seems that population equality at the congressional districting level was "traditional" by the time of the 14th amendment, plans preceding and at the time of *Reynolds* were considerably less egalitarian. That is, the empirical evidence buttresses the notion that population equality had become a norm by the 1860s, not a notion far outside the experience of the 14th amendment's framers.

How about compactness and contiguity? Were they strongly grounded in the American experience of redistricting? In fact, the case for geographic norms is less clear than that for population equality, and as I showed in Section 3.3, different mathematical measurements of geographical criteria may lead us to different conclusions. So we shall examine each in turn, starting with contiguity (Figure 3-6).

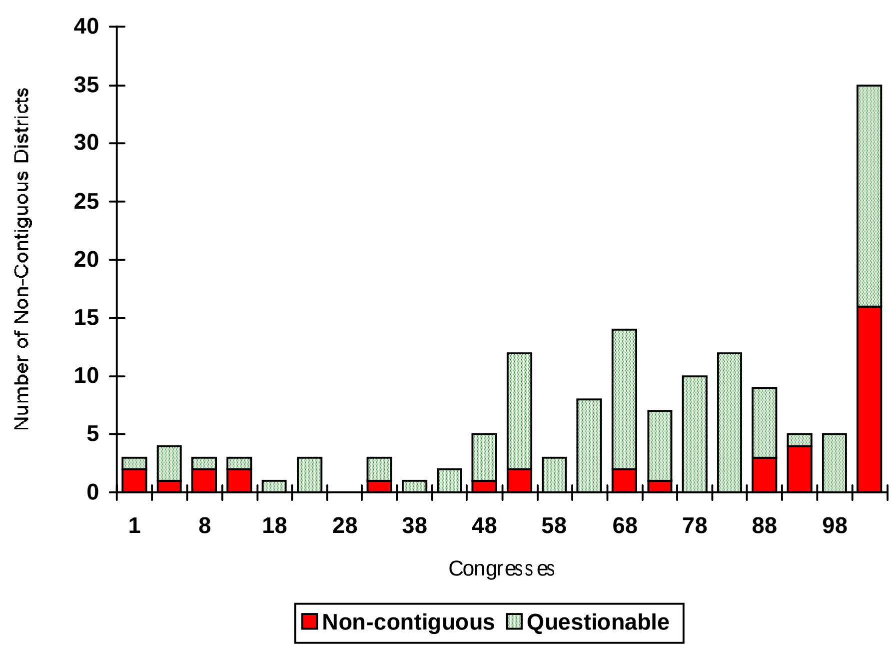


Figure 3-6. Total number of non-contiguous and questionably contiguous districts in each decadal redistricting. Questionable contiguity is evaluated as described in Section 3.1.

Was contiguity always followed in early congresses? The strict answer to this question is a clear negative; the first four decadal redistrictings all had at least one non-contiguous district, and with one exception every decadal redistricting between 1789 and 1913 contained at least one district of questionable contiguity. While congressional requirements for contiguity in the 28th and the 38th–58th congresses seem to have had an initial effect, there were still many non-contiguous districts or questionable districts in most of the decadal redistricting, despite these requirements.

On the other hand, most of the non-contiguous districts were concentrated in a few states; of the 43 questionable or non-contiguous districts in the decadal redistrictings of this period, 16 belonged to New York, 7 to South Carolina, 6 to North Carolina, and 5 to Massachusetts. Furthermore, we can see that an exceptional number of districts in the 1990 redistricting violated contiguity, or were of questionable contiguity. So, at least in the aggregate, modern districts are more frequently discontiguous than was traditional. Are political and natural boundaries traditional borders for congressional districts? Yes. Very few districts divided town and county boundaries. Most were composed of whole counties and towns, or of whole counties subtracting only towns. Districts do begin to divide towns and counties following the third congress, but through the 38th congress the only deviations from this were for entire wards and other similarly sized units in urban areas. Between the 40th congress and the 62nd there were some splits even of these subunits, but only in a handful major cities (New York, Boston, Philadelphia, St. Louis, Baltimore, New Orleans, Chicago).[107] After Reynolds and Wesberry the number of districts that split even political sub-units of counties and cities triples and such splitting becomes widespread outside of major urban areas (Figure 3-7).

Districts Splitting Traditional Boundaries

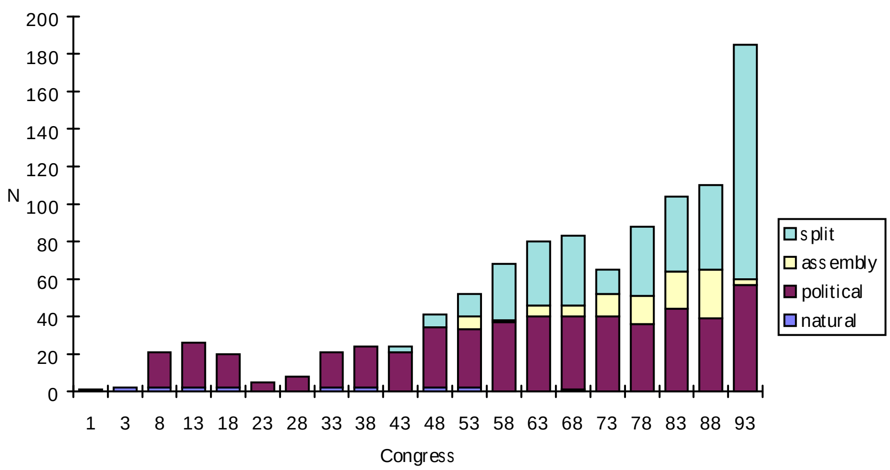


Figure 3-7. Violations of "traditional boundaries" in decadal redistrictings. Questionable districts violated county or town boundaries, but followed boundaries of political sub-units such as wards, election districts, election precincts, or assembly

---

[107] The choice of counties as a districting unit may well have resulted from the fact that the printed decennial census was aggregated to the county level for most of this period. I am indebted to Edward Still for making this point.

districts (the latter listed separately because it was legally required in some states). Split districts typically were split county boundaries in favor of streets or census tracts and blocks.

The extent to which historical political manipulation is responsible for the malapportionment, ill-compactness, and violation of traditional boundaries in this study is an interesting open question. Griffith's (1903) analysis of historical U.S. districts covers the period from colonial time up to 1842, and is the most comprehensive study of this type of which I am aware. In his study, Griffith identifies a number of congressional plans in the period under study as unequivocal attempts to gerrymander. Griffith relies (properly, in my opinion) more on political analysis rather formal indicia to identify gerrymanders. A reanalysis of these plans using my data suggest that formal measures are not consistent indicators of historical gerrymanders.[108] Hence, his conclusion that

---

[108] Some of these attempts were never passed, or were repealed before any elections were held, and hence are not included in my data sources. I analyze the remainder: New York's plan in 1789, Pennsylvania's and Massachusetts's plan in 1802, New York's in 1802-9, Massachusetts's in 1812-14, Virginia's in 1813, Massachusetts's plan in 1822 and in 1833, and Connecticut's in 1835 (Griffith, 1903, pgs. 42, 53, 55, 57-9, 72, 75, 77, 82, 89, 99, 105, 114). (Ohio's 1842 plan is also mentioned, but Griffith declares it out of the scope of his analysis, and does not give enough detail to positively identify gerrymandered districts.)

New York's 1789 plan was extremely malapportioned (in the worst 10% of plans) and violated contiguity, although it's compactness (by the most sensitive method, the minimum normalized perimeter score) was only somewhat lower than average and not in the bottom quartile. Pennsylvania's 1802 plan was even more malapportioned than New York's but was otherwise unexceptionable by formal measures. Massachusetts's 1802 plan was unexceptional by formal measures. New York's 1802-9 plans had "questionable" boundaries, following wards instead of counties (quite possibly, in part, to avoid malapportionment) and it had areas of questionable contiguity, but it was otherwise unexceptional. Massachusetts's 1812-1814 plan had areas of questionable contiguity and somewhat less than average compactness. Virginia's districts had somewhat worse than average malapportionment, and areas of questionable contiguity and extreme ill-compactness. Griffith, however, cautions against interpreting Virginia's odd district lines as indicative of a gerymander, noting that the worst looking district was not gerrymandered: "It (the district shape) indicates rather an indifference to the formation of districts in accordance with geographical considerations" (p. 83). Massachusetts's 1822 plan had a number of "questionable" splits of county and town boundaries, and although Griffith mentions a discontiguity in the plan, my data sources do not show it, perhaps because of the scale of the maps. It was otherwise unexceptional. Finally, Massachusetts's 1832 plan and Connecticut's 1835 plan were unexceptional.

regulating contiguity (etc.) is insufficient to prevent gerrymandering (pg. 118-119) seems well founded given the data and his identification of gerrymanders.

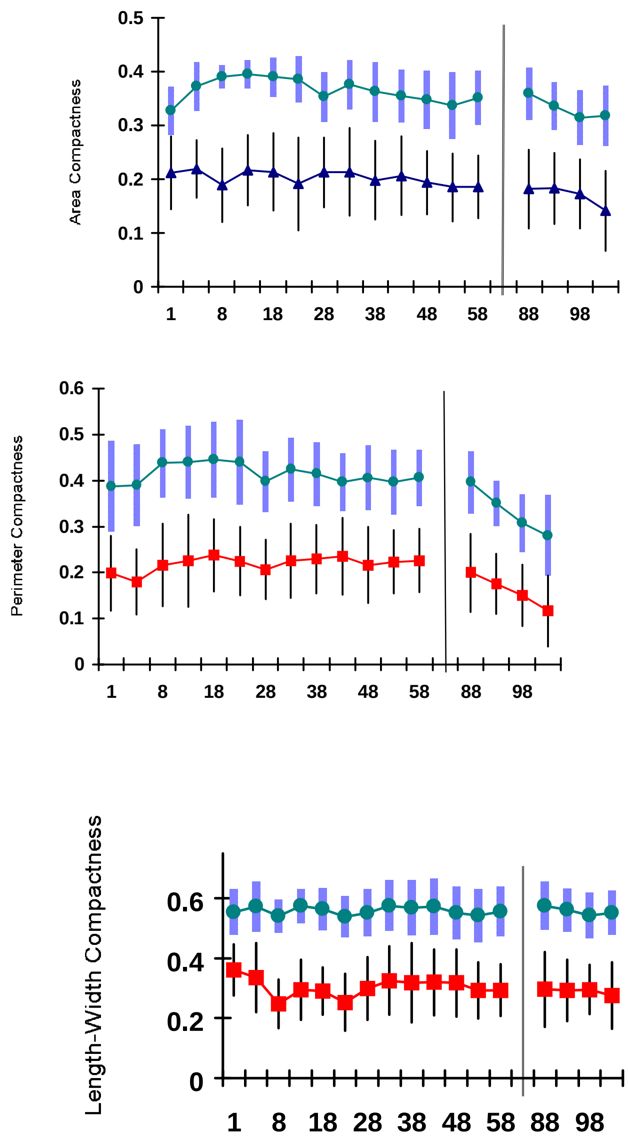
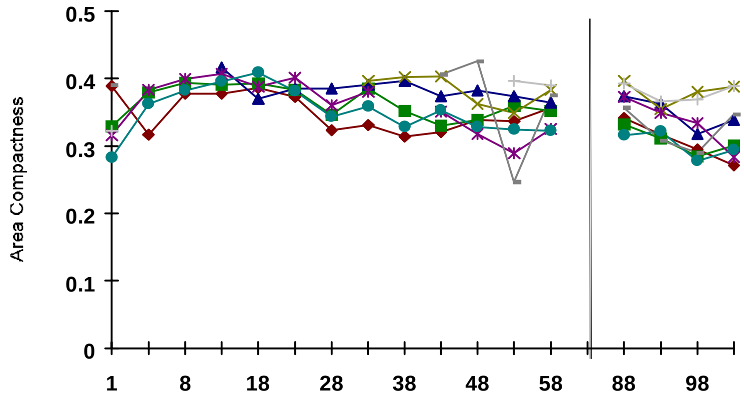
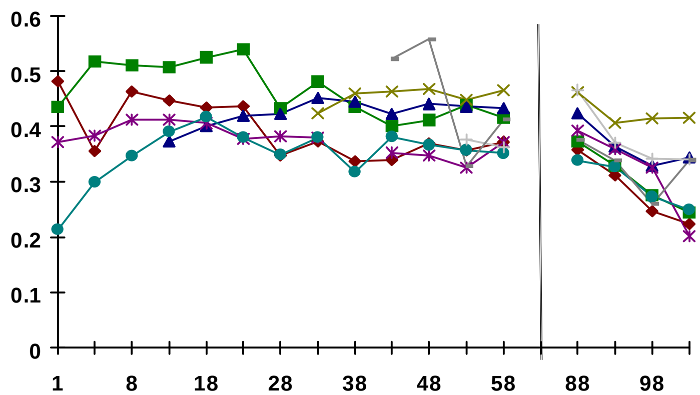


Figure 3-8a-c. Weighted average (by district) of compactness of state congressional districting plans with at least two districts for decadal redistrictings.

Area, normalized perimeter and length-width compactness are measured as described in Table 2. In each graph the top line shows plan compactness as equal to the mean of all districts, and the bottom line shows plan compactness as equal to the worst district. Vertical lines show plus/minus one standard deviation.

In modern times, compactness does seem to have fallen during the 1970's, 80's and 90's, especially when we look at perimeter-based compactness (Figure 3-8). Does this general stasis belie district trends? Figure 3-9a-c show regional variations in mean district compactness (Figure 3-9).





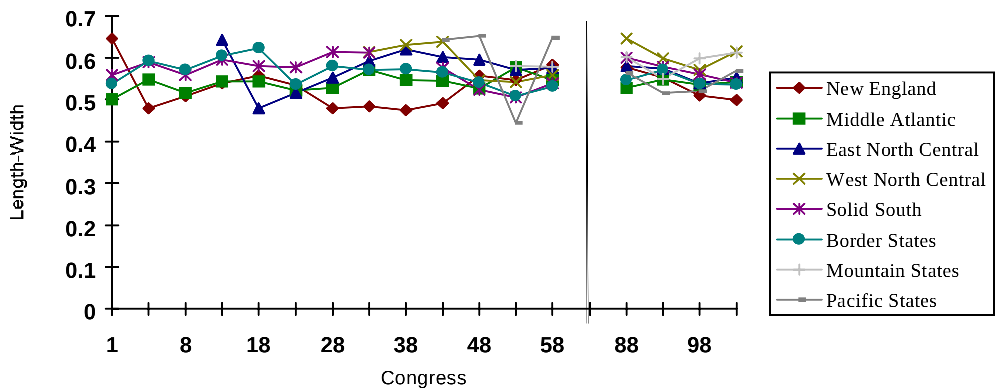


Figure 3-9. Weighted average (by district) of compactness of state districting plans with at least two districts for decadal redistrictings.

Again, besides a small general increase in compactness after the first congress, there seem to be no trends in regional compactness in the early period.[109] Compactness scores for different regions tend to be similar, depending on how compactness is measured, with area-compactness producing the most similar scores and perimeter compactness producing the most regional variance. As in Figure 3-7b, for most regions there is a decline in perimeter compactness over the last several decades. The redistricting plans challenged in *Miller*, *Shaw* and *Vera* were faulted by the Supreme court for failing to conform to traditional principles of compactness. Since *Shaw* in 1993, several other congressional plans have also been recently challenged (Congressional Quarterly Staff 1995; Idelson 1995), and all have been faulted, for, among other things, lack of compactness. Do the plans in the Supreme Court's line of compactness cases[110] violate traditional norms of compactness? Are they less compact than other modern plans? To answer this, we turn to Table 3-7.

---

[109] The pacific states do seem to diverge from the rest, but the data series is very short.

[110] As I previously noted, his line also includes *Hays*, but I exclude *Hays* from this part of the discussion because it was decided issues of standing, not compactness.

		Historical 1789-1913	1970's–90's	1990's Only
Shaw	AC	>1%	1.8%	2.6%
	NORM	0%	>1%	2.1%
	LW	19%	20%	18%
	Min AC	0%	>1%	>1%
	Min NORM	0%	>1%	1.3%
	Min LW	3.6%	3.7%	3.4%
Miller	AC	8.1%	15%	19%
	NORM	>1%	4.8%	11%
	LW	31%	33%	33%
	Min AC	1.8%	3.8%	5.9%
	Min NORM	>1%	4%	9.6%
	Min LW	26%	28%	25%
Vera	AC	48%	61%	64%
	NORM	0%	1%	1%
	LW	78%	78%	80%
	Min AC	8.8%	16%	20%
	Min NORM	0%	>1%	>1%
	Min LW	44%	46%	46%

Table 3-7: Perc**entile Ranking of Challenged Plans. Here I show what percentage of modern and historical districts were of equal or lesser compactness to those districts faulted for compactness by the Supreme Court.**

The districts in *Shaw*, *Miller* and *Vera* span the spectrum of compactness. Compared to historical districts, the districts rejected by the Supreme Court in *Shaw* were ill-compact by almost all measures. On the other end of the spectrum, the districts rejected by the courts as “bizarre” and “irregular” in *Vera* were not, on average, horribly ill-compact by two of the three compactness criteria. For example, by the length-width measure of compactness, the districts rejected by the court in *Vera* scored as well or better than 78% of historical districts. Even the worst of the Texas districts was as compact as 44% of historical districts, by the length-width measure.

It is the perimeter measurement that clearly distinguishes all of the rejected districts from earlier historical districts. The districts rejected by the Supreme Court in this series of cases were less compact, by this measurement, than almost any other district in the period 1789-1913. It is probably this type of ill-compactness that prompted Justice O'Connor to refer to the "(ir)regular(ity)" of districts fourteen times in the plurality opinion for *Vera*.

It is not necessarily fair, however, to compare the compactness of modern plans to historical plans because historical plans did not have to meet the approximate population equality standards imposed by *Wesberry* nor the absolute population equality historically imposed by *Karcher*. In fact, most of the rejected districts rank better relative to modern districts than to historical districts. This is especially true for rankings based on the perimeter measure, almost certainly because the court's decision in *Karcher* has forced states into making myriad minute adjustments to the boundaries of districts in order to exactly balance population. In Texas and North Carolina, for instance, no district varied more than 1 *person* from the average of more than 570,000 people. By contrast, in 1880, the largest and smallest districts in the two states varied by more than 36 and 89 *thousand* people out of an average of approximately 144,000 and 164,000, respectively.

The initial decline in compactness after the 60's, is probably in large part, of increasing court requirement for population equality. The decrease in compactness is not in isolation -- three changes occur immediately following the Supreme Court's decisions in *Reynolds* and *Wesberry*: traditional boundaries are violated in favor of census blocks, tracts and streets, malapportionment decreases and compactness decreases. Given the general stability before the Court's malapportionment decisions, the most straightforward inference from this pattern is that changes in compactness were a result of the splitting of local boundaries by redistricters to meet the Court's new requirements.

The Court continued to tighten its population equality requirements throughout the time period, culminating in *Karcher* v. *Daggett* (1983), which demanded, in essence, absolute equality.[111] Thus, some of the decrease in compactness over this period is almost certainly a result of the necessity to meet these requirements. (Which is not to say that some gerrymanderers did not make a virtue of necessity.)

3.5. Discussion

State and Congressional requirements for population equality, contiguity, and compactness never specified how these properties were to be measured. This failure does not raise problems in a historical analysis of malapportionment, since different measures still lead to the same conclusions. Contrary to what some scholars have argued, gross

---

[111] The Court seems to recognize the connection between maintaining traditional boundaries, increasing compactness and allowing population variance. The opinion in *Karcher* allowed for deviations of population in principle for such reasons as following traditional boundaries, although the Court did not allow such deviations in practice until very recently. In *Abrams* v. *Miller* (1997), the Court seems to have withdrawn, at least for the instant, from their zero-tolerance of population deviations in Congressional districts, in order to support Georgia's "traditional boundaries."

malapportionment was not a traditional feature of congressional districts. Malapportionment decreased and converged after the 28th congress, and reached its low point around the time of the 14th amendment. Although it was generally worst in the South, and best in the Northwest Ordinance states, in almost every region and by any measure, malapportionment at the time of *Wesberry* and *Reynolds* was worse than it had been historically, and far worse that it was when the 14th amendment was drafted and approved.

In his dissent from the *Wesberry,* Justice John Marshal Harlan claimed that the history of congressional regulation of congressional districts contradicts the court's finding that population equality is constitutionally mandated: "This history reveals that the Court is not simply undertaking to exercise a power which the constitution reserved to the Congress; it is also overruling congressional judgment" (page 548). For two reasons, I disagree.

First, the empirical analyses in this chapter show that the congressional regulations of 1842—1911 had at most a small and fleeting effect on congressional districts. Congressional districts were already, for the most part, increasingly well-apportioned and contiguous. This finding is consistent with the hypothesis that Congress was not mandating new and special requirements for districts, but stating commonly held norms that population variations should not be excessive. Second, the floor debates over these apportionment measures also bears out this interpretation. As I showed in Section 3.4.1, despite the long and spirited debate over each of these apportionment measures, the provisions for both equal population and contiguity were readily accepted. The fact that congress chose to regulate district criteria does not imply that such criteria did not also have a constitutional basis, as Pennsylvania Representative Mercur argued (quoted above).

Malapportionment was untraditionally high immediately prior to *Wesberry* and *Reynolds* exceeded that at the time of the 14th amendement, and in decades prior had far exceeded traditional levels. However, the Court's insistence on absolute population equality, especially with the *Karcher* decision, has resulted in a level of malapportionment that is untraditionally low.

In some ways, it is more difficult to measure violations of contiguity than to measure malapportionment. Still, while some marginal cases are difficult to classify in the absence of a precise definition, we can easily discern overall patterns of non-contiguity. Contiguity is not, strictly speaking, a traditional districting criterion. In practically every decadal apportionment from the first through fifty-eighth congress, some districts violated practical criteria of contiguity. Violations of contiguity, however, were few, and concentrated in a small number of states (New York, Massachusetts, and the Carolinas). The violations of contiguity in the latest round of redistricting far exceeded traditional baselines.

Modern districting plans exceed their predecessors to an even greater extent in the frequency with which their districts violate town, counties and other sub-unit boundaries. Except in dense urban areas, early districts followed county and town boundaries exclusively. Districts split larger boundaries with increasing frequency after the 43rd congress, but the frequency of violations of "traditional boundaries" skyrockets following the population requirements imposed by the Court in *Reynolds* and *Wesberry.*

Of the formal criteria examined in this chapter, compactness has been the most controversial. How to define compactness and whether it can be usefully defined are questions that have been debated since the first compactness requirement passed by Congress. (See Section 3.4.1.) Unlike evaluating population equality and contiguity, our conclusions about district compactness depend on which definition of "compactness" that we use. Whether or not there were traditional norms of compactness, and whether the current round of "ugly" districts violate these norms, depends crucially on the precise method that we choose to measure this property.[112]

---

[112] Reviewers of this chapter make the point that the common pre-Reynolds practice of states using county and town lines as units of representation for the state legislatures lead to these districts being historically more malapportioned and more compact than congressional districts.

While my studies confirm that Congressional districts were more equally apportioned than legislative districts (See Section 4.3.), I am unaware of any empirical study demonstrating the second point conclusively. Although Congressional districts were never required to be composed of whole counties, the vast majority of these districts did not, in fact, split such boundaries, so there is little reason to conclude *a priori* that one type of district would naturally be more compact than the other simply because legislative districts

Under the average area and the length-width measures, there seems to be little change in compactness over time, even in the last several decades (although by the "minimum" area measure, plan compactness has dropped slightly). The normalized perimeter-area measure tends to be the most sensitive to differences in plans and over

---

were composed of counties. Furthermore, while it is true that some legislative districts were based upon single counties or cities, counties (and cities) are not themselves necessarily compact, especially in older states and in earlier periods (See, for example, Maryland's, New Hampshire's and Tennessee's counties around the time of the first through fourth congresses), so legislative districts based upon single counties would not necessarily have been compact. Congressional districts containing several counties, could, in some circumstances be more compact than the individual counties comprising the district. (See, for example, Maryland's fourth district in the first congress, and Kentucky's second district in the third congress.) The compactness of legislative districts, and the related question of how county boundaries were created are interesting empirical issues worthy of future study.

More generally, I hesitate to draw conclusions about traditional districting principles for *congressional* districts from observations of *legislative* districts since state and congressional representation have had different historical and legal bases. Congressional districts are constitutionally based upon the representation of population, but legislative districts were sometimes based on different principles (as I show in Table 5).

time. Furthermore, this measures and seems to capture most consistently what the Supreme Court finds as “bizarre” in irregular districts.[113]

A number of districting plans in the 1990’s have been challenged, and some have been overturned, partially on the basis of ill-compactness. By many measures, these do not exceed traditional levels of “ugliness.” By the normalized-perimeter measure, however, these plans are unusually ill-compact, especially when we look at the worst districts in each plan. Under this measure, all plans have become less compact since the Court’s requirements of equal population in districts. These challenged plans are ill-compact even in comparison to other modern plans.

Districts have become uglier in modern times, but this fact has been slow to capture the attention of the Court, perhaps because this change has been, in part, a result of the

[113] Here, I am in general agreement with Pildes and Niemi (1993) key observations about changes in compactness from 1980-1990: Plans in the 1990’s were by some significant measures of compactness, worse than those in the 1980’s, this decline is worst when measured by extreme districts, and this decline is most significant when measured with perimeter-area standards. It is important, however, to place the change in compactness in a larger historical compactness. This chapter shows that decreases in compactness neither started, nor were largest, in the 90's, but instead followed the Court's equal population decisions in the 1960's.

Court's own actions. By completely eliminating malapportionment, rather than returning it to more traditional levels, it eliminated the traditional geographic districting as well.

3.6. Appendix: Corrections to and Omissions from the Data

No source of data is perfect, and this data is limited in four ways. First, population estimates for each district are based on decadal census data, and this data inevitably represents the population at the beginning of the decade more accurately than the population at the end of the decade.

Second, some districts contained political units for which no precise census data exists, either because the political unit was created after the census for that decade, or because the political unit subdivides one or more census aggregation units. In these cases, I adopted the district estimates found in the data source, or, if this were unavailable, I estimated the demographic data myself using census data aggregated at the county level.

This limitation particularly affects districts in major urban areas (primarily Baltimore, Philadelphia, New York City, Boston, St. Louis and Chicago) after 1860, because it was at about this time that many of these ceased to be created entirely from whole counties. For most of these districts it is possible to determine total population accurately by using census information aggregated at the ward level, but other demographic variables have to be estimated (Parsons, et al. 1986).

Third, partially because of this estimation problem, the available demographic data series extends only through the districts of the 62nd congress, and does not resume until the U.S. Census created the Congressional District Data Book (and Atlas) series for the 87th and later congresses. Of course it would be possible to reconstruct reasonably accurate apportionment data for most districts during this gap,[114] using the statutory descriptions of districts found in Martis and Rowles (1987) and ward-level data from the 13th–16th Censuses of the United States; unfortunately, this is a project well beyond the scope of this chapter.

Fourth, the level of detail in district maps varied across time and across sources. I attempted to use the most detailed maps available, but in a few cases lack of detail in the available district maps,[115] or differences in detail in different data sources has affected the accuracy of my measurements.[116]

---

[114] Estimates of more detailed demographic data for urban districts would necessarily be questionable at best, and perhaps useless.

[115] In all of the cases where the primary data source contained several maps of the same district, I used the map which captured the most detail, as long as the complete boundaries for that district could be reconstructed from it.

[116] In general, sources were in agreement on the overall shapes and areas of each district, but the perimeter of convoluted districts could vary significantly with the level of detail contained in each map. This problem is a result of the differences in information that are contained in different maps, and is not an artifact of the methods used to extract the data and is in large part unavoidable because perimeter measurements are, in general,

In this chapter I relied heavily on the *United States Congressional Districts and Data* series (Parsons, et al. 1978; Parsons, et al. 1990; Parsons, et al. 1986) for districts in the period of 1789–1913. In using this data, however, I discovered a number of omissions and errors, most of which I was able to correct using other sources.

This data series omits a number of district maps; the vast majority of these are of urban areas. For many of these maps used the district maps in Martis & Rowles (1982) for the following districts (Table 3-8):

---

sensitive to the accuracy of the measuring device; furthermore, in the case of natural, fractal, boundaries, the "real" length of a shape may be indeterminate.

For example, suppose you were trying to measure the length of a section of the California shoreline, perhaps the section between San Francisco and Los Angeles. If you used a coarse approximation, perhaps by measuring the length of Route 1, which runs along the shore nearby, you would guess that the shoreline is several hundred miles long. If you tried to make more precise measurements by walking along the beach, your path might expand to several thousands of miles. Finer measurements will reveal the shore to be of ever-increasing length. This problem is of most concern when comparing perimeter-based compactness measurements across districts that were measured at widely different scales.

Congress	State and Districts
3rd	PA-1
18th	IN-1–13
23th	PA-1 –PA-3
28th	NY-2–6, PA-1–4
30th	IA-1–2
33rd	MD-4, NY-3–8, PA-1–3
38th	MD-3, PA-14
43rd	MA-3–4, NY-11–14, PA-1–4
44th	PA-1–5
50th	IN-1–15,OH–50-51
58th	NY-18

Table 3-8. Redistricting plans that were reconstructed from other sources.

This data series also omits maps and population data for a number of minor redistrictings. I was able to reconstruct most of these redistrictings using county-level data for the following state plans: New Jersey's redistricting for the 29th congress, Ohio's redistricting for the 29th congress, Georgia's redistricting for the 31st congress, Indiana's redistricting for the 50th congress, and Kansas's redistricting for the 53rd congress. There were a number of omitted minor redistricting plans that I was unable to reconstruct, since they involved extensive changes at the ward level; population variables for the following districts was marked as missing in the data-set[117] (Table 3-9):

[117] In addition, there were several changes to New Hampshire's redistricting plan for the 32nd congress, in which a few towns were shifted among districts. I was unable to reconstruct these districts, but since these changes were very minor I chose to ignore them.

Congress	State and Districts
52th	MD-2–5
55th	MD-1
56th	MA-9–11,MD-3–5

Table 3-9, Redistricting plans that were partially reconstructed from other sources.

I corrected a number of obvious typos and inconsistencies in the population data and maps. Two notable errors were that the population of Tennessee's 5th district in its plan for the 53rd congress, and New York's 6th district in its plan for the 28th congress, were listed as ten times their actual size. Also, a typo shows Howard county in Indiana's 50th congressional map as belonging to two different districts. I corrected these errors.

For the purpose of determining whether a district followed traditional boundaries, Parsons's descriptions make clear whether a district is composed of whole towns and counties but do not always describe a split. Consequently, for all split districts, Martis's descriptions were checked. In a few districts in California's district descriptions for the 92nd—97th congress and New Jersey's district descriptions for the 93rd congress are so lengthy and intertwined that it is difficult to determine, even from Martis's detailed descriptions, whether a district split traditional boundaries. These indeterminate districts were conservatively classified as not splitting "traditional" boundaries. In addition, there are a few small inconsistencies between the two sources that could not be attributed to lack of detail in one of the accounts: Splits around the time of the 40th Congress in Louisiana and in the 43rd Pennsylvania are described somewhat differently by both sources. Martis gives more detail in this case, so his description was used.

Finally, the base maps for Maine and Rhode Island changed across time; maps of these states in the data series show them to have much longer coastlines before 1842 than afterwards, which causes identical districts to appear to have different perimeters and thus different degrees of compactness. To fix this inconsistency, I reconstructed all of the earlier districts for these two states upon a base map created from the post-1842 maps.

For population data after 1912 and before 1963, I relied on the *Congressional Directory*, using the directory first session after each decadal redistricting. There were occasional omissions or obvious typographical errors in these directories, but I was able to correct all of these by examining the directory for later sessions of Congress. In addition, I checked the total of the population in all districts against the recorded total for the state, and found minimal differences for most states and years. Only in five states in these five redistrictings were the differences bigger than 0.1% of the states population, and only in one case did the difference exceed 1%: California 63rd Congress (0.3%), Colorado 63rd Congress (1.1%), Pennsylvania 68th Congress (0.16%), Utah 83rd Congress (0.25%).

For districting plans in the period 1963–1993, I relied primarily on the district data books published by Congressional Quarterly (as described in Section 3.2). For the most part, these books provide complete and concise summaries of U.S. Census data and district maps. In a small number of cases, however, district maps in this data series were incomplete. I corrected most of these omissions by constructing maps of the district from maps contained in Martis & Rowles the U.S. Census Congressional District Atlas data series. Unfortunately, some of the maps in this series were fragmentary, and I was forced to omit a small number of districts from the analysis (Table 3-10).

Reconstructed Districts	
Congress	**State and Districts**
93rd	CA-23,25,32,34, IL-12, NY-6
98th	IL-1,2,4,7,9,10,12, NY-14,19
Omitted Districts	
Congress	**State and Districts**
93rd	MI-14,15,17, OH-22, PA-13
98th	CA–5,6,21,33
103rd	NY-13

Table 3-10, Redistricting plans that were reconstructed from other sources or omitted.

Chapter 4. Modeling the Effect of Mandatory District Compactness on Partisan Gerrymanders

4.1. The Increasing Importance of Compactness in Redistricting

From the time that geographical districting was first used in the United States, irregularly shaped districts have been subjects of popular attention, evoking criticism from the press, but little action from the courts. In recent years, however, the courts have begun to scrutinize district lines. To aid the courts in this scrutiny, the academic community has developed formal methods of measuring the geographic compactness of districts.

I use a three-stage model to examine the electoral effects of formal district criteria. First, I equate the task of drawing compact districts to an optimization problem, which I solve with combinatorial optimization techniques. By treating the problem in this way, I am able to draw thousands of compact district plans that are free from personal bias. Second, I generate many possible population maps, according to different clustering functions. With these maps I abstract away the eccentricities of any one local area, and focus on the electoral effects of general population characteristics on redistricting. Third, I examine the electoral outcomes that would be most likely under each plan, and I relate these outcomes to the geographical compactness of the district.

District planners now may want to go in the direction set by the courts and create compact plans. However, how will they tell whether their plans are compact? If well-known, effective, fair compactness standards existed, the district planner’s job would be simple. Unfortunately, more than thirty compactness measures have been proposed, and none of these measures has been rigorously examined. Scholars disagree about the consistency of these measures, their effectiveness in preventing electoral manipulation, and their neutrality.

Current empirical research does not help district planners to determine which compactness measures to use, which are effective, and which are neutral. It is to this research we now turn.

It is a common assertion that compactness standards will constrain partisan gerrymanderers. This is a straightforward prediction, but misses the point. It is a simple mathematical truth that the maximum of a constrained optimization problem is less than or equal to the maximum of the same unconstrained problem. Hence if one thinks of redistricting as such a mathematical problem, in which one group has control of the redistricting process and acts single-mindedly to maximize a single goal, then any sufficiently restrictive constraint on redistricting plans will reduce the ability of that group to obtain its goal. This argument applies equally well whether the group is a team of Republican redistricting experts or a minority Political Action Committee, and whether the goal is to maximize the probability of partisan control of the legislature, or to maximize the number of minority opportunity districts.

There are many ways of constraining gerrymandering, including eliminating redistricting altogether; the important question is, what are the consequences of constraining redistricting in this way: Can compactness be measured consistently and sensibly? If so, which compactness measures should we use? How restrictive do compactness standards have to be in order to have an effect? To what extent will compactness standards limit redistricting for “good government” goals? Are compactness standards neutral — or will they systematically benefit certain political groups?

Several authors have examined the question of consistency with formal methods (Young 1988), and I will not repeat that work here. I will focus, instead, on the questions of effectiveness and neutrality, which have been debated so recently and hotly. I will pay special attention to the interaction of compactness rules and population characteristics.

4.2. Modeling Partisan Gerrymanders Under Compactness Rules

This chapter uses computer simulation to model the effects of geographical compactness on politics. Formal and empirical analysis in this area is limited. Formal models of redistricting are extremely simplified. Although formal models of redistricting exist, most of these ignore geography altogether, and those that include it do not model compactness or natural population distributions.

Empirical analysis of districting criteria cannot avoid selection bias. Real districts are not random samples but intentional creations, and there are many potential causes for “ugly” districts. Ill-compact districts may be caused by geographical constraints; by an underlying unevenness in the distribution of population across a state; by attempts to follow “natural” political boundaries; or by political attempts to manipulate lines for the benefit of communities of interest, racial minorities, political parties, or incumbents. These causes are difficult to measure and they may interact in complex and confounding ways — district lines drawn to protect incumbents in one district may be compact, in and of themselves, but may cause a neighboring district, drawn in absence of any political motive, to be ill-shaped. For these reasons, an unguided analysis of ex-post non-compact districts says little about the effects of compactness as a restraint on the political process ex-ante.

We can use simulation to draw districts that are driven only by compactness, and we can create a nearly unlimited numbers of these. Simulation allows us to abstract away from the geographical and political eccentricities of any single plan, and to directly analyze the relationships among the shapes of districts, the distribution of political groups, and the outcomes of elections – relationships that can be used as hypotheses to guide empirical analysis. As two of the strongest advocates of compactness write: “Enough knotty statistical issues must be overcome that probably the only way to settle this point (the effect of compactness standards) is through … running thousands of computer models of compact districts and seeing what happens” (Polsby and Popper 1991, 335 fn.).

Historically, computers have been frequently used to create redistricting plans — but usually as a tool to assist human planners (Browdy 1990a).[118] In Chapter 2, I showed

---

[118] There are some notable exceptions to this usage: Shepard and Jenkins (1970) and Taylor (1973) also use automation procedures to examine a range of districting options, but apply their techniques to the analysis of a particular election, rather than to an election rule, while Engstrom and Wilder (1977), Taylor and Johnston (1979), and O’Loughlin

how individual districts could be mathematically characterized. In this chapter, I treat redistricting, as a whole, as a mathematical set-partitioning problem, and use automated districting techniques to generate a series of arbitrarily drawn district plans. I complement these automated districting techniques with general combinatorial optimization algorithms that have been used successfully on similar problems in computer science. These same algorithms can be applied to a variety of formal districting rules. Because compactness standards are central to the current debate over redistricting, I use this procedure to focus upon plans drawn under equal population and compactness rules.[119]

(1982) argue that districts created automatically should be used as a benchmark with which to detect gerrymandering.

[119] You should note that, in these simulations, I model explicitly only compactness and equal population rules. In particular, I do not require plans to be contiguous. I do this for several reasons:

• First, in order to isolate the effects of compactness, the constraint set and value functions were kept as simple as possible.

• Second, contiguity is often, in practice, an ill-defined or vacuous requirement. Practically any set of regions can be made contiguous if lines are drawn finely enough.

4.2.1. Mathematically Characterizing Redistricting

If you were a mathematically-inclined district planner, you might characterize redistricting as a partitioning problem. (See, for example, Gudgin and Taylor (1979)) You would simplify the problem a bit by pretending that the state that you wish to redistrict is composed of indivisible[120] census blocs. Then, you would write out a function to evaluate partitions of blocs.[121] Finally, you would solve for the maxima of the problem — you would find the partition with the highest value. If you could perform this procedure then you would have the best district plan possible. Adding compactness standards does not make this problem more difficult to formulate. You can bring

---

• Third, compactness requirements encompass contiguity: the maximally compact plan will not be measurably and avoidably noncontiguous.

[120] This is not far from the truth since most population data is, at best, limited to the census-bloc level of detail.

[121] A partition divides a set into component groups that are exhaustive and exclusive. More formally:

For any set $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$, a *partition* is defined as
a set of sets $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_k\}$ $s.t.$
(1) $\forall x_i \in \mathbf{x},\ \exists \mathbf{y}_j \in \mathbf{Y}, s.t.\ x_i \in \mathbf{y}_j$
(2) $\forall i, \forall j \neq i, \mathbf{y}_j \cap y_i = \varnothing$

compactness standards into the problem either as constraints to optimization, or as supplements to your value function.

This characterization of redistricting is useful whether you are trying to find the least biased plan or the most effective gerrymander: If you are an altruistic social planner, you would use a value function that weighed all of the social benefits and costs of redistricting. An altruist might include such factors as preserving county boundaries, maintaining the competitiveness of districts, and minimizing the “bias” of the plan in your value function (Lijphart 1989). A partisan, on the other hand, might attempt to maximize the number of safe party seats, or the probability of their party being in control of the legislature. Alternatively, a self-interested incumbent might attempt simply to maximize the probability of retaining her seat in upcoming elections.

Because altruistic social planners are likely to be outnumbered by partisans, incumbents, and other self-interested individuals, we may wish to impose rules on the redistricting process. Whether we require that all districts have the same population, that they respect political boundaries, maintain geographic contiguity, or that they comply with compactness criteria, we can represent these requirements as constraints on our optimization problem. Approaching redistricting mathematically has two advantages: This approach can help us to draw better districts, and it can help us to predict the effects of particular redistricting rules.

4.2.2. Analyzing The Effects Of Redistricting Rules

How can we use a mathematical characterization of redistricting to predict the effects of redistricting rules? Suppose you are a party leader, intent on producing a partisan gerrymander, and suppose that a citizens' group introduces an initiative requiring all districts to be compact. Should you expend political resources to fight this initiative?

To make this decision, you will have to estimate the effectiveness of the partisan gerrymander you expect to obtain when there are no rules, and then you must weigh that estimate against the effectiveness of the gerrymander which you expect to obtain if you are forced to draw contiguous districts. In mathematical terms, you subtract the value of the optimal partition of the constrained problem from that of the optimal partition of the unconstrained problem. If the difference is big enough, then you should fight the contiguity rule.

In this example, we assumed that the best plan that is found would then be *chosen*. This corresponds to the situation in which an organized group has substantial control over the districting process. In essence, this is a "game" played between a party and the courts. The controlling party submits a redistricting plan, which the court may accept, modify or reject.[122]

4.2.3. Creating Arbitrary Redistricting Plans

While we can easily *formulate* redistricting as a partitioning problem, this problem may be difficult to solve. Political scientists have used a number of different methods to

---

[122] Some caution would be warranted in extending this model of partisan gerrymanders to incumbent gerrymanders, in which each incumbent vies for a district that maximizes her chance of reelection. Unlike a partisan gerrymander, created by party leadership, the incumbent gerrymander may not be the result of an individual choice *per se*, but the result of a strategic game with multiple players. In practice, even if the optimal incumbent gerrymander were known, it might not be the plan that emerges from the smoke-filled rooms in the back of the state house.

Despite the possibility of game-theoretic complications in incumbent gerrymanders, it is valuable to understand how district rules affect the optimal plan. In some cases, a plan will, in effect, be chosen by party leadership or by some other unified group. Even if the redistricting process is best modeled as a game in this case, we still must understand the payoffs to players under different rules (i.e., the expected value of the optimal partition), before we can analyze equilibria of the game.

search for compact districts. At the same time, computer scientists have developed similar techniques to search for optimal partitions.

Exact methods systematically examine *all* legal districts either explicitly or implicitly. Explicit enumeration ("brute force" search) methods literally evaluate every possible plan (examples include (Gudgin and Taylor 1979; Shepherd and Jenkins 1970)),[123] but more sophisticated methods, such as "branch and bound," exclude groups of solutions that are obviously sub-optimal.

Unfortunately, as the number of population blocs increases, the number of potential plans grows so rapidly that no computer can evaluate all of them explicitly. Furthermore, redistricting problems belong to a set of problems for which it is widely believed that no guaranteed, feasible, optimization methods exist.[124]

---

[123] A close examination of these algorithms reveals that in order to make the programs finish in a reasonable amount of time the authors use unproven "short-cuts," making them, in actuality, heuristic. (See Chapter 5.)

[124] Technically, these problems can be proved to be in the set of problems that computer scientists have labeled "NP-Complete." (See Chapter 5.) Most computer scientists believe that to solve an NP-complete problem exactly requires you to spend computation time that grows exponentially with the problem's size ("n" above), although

In the larger simulations, because of these barriers to finding an exact solution to larger problems, I turn to *heuristic* methods for finding compact districts. These methods by definition, do not guaranty an optimal solution, but often perform well in optimization. Most heuristics for locating optimal partitions are based on the principle of iterative improvement. In the simplest of these methods, known as *hill climbing*, you start with a set of randomly generated redistricting plans and repeatedly look for small changes to the plan that improve it — stopping when there are no small alterations that can yield an improvement. Several previous researchers have used variants of hill climbing methods to draw new district plans or to make improvements to existing districts (Liittschwager 1973; Moshman and Kokiko 1973; Nagel 1972; Rose Institute of State and Local Government 1980; Vickrey 1961; Weaver and Hess 1963). I use a variant of hill climbing similar to Nagel's method. In addition, where it is possible to derive the most compact plan from geometric arguments. I use a simple "descent" algorithm: I start with the most compact plan and use hill climbing to create less compact variants of it. (See the appendix to this chapter.) In the illustration below, I show an example of this process (Figure 4-1).

---

there may sometimes be ways to more quickly find approximate solutions. See Papadimitriou 1994 (1994) for a general introduction to NP-completeness.

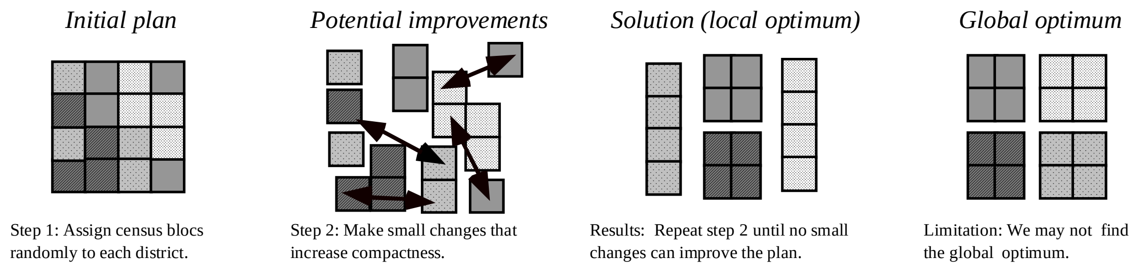


Figure 4-1. An example of creating a compact plan through "hill climbing."

Simple hill climbing methods, however, sometimes produce results that are far from optimum, because these methods are trapped easily in local optima. To minimize this problem in the simulations, I test hill-climbing methods against a number of methods that have been successfully used to solve similar problems in other fields: simulated annealing, genetic algorithms, and Monte Carlo methods. This variety of different methods helps to ensure that results are not being driven by quirks in the optimization process.

Simulated annealing is one of the most successful of combinatorial optimization methods. It is based on a mathematical analogy to the slow cooling of metal. If the value function being optimized is sufficiently well behaved, simulated annealing asymptotically converges to the optimum value (van Laarhoven and Aarts 1989). It has been previously recommended for use in redistricting (Browdy 1990b). *Genetic algorithms* are search algorithms based on an analogy to natural selection and genetic combination. Potential solutions to the optimization problem are defined as genetic strings, which can be mutated or "crossed" with other strings. A group of potential solutions then competes to survive and reproduce in the next generation.

Chandrasekharam (1993) demonstrates the effectiveness of genetic algorithms for graph-partitioning problems, which are somewhat similar to the redistricting problem. For details on the algorithms used in this chapter, see the appendix.

I cover the range of procedures that are available for drawing compact districts. Should compactness standards be legally mandated, district planners will have little choice but to turn to such techniques. Thus the plans that I produce, though simpler than real district plans, are similar in principle to those that will be produced should compactness standards become widespread.

4.2.4. Measuring Compactness

All of the optimization methods that I have discussed are flexible enough to accommodate a variety of value functions. Yet, choosing a particular compactness measure was a special challenge because previous researchers have proposed over thirty distinct measures of compactness (Niemi, et al. 1991; Young 1988). Neither the courts nor political scientists recognize a single standard for measuring compactness. In addition, although many states require compactness, only three states (Iowa, Colorado, and Michigan) define the term (Grofman 1985). What compactness standards represent the set best?

Most measures in the compactness literature fall into one of three categories: "area-based" measures, "perimeter-based" measures, and "population-based" measures. For the simulations, I chose a compactness standard from each of these categories. In addition, each of the measures that I use duplicates, as closely as is practical, legal standards of compactness the United States:

To measure the compactness of a district's area, I compared the area of the district to the area of the smallest box bounding that district.[125] A plan's compactness is defined as the mean compactness of its districts. This measure duplicates, as much as possible, a compactness requirement used in Iowa and in Michigan.[126]

---

[125] Some similar compactness measures use a bounding circle or convex polygon. While these other shapes are theoretically preferable, I use a bounding box here because of the discreteness of the simulation, map, and the desire to have my standard duplicate current legal standards. Furthermore, given the large granularity of population blocs in this simulation this measurement is quite similar to comparing districts to the bounding circle, while being more efficient to compute.

[126] These measures existed in state statutes or constitutions at the time of writing, but are usually not enforced. The 1980 Iowa General assembly Bill generally defines compactness as, "Compact districts are those which are square, rectangular or hexagonal in shape to the extent permitted by natural or political boundaries." The Michigan constitution also specifies, generally, that its state house districts should be "as nearly square in shape as possible." Iowa also offers several operational definitions of compactness, the first of which is "the absolute value of the difference between length and width" (Grofman 1985, 180 fn.). While this absolute value is not equivalent to my

To measure the compactness of a plan's boundaries, I calculated the total perimeter of all its districts; the best plan minimizes this total. Colorado currently uses this measure to evaluate districts (Grofman 1985).

To measure the compactness of a district's population,[127] I calculated the moment-of-inertia for the district's population.[128] A plan's compactness is defined as the mean compactness of its districts. This approach is similar to the measure in force in Iowa, which has the only population measure currently in effect in the U.S.[129]

---

bounding-box measure, it is a similar, if cruder, attempt to capture the squareness of a district.

[127] Since in the simulations, all population blocs have the same weight (although the partisan proportion in each bloc may vary), population-based measures produces results identical to analogous area-based measures.

[128] Formally, this is $\frac{\mathrm{P}}{\sqrt{\sum_{x \in X} \left(p_x d(x)\right)^2}}$, where P is the total population of the district, p_x is the population of a particular census bloc, X is the set of all census blocs in the district, and *d()* is the geographical distance from the center of the census bloc to the population-center of the district.

[129] The Iowa measure calls for taking the ratio of the "dispersion of population" around the population center of the district, to the dispersion around the geographic center

Where possible I normalized their scores to fall within the (0,1] interval. Plans that have a score of “1” are as compact as possible. (This normalization was not possible for the perimeter measure, since it is not always possible to know the value of the perimeter of the most compact plan.)

4.2.5. Simulated Politics

In the simulation, *m*x*n* grids represent maps. Two different political groups populate each map. To separate the effect of compactness from the effects of equal population standards, each census bloc is normalized to have one hundred voters, so that only the proportion of each type of voter varies across population blocs. Each group runs a candidate in each district, and members of that group will vote “sincerely” for a candidate identified with that group. These assumptions best fit polarized, bipartisan elections.

Although I assume political groups have symmetric voting behavior, I allow them to be distributed across the state “map” differently. I duplicated the simulations using three different models to determine the political composition of each census bloc.

---

of the district. Since in this simulation I use only maps with uniform population densities, these two centers are always identical, and hence this ratio will always be one. Under these conditions, the population moment-of-inertia measure that I use better captures population dispersion.

I first performed a set of simulations using a very simple population model that I will refer to as the “uniformly-random” distribution. Unless otherwise noted a total of 10,000 plan/population distribution combinations were examined for each simulation run. In this model, the population of each census bloc is drawn from the same normal distribution.[130]

Clustering is a feature of realistic population geography models (Garner 1969). And in the second set of simulation runs, I used a “clustered” distribution to model the distribution of political groups. In this “clustered” distribution one political group is concentrated into *r* compact clusters, each of size *s*, and each randomly located. Similar cluster models have been used previously to explain voting behavior; in particular, Gudgin & Taylor (1979) find that the well known “cube law” of elections can be explained by a variation of the cluster population model.

In the third set of simulation runs, I use a more complicated clustering process that is based on Schelling’s (1978) neighborhood formation model. In Schelling’s model, persons in two different groups are at first randomly distributed on a map, then if an individual is surrounded by too many individuals of the other type, they can move to any adjacent square, if they prefer. In each round, every person is offered an opportunity to

---

[130] I duplicated each of the simulations in this model, substituting uniform distributions (with the same mean) for normal distributions, but the simulation results were indistinguishable.

move; when no one can improve their location by moving, the neighborhood is 'stable' (Schelling 1978). My model is similar to Schelling's, although it is aggregated at the census bloc level.[131] I show a typical population distribution that this model generated (Figure 4-2):

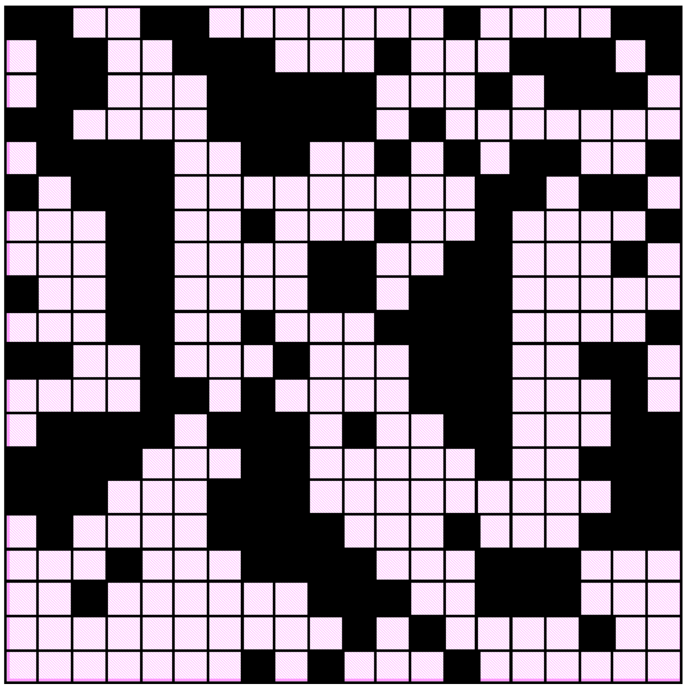


Figure 4-2. A map of 20x20 census blocs, with two political groups distributed across it using a Schelling distribution. The black squares represent census blocs primarily occupied by the minority political group.

[131] Schelling requires that individuals move into empty spaces, whereas I allow two willing individuals to trade places.

4.3. Evaluating The Electoral Effects Of Compactness: Simulation Results

The results from these thousands of simulations reveal three interesting properties of compact plans and of compactness standards. First, the distribution of compact plans shows that compactness measures are useful only for comparing similar plans, but not for making absolute measurements of plans. Second, the simulation shows the difficulty of drawing compact plans under some measures — we should avoid these particular measures if we want to minimize the potential for gerrymandering. Finally, the simulation shows that district compactness can systematically influence election results.

4.3.1. Compactness Measures Are Relative Measures

I used an exhaustive analysis of districts to generate all possible plans for a number of small maps. I use a box-plot to compare the distribution of compact plans under the perimeter measure, for different map sizes, shapes, and numbers of districts (Figure 4-3):

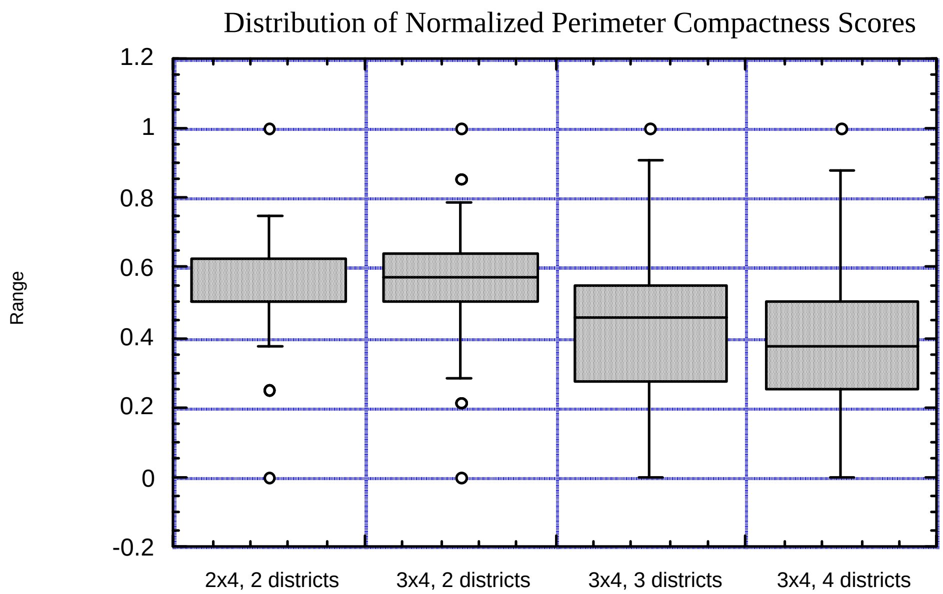


Figure 4-3. Box plots of Plan Compactness. The perimeter measure is normalized to the (0,1) interval.

Looking at any one of these box-plots, you can easily see that compact plans are scarce — most plans fall *far* short of the optimum. Furthermore, the scores of most plans cluster in a very narrow range of compactness values.

If you compare the box-plots for different maps, you can see that while the distribution of scores for each map is similar, the values of the minimum, maximum, and median scores are quite different. Given only the compactness scores of two plans, you can make reasonable comparisons between them only if these plans partition the same map into identical numbers of districts — a compactness score is meaningless outside of its specific context. For example, a score of “.5” is in the bottom decile of plans for the first map in Figure 4-3, and in the *top* decile for the last map.

Some authors have proposed that the courts mandate a minimal level of compactness for district plans. Given the shape of the distribution of compactness scores observed here, the effectiveness of compactness standards for limiting manipulation is likely to be very sensitive to the particular minimum level specified. If the minimum level is set high, the vast majority of plans will fail to meet the standard — it may be difficult to draw any plans at all. If, on the other hand, it is set at the middle of the distribution, the ability to gerrymander may be virtually unaffected.

Empirical studies of compactness scores must also take note of both their nonlinearity and their sensitivity to geographic context. Suppose that your ability to gerrymander is roughly proportional to the number of plans from which you can choose: Then, you will find it immensely more difficult to create an effective gerrymander that scores in the top 99th-percentile than to draw a plan with a slightly lower relative score. Furthermore, since compactness scores will depend on state boundaries, you may find it easy to create a gerrymander that scores “.90” in a state with regular boundaries, like Iowa, and impossible to create any plan at all that scores above “.75” in a state like Maryland. Differences in population distribution can be expected to further cloud such comparisons, as the ability to draw compact districting plans will be affected by equal population constraints. In general, comparing the compactness of plans across different states has little value.

4.3.2. Some Compactness Standards Make Detection Of Gerrymanders Difficult

In Table 4-1, I compare the performance of each optimization method. (This measure is limited to cases where the optimal plan can be deduced from regularities in the shape and population distribution.) Both the hill-climbing method and the genetic algorithm were equally successful in finding optimal plans, although the genetic algorithm was too computation-intensive to use on the larger maps.[132] Unlike these two methods, both the Monte Carlo procedure and simulated annealing performed poorly.[133]

---

[132] The time required to find a solution using the hill climbing method seemed to grow quadraticly in the number of census blocs (in time-complexity notation) while the convergence rates for genetic algorithms and simulated annealing grew at an even faster rate. Consequently, for maps larger than 5x5, I used descent and hill-climbing methods exclusively.

[133] The annealing procedure that I used sometimes made trades that would cause the population of the districts to become unbalanced, lowering the overall score of the plans. Once it did this, it was usually unable to recover because future changes to the plan were unlikely to bring the plan back into balance. Although I felt constrained to use well-documented and generalized algorithms, in practice, it should be possible to modify the algorithm for better performance with given redistricting goals.

Grid Size	# of Districts	Number of Possible Plans	Measure	Best Possible Score	Mean of 1000 Random District (std dev)	Hill-Climb mean (std dev)	Anneal Mean[134] (std dev)	Genetic Mean (std dev)
3x4	4	15400	perimeter	32	42 (3.0)	32.1 (0.41)	---------	32 (0)
3x4	4	15400	area	0.75	0.33 (0.07)	0.73 (0.07)	0.45 (0.096)	0.73 (0.03)
3x4	4	15400	moment [135]	1.5	0.98 (.12)	1.0 (0.02)	1.2 (0.1)	1.49 (0.06)
5x5	5	5.2×10^{12}	perimeter	unknown	86 (4.6)	54.8 (2.8)	---------	55.5 (4.3)
5x5	5	5.2×10^{12}	area	unknown	0.23 (.03)	0.42 (0.065)	0.02 (0.07)	0.42 (0.046)
5x5	5	5.2×10^{12}	moment	unknown	0.61 (0.04)	0.99 (0.001)	---------	1.03 (0.05)[136]
8x8	4	5.0×10^{53}	perimeter	64	203 (9.3)	90.5 (8.7)	---------	----------
8x8	4	5.0×10^{53}	area	1.0	0.25 (.006)	0.27 (0.015)	---------	-----------
8x8	4	5.0×10^{53}	moment	1.0	0.34 (.007)	0.65 (0.02)	---------	-----------
9x9	9	1.5×10^{65}	perimeter	108	295 (7.3)	150 (9.8)	---------	-----------
9x9	9	1.5×10^{65}	area	1.0	0.12 (.006)	0.18 (0.026)	---------	-----------
9x9	9	1.5×10^{65}	moment	1.0	0.32 (.009)	0.87 (0.03)	---------	-----------
20x20	8	3.0×10^{728}	perimeter	160	1508 (13)	-------------	---------	-----------

[134] Annealing returned plans that violated equal population constraints, these plans were assigned a compactness value of 0.

[135] If measured over a continuous area, the moment of inertia measure for a shape can be no larger than one, but this condition is violated in very small discrete approximations.

[136] Sample size in this case was 148, because computation exceeded time limit.

20x20	8	3.0×10^{728}	area	1.0	0.06 (.001)	-------------	---------	-----------
20x20	8	3.0×10^{728}	moment	1.0	0.13 (0.0)	0.36 (0.003)[137]	---------	-----------

Table 4-1. Performance of algorithms using different measures of compactness.[138]

In fact, the most striking differences in this chart are not among methods, but among compactness measures. I show the ratio ("approximation ration") between the mean value of the plans created using the best optimization method to the value of the most compact plans possible. Notice that, in general, these methods were much more successful at finding compact plans under the perimeter standard and moment-of-inertia standard than under the area-based standard. What does this tell us about the properties of these different standards? (Figure 4-4)

[137] Sample size in this case was 287, because computation exceeded time limit.

[138] For each table entry I performed 1000 simulation samples, unless otherwise noted. When a cell filled is filled with dashes it means that the specified algorithm was not able to complete a significant number of iterations before the time limit (several days) expired.

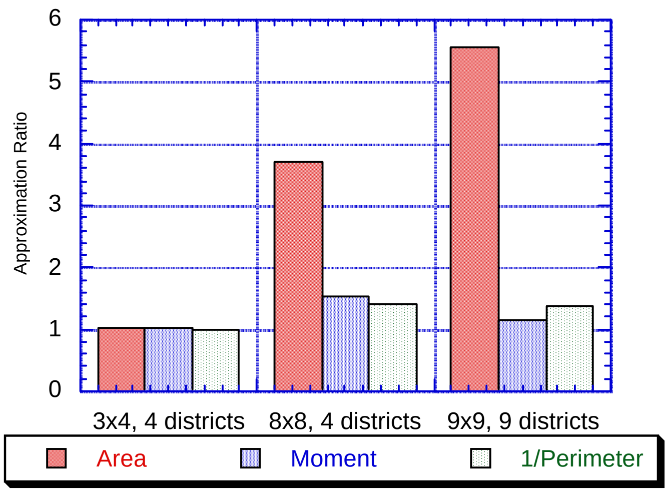


Figure 4-4. To obtain an approximation ratio, I divide the best possible score by the mean reached by the best algorithm.[139] The best possible ratio is one, which means that the algorithm always reaches the best possible solution.

Remember that all of these optimization methods rely upon iterative improvement. In other words, they operate through change that is gradual and limited. Since, as we have seen, these methods work well for the perimeter-based standard, we can conclude that the perimeter and moment standards are sensitive to small changes in a district plan; on the other hand, the area-based measure is much less sensitive to small changes— to improve the plan's area compactness we need to change districts radically.

---

[139] I use the inverse of the perimeter here because the perimeter measure grows when a shape becomes less compact, unlike the others.

Changing districts radically can be politically difficult and can interfere with other redistricting goals, such as preserving natural boundaries and communities of interest. While the simulations ignore these concerns, the courts should not. Because of these difficulties, the courts will find the perimeter-based standard easier to manage than the area-based standard. (In coming to this conclusion, I retain the assumption from Section 4.2.2 that a single party substantially controls the redistricting process, and is able to create plans that the court must then either approve, modify, or reject.)

Furthermore, for most real district maps we will not know the value of the most compact plan beforehand, a situation that is exacerbated by the area-based measure. Since it is likely to mislead us with plans that are locally optimal, but which fall far short of the most compact plans, the area-based measure allows gerrymanderers much greater leeway in designing their districts. Altruistic district planners will suffer as well, as they may expend unnecessary effort trying to improve a plan that is already very close to optimal.

4.3.3. Compactness Standards May Create Opportunities For Political Manipulation.

In addition to failing to prevent gerrymanders, there is a further consideration that has not been suggested in the literature as yet: The process of evaluating plans under a compactness standard might well induce strategic behavior that would harm the reapportionment process.

Finding the maximally compact plan is, as I have indicated, a very difficult mathematical problem. In practice, it will often not be possible to determine whether a plan is "optimally compact," especially for plans composed of a large numbers of census blocs. Moreover, as I argued in the Section 4.3.2, the simulation results indicate that if we do not know the value of the optimal plan, we cannot set reasonable an "absolute" compactness limit.

Instead of using some absolute measuring value, we will have to compare plans against each other, or simply search for "improvements" to any proposed plans. It has even been suggested that when two plans are proposed, the most compact should automatically be implemented (Polsby and Popper 1991).

Unfortunately, in a strategic political environment, in which plans are compared only with each other, the very shape of districts becomes valuable information to your opponents: If you hide your plan, there is a chance that opponents will mistakenly believe their plan to be the most compact, which is to your advantage. Whereas if you reveal your plan, you give up this strategic advantage without gaining anything. In sum, because compactness standards give district planners an incentive to hide information, these standards may increase political manipulation.

4.3.4. Compactness Standards Are Not Politically Neutral

In this next section, I will show that there is a systematic relationship between compactness standards, population distribution and electoral advantage. The specific effect that these standards will have on redistricting, however, will depend on the political institutions used to create districts. Here I examine the effects of compactness on partisan gerrymanders and on automated redistricting plans.

Arbitrarily Selected Compact Districts

Polsby and Popper 1991 claim that if the court adopts a policy of automatically accepting the most compact districting plan proposed to them, then through competition among political groups, gerrymandering will disappear. This view of arbitrary compact district plans is relatively recent, but scholars have long argued that we should simply use a computer to generate arbitrary district plans, following only the principles of compactness, contiguity and population equality (Harris 1964; Kaiser 1966; Weaver and Hess 1963). Suppose that we did manage to create districts arbitrarily, following only the principles of compactness and population equality, as these scholars desire. Would this be a neutral solution to the gerrymandering problem? Can we reach color-blindness if we choose the first horn of Justice Souter's dilemma by awarding primacy to compactness standards?

In Table 4-2 I show the correlation between compactness and electoral results in such a case. Since neither compactness scores nor seats were distributed normally, I also report Somers's d along with the correlation measures. Somers's d is similar to Kendall's Tau, except that it treats ties asymmetrically, ignoring ties on the dependent variable (number of minority seats, in this case). I use it in this case because of the number of minority districts takes on only a few values, leading to many ties that would distort Kendall's measure. See (Liebetrau 1983) for a discussion of these measures.

Grid Size	Number of Districts	Cluster Size	Number of Clusters	Minority Percentage of Population	Mean Minority Controlled Districts (std dev)	Correlation Between Minority Controlled Districts and Compactness (Somers's d)
5x5	5	1	3	12%	0.02 (0.13)	0.0 (-0.03)
5x5	5	1	5	25%	0.34 (0.49)	0.0 (-0.01)
5x5	5	1	12	48%	2.37 (0.6)	0.06 (0.00)
5x5	5	4	1	16%	0.35 (0.47)	0.39 (0.44)
5x5	5	4	2	32%	1.17 (0.63)	0.36 (0.35)
5x5	5	4	3	48%	2.37 (0.6)	0.07 (0.08)
5x5	5	9	1	36%	1.46 (0.6)	0.32 (0.26)
8x8	4	1	26	40%	0.47 (0.43)	0.01 (-0.02)
8x8	4	9	3	42%	0.93(0.65)	0.46 (0.47)
8x8	4	16	1	25%	0.24(0.42)	0.46 (0.61)
8x10	8	9	3	34%	1.38(0.89)	0.58 (0.52)
20x20	8	1	100	25%	0.04(0.20)	0.0 (0.0)
20x20	8	4	40	40%	3.1 (1.2)	0.29 (0.23)
20x20	8	9	18	41%	3.84(1.85)	0.43 (0.33)
20x20	8	36	2	18%	1.38(1.03)	0.66 (0.65)
20x20	8	36	4	36%	4.1 (1.56)	0.58 (0.52)

Table 4-2. The effects of perimeter compactness on the representation of clustered minorities.[140]

(10,000 Samples were performed for each grid/district combination)

---

[140] I generated ten thousand district plan-population combinations for each set of district and population parameters. Since I controlled for the other parameters by keeping them constant across runs, a standard correlation measure adequately represents the linear association between the number of districts captured by a minority and the compactness of districting plans. Genetic algorithms, hill climbing and descent methods were used for the 5 by 5 case, while hill-climbing and descent methods were used for all other cases. Since the results from each method were similar, only hill-climbing results are reported.

As I had suggested earlier, the electoral effects of compactness depend upon the geographic distribution of political groups. When all political groups are thoroughly geographically mixed, no district, compact or not, can contain a majority of a minority group. Even when political groups form small geographic clusters, if these clusters are dispersed geographically, as in the uniform and normal population distribution models, then compact districts are no more likely to elect candidates from one political group than from another. This is not because the district drawing process is politically neutral in these circumstances. On the contrary, in these cases compactness has no effect on electoral outcomes because geographical districting itself embodies such a powerful majoritarian bias[141] that minority political groups are unlikely to win seats under *any* circumstances.[142]

---

[141] I use the term "bias" here in the descriptive, rather than normative sense — arbitrarily drawn districts tend to award district-share in excess of the majority's share of the population. See Grofman (1982) for a similar analysis of the majoritarian bias inherent in redistricting.

[142] Also important will be the relative geographic size of the district, minority clusters, and census blocs. For example, census blocs that were large relative to the size of minority clusters could make it difficult or impossible to create a minority-majority district, even intentionally.

Consider the 12th entry in the table, which describes a simulation run on a 20 by 20 grid where the minority political group populated 100 of four hundred census blocs in the state. Although the minority group makes up 25 percent of the voting population of the state, it loses nearly every election simply because it has the misfortune not to be geographically concentrated. The only way for minority political groups to win any seats in these circumstances would be if we tailored districts expressly to their boundaries, linking small concentrated clusters or minorities. While districts created in this way will almost certainly be noncompact, for such purposes *non*compactness will be a necessary condition, but almost never a sufficient one. Compactness makes it impossible for dispersed minority political groups to gain representation. I will return to this issue in Section 4.3.2.

These results may understate the electoral effects of compactness on minority representation when we consider the assumptions we made in the simulation about turnout. In the simulations, we assumed that each political group turned out to vote at the same rate and voted strictly for their own party. This is a simplification that helped to reveal the general dynamics of voting in compact districts, but which may bias our predictions. In particular, if the minorities that are geographically dispersed also turn out at a lower rate than the majority political group, or if they have a higher rate of cross-over voting, it will be even more difficult to draw compact districts that would allow minorities an opportunity to elect a candidate of choice.

In contrast, compactness helps combat majority bias when minority political groups are geographically concentrated. As Figure 4-5 shows, there is a strong positive correlation between a plan's compactness and the number of seats captured by such a political group. Why do we see such a correlation? The explanation for this is straightforward: Fortuitously, when both districts and minorities are very compact, a concentrated minority will sometimes fall completely within the district lines of the "optimal" plan. By contrast, under any other circumstances short of a purposeful minority gerrymander, the majoritarian bias inherent in geographical redistricting makes minority controlled districts extremely unlikely (Figure 4-5).

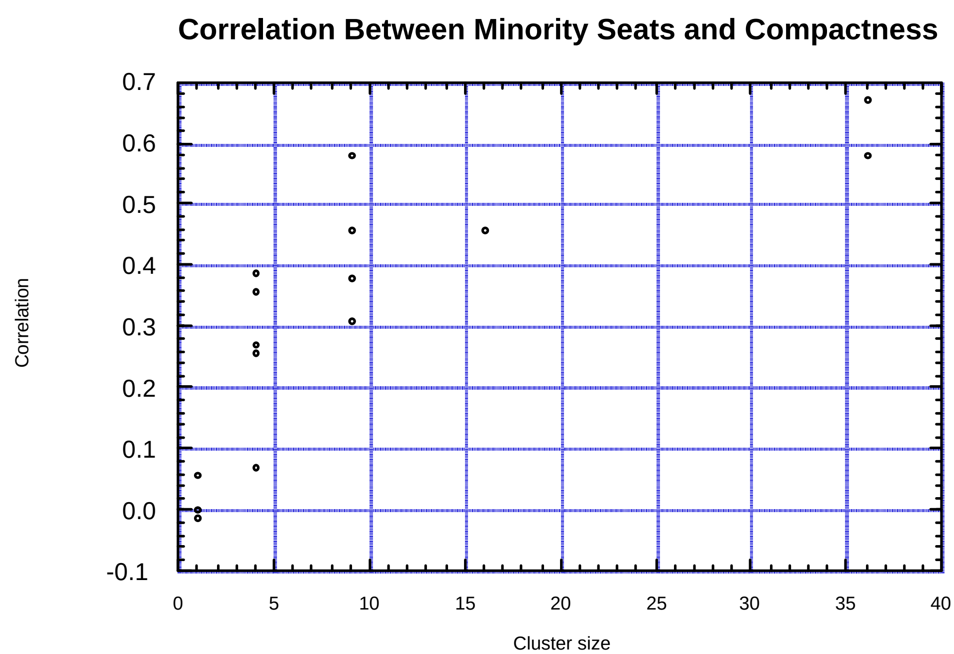


Figure 4-5. Correlation between minority seats and perimeter compactness (based on data in Table 4-2).

This phenomenon is not isolated to perimeter compactness, to small numbers of districts, or to compact clusters. Table 4-3 shows us the same patterns when we use the moment-of-inertia measure for compactness. Table 4-4 shows us somewhat weaker patterns when minorities are grouped in less compact clusters and into more districts.

Grid Size	Number of Districts	Cluster Size	Number of Clusters	Minority Percentage of Population	Mean Minority Controlled Districts (std. dev.)	Correlation Between Minority Controlled Districts and Compactness (Somers's d)
8x8	4	1	26	40%	0.47 (0.53)	-0.01 (-0.01)
8x8	4	4	6	38%	0.48 (0.55)	0.14 (0.15)
8x8	4	9	3	42%	1.2 (0.73)	0.52 (0.52)
8x8	4	16	1	25%	0.41 (0.49)	0.53 (0.57)
8x10	8	9	3	34%	1.38 (0.93)	0.62 (0.55)
20x20	8	4	40	40%	0.76 (0.72)	0.15 (0.12)
20x20	8	9	18	41%	1.46 (0.86)	0.25 (0.19)

Table 4-3. The effects of moment-of-inertia compactness on the representation of clustered minorities.

(10,000 Samples were performed for each grid/district combination.)[143]

[143] Hill-climbing and descent methods were used for these cases. Since the results from each method were similar, only hill-climbing results are reported.

	Grid Size	*Number of Districts*	*Minority Percentage of Population*	*Mean Minority Controlled Districts (std dev)*	*Correlation Between Minority Controlled Districts and Compactness (Somers's d)*
Perimeter	8x10	4	34	0.33 (0.48)	0.10 (0.12)
	20x20	16	10	0 (0)	0 (0)
	20x20	16	25	0.08 (0.28)	0.08 (0.16)
	20x20	16	40	2.88 (1.14)	0.25 (0.19)
Moment	8x10	4	34	0.37 (0.49)	0.01 (0)
	20x20	16	10	0 (0)	0 (0)
	20x20	16	25	0.08 (0.28)	0.04 (0.06)
	20x20	16	40	2.87 (1.11)	0.13 (0.09)

Table 4-4. The effects of compactness on the representation of Schelling-distributed minorities.

(10,000 Samples were performed for each grid/district combination.)[144]

Compact Partisan Gerrymandering

The previous section examines the effects of compactness when the creation of district plans is arbitrary, in one common sense of the word. Arbitrary district plans, although a distinct possibility in the future, have yet to come into wide practice. On the other hand, many authors past and present claim that partisan gerrymandering is widespread (Congressional Quarterly Staff 1993; Griffith 1974). Furthermore, although there are other types of gerrymandering (incumbent gerrymandering, for example), partisan gerrymandering is likely to become more pervasive as term limits and spending

---

[144] Hill-climbing and descent methods were used for these cases. Since the results from each method were similar, only hill-climbing results are reported.

limits weaken the individual incumbent relative the political party. What are the effects of compactness when it is applied to partisan gerrymanders?

Obviously, if one party is in complete control of drawing districts all of the time and in all places, its ability to create districts will be limited by practically any restriction, compactness included. Suppose, however, that different parties substantially control the redistricting process at different times and different places. If the courts continue to use compactness as a red flag to mark plans for judicial review, the party in control may try to produce the most compact partisan gerrymander that they can. How will compactness affect electoral results in these circumstances?

Table 4-5 and Table 4-6 compare the relationship between compactness and seats for the minority party when compact districts are arbitrarily selected to the case where partisans try to gerrymander compactly. In this second case, I altered the simulations to find, for each party, the most compact plan subject to the constraint of partisan seat maximization.[145]

---

[145] Gerrymandering to capture the maximum number of seats is only one possible partisan objective. In the real world, where there is uncertainty over voters' behavior, partisans might try to maximize the probability of controlling the legislature instead. I use seat maximization in this simulation because I have assumed certainty, and because for some minority population distributions, it may be impossible a-priori to capture the legislature.

The effects in these tables are somewhat more complicated than in Table 4-3 and Table 4-4. Here we see two major effects. First, we see the effect that we would expect to see, given these tables — if the minority party is populous and compact enough, the minority party benefits from a compactness rule: They will be able to produce maximal gerrymanders that are more compact, on average, than the maximal gerrymanders for the majority party, given the same population distribution. (See rows 2–4, 9, and 12–14 in Table 4-5.)

	Grid Size	*Number of Districts*	*Cluster Size*	*Number of Clusters*	*Minority Percentage of Population*	*ARBITRARY PLANS Correlation Between Compactness and Minority Controlled Districts (Somers's d)*	*COMPACT GERRY - MANDER Correlation Between Compactness and Minority Controlled Districts (Somers's d)*
Perimeter	8x8	4	1	26	40%	0.01 (-0.02)	-0.20 (-0.19)
	8x8	4	9	3	42%	0.46 (0.47)	0.23 (0.22)
	8x8	4	16	1	25%	0.46 (0.61)	0.18 (0.22)
	8x10	8	9	3	34%	0.58 (0.52)	0.19 (0.19)
	20x20	8	1	100	25%	0.0 (0.0)	0.00 (0.00)
	20x20	8	4	40	40%	0.29 (0.23)	-0.23 (-0.22)
	20x20	8	9	18	41%	0.43 (0.33)	-0.02 (-0.02)
	20x20	8	36	2	18%	0.66 (0.65)	-0.15 (-0.27)
	20x20	8	36	4	36%	0.58 (0.52)	0.06 (0.06)
Moment of Inertia	8x8	4	1	26	40%	-0.01 (-0.01)	-0.58 (-0.65)
	8x8	4	4	6	38%	0.14 (0.15)	-0.45 (-0.44)
	8x8	4	9	3	42%	0.52 (0.52)	0.14 (0.16)
	8x8	4	16	1	25%	0.53 (0.57)	0.09 (0.12)
	8x10	8	9	3	34%	0.62 (0.55)	0.27 (0.24)
	20x20	8	4	40	40%	0.15 (0.12)	-0.63 (-0.64)
	20x20	8	9	18	41%	0.25 (0.19)	-0.56 (-0.55)

Table 4-5. The effects of compact gerrymanders on clustered minorities.[146]

(500 samples were for each grid/district combination.)

[146] Hill-climbing was used for these cases.

	Grid Size	*Number of Districts*	*Minority Percentage of Population*	*ARBITRARY PLANS Correlation Between Compactness and Minority Controlled Districts (Somers's d)*	*COMPACT GERRYMANDER Correlation Between Compactness and Minority Controlled Districts (Somers's d)*
Perimeter	8x10	4	34	0.10 (0.12)	-0.16 (-0.13)
	20x20	16	10	0 (0)	0.00 (0.00)
	20x20	16	25	0.08 (0.16)	0.09 (0.08)
	20x20	16	40	0.25 (0.19)	0.18 (0.16)
Moment of Inertia	8x10	4	34	0.01 (0)	-0.47 (-0.46)
	20x20	16	10	0 (0)	0.00 (0.00)
	20x20	16	25	0.04 (0.06)	-0.51 (-0.50)
	20x20	16	40	0.13 (0.09)	-0.73 (-0.67)

Table 4-6. The effects of compact gerrymanders on Schelling-distributed minorities.

(500 samples were for each grid/district combination.)[147]

On the other hand, remember that if the minority party is weak or dispersed, compactness did not help them very much when districts were created automatically. As I noted in Section 0, the minority would need noncompact districts to capture seats. Hence we see in these cases that compactness harms the minority party: The majority party will be able to produce maximal gerrymanders that look much better than the gerrymanders produced by the minority party.[148]

---

[147] Hill-climbing was used for these cases.

[148] Note that compactness can obviously have no effect where the minority party population is small enough and scattered enough that they cannot capture any districts,

4.4. Discussion

The simulation shows us features of compactness standards. First, compactness effects are nonlinear. Electoral manipulation is much more severely constrained by high compactness than by moderate compactness. Any empirical study of the relationship between gerrymandering and compactness must use models that can accommodate these nonlinearities.

Second, compactness effects are context-dependent. The difficulty of drawing compact plans is significantly affected by the shape of the state being divided, as well as by the compactness measure used. Similarly, differences in population geography may affect the difficulty of drawing compact, equal-population plans. Therefore, comparisons of compactness between states are misleading: statistical studies of the electoral effects of compactness should take this into account – for example, using time-series rather than comparing compactness across states.

Third, compactness standards can have asymmetric effects on different political groups if those groups are distributed in geographically different ways. The geography of district lines alone is not sufficient to diagnose a gerrymander: A majority which is purposefully attempting to exclude a large but geographically diffuse political minority from the political arena will want to draw districts that are as compact as possible — even if given substantial opportunity to gerrymander. (See row 5 in Table 4-5 and rows 2 and 6 in Table 4-6.)

whereas the same majority, with the same purpose, facing a geographically concentrated minority, will want to draw noncompact districts. To correctly predict the electoral effects of a set of district lines, you must know the geographical distribution of all the relevant political groups. Geography does matter — but you must interpret it within a political context.

Fourth, this study shows that compactness can also significantly disadvantage geographically *concentrated* minorities, in the context of a partisan gerrymander. We might expect compactness to have the greatest political effect on racial and ethnic minorities, and on the parts of the Democratic Party that they disproportionately support, because, these minorities are disproportionately concentrated in large cities.[149]

The electoral effects of compactness, however, will not be limited to these groups; the simulations results apply in general to any large minority group that is politically cohesive and geographically concentrated – including any "community of interest" that has these characteristics. The argument that "communities of interest" that are *not* compact are likely to be hurt by compactness standards is fairly straightforward. Minorities that are geographically dispersed are already at a disadvantage in the redistricting process, especially where the building blocks for districts are large, and

---

[149] Only 19% of the white population resides in cities of 100,000 or more, while 57% of African-Americans, 56% of Hispanics and 53% of Asian-American's reside in these dense urban areas (G.P.O. 1993).

compactness is likely to make this disadvantage even greater. What has not previously been noted, however, and what these simulations show, is that even “communities of interest” that are compact can be systematically damaged by compactness standards, in some institutional environments. This contradicts the optimistic assumption that some proponents of compactness have made (Polsby and Popper 1993), that compact communities are likely to benefit from compactness standards.

(Allegations that compactness requirements harm dispersed minorities are often based on the implicit assumption that in the absence of a compactness requirement, districts will be drawn so as to maximize the total number of minority districts. If this is the case, any requirements on districting at all will tend to reduce minority seats, as I discuss in Section 4.3.)

The Court has declared that optimal compactness is not required -- redistricting is not a “beauty contest.” If compactness is to be used at all, however, such beauty contests are going to be difficult to avoid, for several reasons: The extent to which compactness is likely to vary dramatically with the particular level of compactness that is considered “good enough.” Moreover, to set this threshold appropriately, the Court would have to consider separately the particulars of each measure of compactness, and the geography of each state. The only practical method of implementing effective compactness measures are likely to require the beauty contests that the court wishes to avoid.

Such beauty contests would have little to recommend them. Compactness standards, rather than being the neutral standard that the court envisions, are likely to have distinctly partisan effects. These simulation results contradicts the view of compactness advocates and bears out Lowenstein's (1985) assertion that compactness is not a partisan-neutral standard because of the way that Democrats are concentrated geographically.

These simulations also raise questions for the use of ill-compactness as evidence of political gerrymandering. Parties with distributed support are likely to able to draw districts to their advantage without violating triggering any alarms.

The effects of compactness rules will also depend upon an important factor that the Court has largely ignored in recent cases -- the larger institutional mechanism for drawing districts. Compactness rules will have different effects — rules that benefit geographically concentrated minority parties under a system of automated redistricting may have different effects under a system dominated by partisan gerrymandering. The apparent mathematical universality of compactness rules is, in fact, illusory — to understand the political effects of compactness requires an intensely local appraisal.

Appendix: Computer Techniques

The optimization routines used in this chapter were written in the *C* language, and run on a variety of Unix workstations. Although, the *C* implementations were, for the most part, specific to this chapter, they are based on a number of publicly documented algorithms. This appendix details the Monte Carlo, genetic, simulated annealing, and genetic algorithm used in the chapter.

The Monte Carlo procedure I used was quite simple — I create a district by adding population units at random, until the districts are approximately the correct size. Hill climbing can build on this random assignment.

My hill climbing procedures used the “greedy” method described in (Nagel 1965): each round the program examines every possible combination of single shifts (moving a census block from one district to another) and single trades (i.e., a pair of compensating shifts between two districts). After examining all possibilities, the trade that most improved the plan was executed, and the process repeated. The program stopped when it could find no further improvement to make. Reverse hill climbing operated similarly, except that first changes were made to decrease fitness of the maximally compact plan, until the program reached the target level of non-compactness. And second, to avoid deterministically reusing the same set of non-compact plans, and the fitness decreasing changes were chosen randomly rather than greedily.

The genetic algorithm optimizer was based upon the algorithms described in Goldberg (1989). Following this terminology, each district was encoded as a haploid string with separate positional information (i.e., a string of {district assignment, census block} pairs). Each population (population=750) of strings was subjected to mutation (p=.001), inversion (p=.01), and pmx-crossover (p=.05), as described in Goldberg. Strings were then chosen for the next generation by repeated random drawing (with replacement) of pairs of strings, the fitter string succeeding to the next generation. This process was repeated for one hundred generations.

Simulated annealing code came directly from Lester Ingber's Adaptive Simulated Annealing (ASA) package (v. 2.2), as described in (Ingber 1989). I started by representing each census block as an integer, in the ASA framework, which was assigned a district number by the annealing process. In view of the poor performance of the annealing algorithm, I then modified the code slightly, to create a "combinatoric" type to represent each census block. The only difference between the "integer" and "combinatoric" types was that for combinatoric types only the probability of the variable being changed lowered with the annealing temperature; whereas for integer types, the probablilty and the magnitude of the change lowered. The logic behind the change was that since district numbers are arbitrary, changing a census bloc from district 1 to district 2 is not necessarily a "smaller" change, in any relevant measure, than shifting a block from district 1 to district 10. These results are reported in the chapter, and were only slightly better than the original results using integers.

The simulation programs in this chapter used uniform and normal random numbers. Simulations are often susceptible to quirks in random number generators, and the random number generators in many standard *C* libraries are poorly written and non-portable across platforms. Because of this, I used the portable *Ranlib* package written by Barry W. Brown, and James Lovato at the Department of Biomathematics, The University of Texas, M.D. Anderson Cancer Center. The algorithms that this package uses to generate uniform and normal random number routines are described in (L'Ecuyer and Cote 1991) and (Ahrens and Dieter 1973).

Chapter 5. Is Automation the Answer? — The Computational Complexity of Automated Redistricting

5.1. Redistricting and Computers

> "There is only one way to do reapportionment — feed into the computer all the factors except political registration." - Ronald Reagan (Goff 1972)
>
> "The rapid advances in computer technology and education during the last two decades make it relatively simple to draw contiguous districts of equal population [and] at the same time to further whatever secondary goals the State has." - Justice Brennan, in *Karcher* v. *Daggett* (1983)

Ronald Reagan was not the only recent politician or academic to assert that computers can remove the controversy and politics from redistricting. Computers can prevent gerrymandering by finding the "optimal" districting plan, given any set of values that can be specified, claim proponents of automated redistricting. The Supreme Court seems to express a similar sentiment with the emphasis that it put on such mechanical principles as contiguity and compactness in the recent redistricting cases of *Miller* v. *Johnson* (1995) and *Shaw* v. *Reno* (1993).

In Chapter 4, I showed that the mechanical application of compactness standards has previously unanticipated partisan consequences; in this chapter I examine the general automation of the districting process, and its consequences. Will we soon be able to write out a function that captures the social value of a districting arrangement, plug this function into a computer, and wait for the "optimal" redistricting plan to emerge from our laser-printers? I argue that this rosy future is unlikely to be realized soon, if at all, because we are unlikely to solve three problems that face automated redistricting.

In Section 5.2, I show that current redistricting methods are not adequate for the purposes of automated redistricting. Current automation techniques must resort to unproved guesswork in order to handle the size of real redistricting plans. Before automated redistricting produces trustworthy results, large gaps must be filled.

Proponents of automation assume that despite current shortcomings, finding the optimal redistricting plan simply requires the development of faster computers. In Sections 5.3 and 5.4, I show that this assumption is false — in general, redistricting is a far more difficult mathematical problem than has been yet recognized. In fact, the redistricting problem is so computationally difficult that it is unlikely that any mere increase in the speed of computers will enable us to solve it.

Even if these complexities are overcome, automated redistricting faces a serious limitation: To use automated redistricting we must write out a function that meaningfully captures the social worth of districts, and that at the same time can be put into terms rigid enough for computer processing. In Section 5.5, I argue that if we do this we will have to ignore values that are based upon subtle patterns of community and representation, which cannot be captured mechanically.

5.2. Current Research on Automated Redistricting

Although the literature on automated redistricting is at least thirty-five years old, it has seen a recent resurgence. This research generally falls into two categories: The first category addresses the merits of automated redistricting *per se*, and the second category suggests methods we can use to create districts automatically.[150] In this section I briefly summarize the previous research in each of these two categories.

5.2.1. Arguments for Automating the Redistricting Process

In one of the earliest papers on this subject, Vickrey proposed that districting be automated, and that this automation process be based upon two specific values: population equality and geographical compactness. Under his proposal political actors would be permitted to specify or add criteria to a goal function for redistricting, but they would not be permitted to submit specific redistricting plans. Then plans would be created *automatically*, with no further human input, from census blocks, to meet the goal

---

[150] There is, as well, a third category of literature which indirectly touches on automated redistricting. Authors in this third category typically suggest a *particular* criteria for drawing optimal districts— much of the literature on geographical compactness falls into this category.

In particular, several authors have argued that gerrymandering can be eliminated by drawing districts which are maximally compact (Harris 1964; Kaiser 1966; Polsby and Popper 1991; Stern 1974; Wells 1982). (Also see Young (1988) Niemi et al. (1991) for a survey of other compactness measures.) While these authors focus primarily on the *criteria* for evaluating districts, their core argument is the same as that examined above — i.e., that redistricting can be performed best by automatically optimizing a pre-specified representation function.

function. At its heart, automated redistricting is an attempt to push all decision making to the beginning of the redistricting process.

Vickrey (1961) asserted that automated redistricting provides a simple and straightforward way to eliminate gerrymandering. More recently, Browdy (1990a) followed and extended Vickrey's arguments, and created what seems to be the best case for automated redistricting. Five main arguments are offered in the literature, and these can be easily summarized:

Argument 1. Automated redistricting, in and of itself, creates a *neutral* and unbiased district map (Forrest 1964; Harris 1964; Kaiser 1966).

Argument 2. Automated redistricting *prevents manipulation* by denying political actors the opportunity to choose district plans, while simultaneously producing districts that meet specified social goals (Browdy 1990a; Stern 1974; Torricelli and Porter 1979; Vickrey 1961; Wells 1982).[151]

Argument 3. Automated redistricting *promotes fair outcomes* by forcing political debate to be over the general goals of redistricting, and not over particular plans, where selfish interests are most likely to be manifest (Vickrey 1961).[152]

---

[151] Compactness advocates make a similar argument.

[152] In the compactness literature quoted above, it is argued that the compactness criteria themselves make the automated process fair.

Argument 4. Automated procedures provide a *recognizably fair process* of meeting any representational goals that are chosen by the political process[153] (Browdy 1990a; Vickrey 1961). Browdy also argues that such procedural fairness will help to curtail legal challenges to district plans.

Argument 5. Automated redistricting *eases judicial and public review* because the goals and methods of the districting process are open to view; and because automation process creates a clear separation between the intent and effect of redistricting (Browdy 1990a; Issacharoff 1993) .

5.2.2. Criticisms of Automated Redistricting

Automated redistricting has been criticized as well as praised. Previous authors have raised two central objections to automated redistricting.[154] The first argument against

---

[153] Polsby and Popper (1991) argues similarly that a mechanistic application of formal compactness standards is inherently fair.

[154] There are also a number of papers arguing against particular formal measurements, rather than against automated redistricting. Specifically, compactness standards have been subjected to intense scrutiny. For an introduction to some of the issues surrounding the use of these standards, see Lijphart 1989, Lowenstein and Steinberg 1985, Mayhew 1970 and Chapter 3 in Cain 1984. As most of these arguments are directed against the use of particular geographical criteria and not against automated redistricting in general, I have not included these papers in the preceding summary.

automatic redistricting was originally expressed by Appel (1965). He protested that automated redistricting should not be viewed as inherently objective. He argues that redistricting standards and processes embody political values and that automation of this process hides the fundamental conflict over values. Dixon (1968) as well, pointed out that automated processes, even if based on nonpolitical criteria, may have politically significant results. More recently, Anderson and Dahlstrom (1990) cautioned that political consequences of redistricting goals makes redistricting, whether it is automated or not, inescapably political.

I believe this objection to be both correct and unavoidable. Automation is a process for obtaining a given set of redistricting goals. Neutrality, however, is a function of *three factors*, the process selected, the goals themselves, and the effects of seeking to obtain those goals in a particular set of demographic and political circumstances. There is no general consensus over what "objectively neutral" goals are, or whether they exist[155] at all; therefore, no amount of automation can make the redistricting process "objectively neutral."

---

[155] Much doubt has been expressed as to whether such goals exist. Furthermore, the fundamental conflicts between some of the more commonly proposed goals make such a consensus unlikely. For a discussion of the most commonly proposed goals and some the conflicts between these, see (Cain 1984; Dixon 1968; Grofman 1985; Lijphart 1989).

This objection, however, only applies to the first, and most extreme, claim for automated redistricting. Many proponents of automated redistricting do not make this type of extreme claim, and instead explicitly acknowledge the political nature of redistricting goals. They propose that the automated process be used to neutrally and effectively meet goals generated previously by a political process (Browdy 1990a; Issacharoff 1993; Vickrey 1961). This proposal seems to meet at least the first objection above.[156]

Anderson, echoing Dixon (Dixon 1968), made a second argument against automated redistricting by drawing attention to the legislative process used to select automatically generated plans. He argues that the legislature's willingness to accept the plans that are generated by an automated process will be politically motivated — reintroducing political bias into the process (Anderson and Dahlstrom 1990).

While I believe Anderson to be correct, this specific objection does not seem to me to be strong. If we mandate that the legislature must accept the results of the automatic process, we can prevent this particular attempt to reintroduce bias into the system. In general, however, I believe that researchers have largely underestimated the potential for political biases to become part of the automation process.

[156] And, indeed, Dixon, who argues against the neutrality of automated redistricting, freely acknowledges its usefulness for this type of situation (Dixon 1968).

5.2.3. *The Core Argument — Automation as a "Veil of Ignorance"*

In the arguments for automated redistricting, automation essentially plays the role of a Rawlsian (Rawls 1971) "veil of ignorance" which creates fairness by hiding each actor's position in the final outcome. Like the Rawlsian version, the "veil of automation" attempts to hide the final outcome (i.e., redistricting plans) from those bargaining over the social contract (i.e., redistricting goals and procedures). Like the more general veil of ignorance, the automation process claims to prevent manipulation by promoting a recognizably fair method that will, on average, promote fair outcomes.

Vickrey, in one of the first arguments for automated redistricting, particularly emphasized that in order for the automated process to be successful at promoting fairness, it must be *sufficiently unpredictable*[157] — it should not be possible for political actors to deduce the results of the redistricting goals over which they bargain (Vickrey 1961). This property is essential, for if it does not obtain, then the choice of objective functions collapses into a choice of individual plans, and the incentive to gerrymander

---

[157] Here I use the term "unpredictable" where Vickrey originally used "random." Vickrey's concern was that it should not be obvious to the political actors what exact results derive from a particular value function. Enough randomness in the process would certainly ensure this concern is met, but the process need not be random to do this. For example, if the process is sufficiently *chaotic*, it is not random, but it may still be, for all intents and purposes, unpredictable — satisfying Vickrey's central concern.

remains unameliorated by the automation process. If we can predict the plans that will result from our values, we can pierce the veil of automation.

While proponents of automated redistricting have recognized this need for unpredictability, they have not mentioned the danger from unpredictable results. An automated process for creating districts in accordance with agreed upon values must predictably achieve (or at least approach) the goals that were agreed upon in the bargaining stage, or lose legitimacy.

The automated redistricting process must maintain a delicate balance. To prevent manipulation while maintaining fairness, the automated process must predictably implement the redistricting goals that we have agreed upon in the bargaining process; but it must be unpredictable in every other dimension that is of interest to the bargaining agents. These are difficult requirements to satisfy when the bargaining agents are individuals who narrowly seek specific, hand-tailored gerrymanders. They become even more difficult to meet when bargaining agents represent interest groups or political parties unconcerned with particular incumbents, because such agents are interested in far more general properties of redistricting plans. Can automated redistricting methods reliably produce plans that exclusively embody any specific set of redistricting goals?

5.2.4. Why Current Methods Are Inadequate for Automated Redistricting

Initially, many researchers expressed optimism about the ease of achieving redistricting goals through automation (Nagel 1965; Torricelli and Porter 1979; Vickrey 1961; Weaver and Hess 1963). Vickrey best captures this initial hopefulness:

> "In summary, elimination of gerrymandering would seem to require the establishment of an automatic and impersonal procedure for carrying out redistricting. It appears to be not all difficult to devise rules for doing this which will produce results not markedly inferior to those which would be arrived at by a genuinely disinterested commission." - William Vickrey (Vickrey 1961)

While optimism has now dulled somewhat, because it has been recognized that purely automated redistricting techniques remain generally unsatisfactory (Backstrom 1982), many authors still assume that automation of the redistricting process is within reach (Anderson and Dahlstrom 1990; Browdy 1990b; Polsby and Popper 1991).

In his original paper, Vickrey sketched a method for performing automated redistricting, but did not give develop a precise implementation of this method. Much of the following work assumed the benefits of automated redistricting and focused primarily on providing criteria and methods to use in such automation. Liittschwager (1973) applied Vickrey's method to the Iowa redistricting process. Similarly Weaver and Hess (1963), and Nagel (1965) developed methods and/or measures for drawing districts in accordance with principles of population equality and geographical compactness.[158] More

---

[158] See also Chapter 6 in Gudgin and Taylor 1979 (Gudgin and Taylor 1979), and Papayanopoulos (1973) for a review of early attempts at automated redistricting.

recently, Browdy (1990b) proposed that the method of *simulated annealing* may be generally applicable to the problem of drawing optimal districts.

Many different methods have been used or suggested for finding optimal districts. We can put these techniques into two broad categories: *exact* methods and *heuristic* methods. In the remainder of this section, I review the methods used to search for optimal districts. Although useful for assisting humans, current methods cannot satisfy the goals of automated redistricting.

Limitations of Exact Methods

Exact methods systematically examine all legal districts, either explicitly or implicitly. Explicit enumeration, or "brute force" search methods literally evaluate every district. More sophisticated methods such as implicit-enumeration, branch-and-bound, or branch-and-cut techniques exclude classes of solutions that can be inferred to be sub-optimal without an explicit examination. Finding the optimal districts in this case is then merely a matter of sorting the list of district scores. These methods have been used by several authors to approach very small redistricting problems (Garfinkel and Nemhauser 1970; Gudgin and Taylor 1979, Chapter 6; Papayanopoulos 1973; Shepherd and Jenkins 1970).[159]

---

[159] A close examination of these algorithms reveals that in order to make enumeration complete in a feasible amount of time, "short-cuts" are used where some sub-classes of partitions are *assumed* to be unreasonable, and are disregarded without examination and

Exact methods have two major shortcomings: First, no exact method has been developed that will solve redistricting problems for a reasonably sized plan; and as I will show in Section 4, the mathematical structure of the redistricting problem makes it unlikely that exact methods will be developed in the future that will be able to solve reasonably sized plans. Second, even if we find an exact method that works for real plans, then anyone could use that method to determine the precise plan corresponding to a particular set of redistricting goals; thus, exact methods would have completely predictable results, violating our requirements for an automated redistricting method.

Limitations of Heuristic Methods.

Heuristic procedures use a variety of methods to structure the search for high-valued redistricting plans. None of the heuristic algorithms guarantees convergence to the optimal district plan in a finite amount of time. At best, they are good guesses.

All of the general redistricting heuristics cited in the literature are based upon making iterative improvements[160] to a proposed redistricting plan. The single most

---

without proof of sub-optimality. Restrictions such as "exclusion distance" in Garfinkel and Nemhauser (1970) or limiting examination to "amalgamations" in Shepherd and Jenkins (1970) must be formally classified as heuristic rather than exhaustive.

[160] In the field of computer science, heuristic algorithms are divided into two categories: *iterative improvement*, as above, and *divide-and-conquer* methods. These two

popular method seems to be *hill climbing* and its variants, although a few researchers add more sophisticated features of *neighborhood search* techniques:[161]

• *Hill climbing* methods work by making small improvements on a potential solution until a local optimum is reached. Hill climbing often starts with the current district plan or with a randomly generated plan, and it makes improvements through repeatedly trading[162] census blocks between districts (Moshman and Kokiko 1973; Nagel 1965). Another variant of this method selects arbitrary census tracts to form the nuclei of each district, and then repeatedly adds tracts that most improve[163] the current district until each district is

---

general broad categories include variants such as: simulated annealing and genetic algorithms, which will be described in Section 5.4.

[161] In addition to general redistricting methods, there are a number of special purpose methods of note: Tobler develops an iterative graphical remapping process to generate districts of equal population (Tobler 1973). This process gradually distorts geographical maps to create new maps where population is equivalent to area, facilitating the creation of districts with equal population.

[162] Usually, improvements are made sequentially for each district, but if the number of census tracts is small, several trades may be examined simultaneously.

[163] Often this is simply the tract that is closest to the selected district center (in whatever metric used). The Weaver and Hess algorithm uses linear-programming techniques to select population units to add to the district center.

fully populated (Bodin 1973; Liittschwager 1973; Rose Institute of State and Local Government 1980; Taylor 1973; Vickrey 1961; Weaver and Hess 1963).

• *Neighborhood search* methods are all similar in that they seek to improve potential solutions by examining the value of “nearby” solutions; in this they are similar to hill climbing. Unlike hill climbing algorithms, sophisticated techniques in this class use various techniques to attempt to avoid becoming stuck at local optima. Browdy (1990b) suggests the use of *simulated annealing*,[164] a member of this class of methods.

The main difficulty with all heuristics is that they are, at heart, informed guessing procedures. This is not necessarily bad — when you are faced with a difficult problem, you may be able to find an adequate solution cheaply much of the time by guessing. If we use guessing procedures to decide political questions, however, we must show that our guesses are unbiased and likely to produce good solutions. No researcher in this field has been able to show, either theoretically or empirically, that the districts produced by their methods are near optimal, or that they are unbiased. On the contrary, many heuristic methods produced results that are clearly sub-optimal (Bodin 1973), or depend strongly

[164] This and other neighborhood search algorithms will be discussed in more detail in Section 4.

on starting conditions (Browdy 1990b; Nagel 1965; Weaver 1970; Weaver and Hess 1963).[165]

Are the inadequacies of current automated redistricting techniques merely temporary? Will improvements in software design and increasingly powerful computers make automated redistricting easy? In the next section, I will show that automated redistricting has been unsuccessful not only because of current techniques but because of inherent complexities in the structure of the redistricting problem. Furthermore, I will show that these complexities are unlikely be overcome simply through the use of faster hardware or more clever software.

5.3. Automated Redistricting May be Intractable

To find “optimal” redistricting plans, as the advocates of automated redistricting suggest, we must first formulate the redistricting problem mathematical terms and then solve this mathematical problem. In this section, I will show that, regardless of the formulation, the redistricting problem is formally computationally intractable — it is practically impossible to solve exactly.

[165] Note that Browdy’s proposal for using *simulated annealing* has yet to be implemented. I will discuss this suggestion in detail in Section 5.4.

5.3.1. Redistricting is a Large Mathematical Problem

We can mathematically characterize the redistricting problem in a number of different ways. One simple way to mathematize redistricting is to think of it as a *set partitioning* problem. I will use this particular characterization extensively in this chapter. While there are other characterizations that we could use, such as *graph partition, polygonal dissection* and *integer programming* (See the appendix to this chapter.), the results in this section are not dependent on the characterization we choose, since these characterizations are computationally equivalent. (See Section 5.4.)

In particular, we will characterize redistricting as a combinatorial optimization problem:[166] Imagine that census blocks are indivisible,[167] and that you have complete information about voting and demographic information for every census block in your

[166] This in itself is not original to this chapter. Redistricting has been implicitly characterized as a combinatorial optimization problem from Vickrey (1961) onward. See Gudgin and Taylor (1979) and Papayanoupoulos (1973) or previous explicit characterizations of this problem.

[167] Population units are assumed to reflect the most accurate and detailed information practically and/or legally available, unless otherwise specified. For most of this chapter, population units can be read as “census blocks” without too much loss of generality.

state. The redistricting problem is to *partition*[168] the entire set of units into districts such that a *value function* is maximized.[169] This partitioning problem may be complicated by the addition of a set of constraints on districts. These constraints, such as contiguity, may limit the set of legal plans.[170]

---

[168] A partition divides a set into component groups which are exhaustive and exclusive. More formally:

For any set $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$, a *partition* is defined as
a set of sets $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_k\}$ $s.t.$
(1) $\forall x_i \in \mathbf{x},\ \exists \mathbf{y}_j \in \mathbf{Y}, s.t.\ x_i \in \mathbf{y}_j$
(2) $\forall i, \forall j \neq i, \mathbf{y}_j \cap y_i = \varnothing$

[169] More formally:

Given:
• a set of census blocks $\mathbf{x}$
• the set of all partitions of $\mathbf{x}$, Y
• a value function on partitions, V(y)
The optimal district plan is
$$\mathrm{D}^* = \max_{y \in Y}(V(y))$$

[170] These can be represented as formal constraints on membership in the set of allowable partitions in note 17 above, or may for some approximations, simply be incorporated in the value function to be optimized.

The redistricting problem poses special difficulties because the size of the solution set can be enormous. In general, it will be impossible to attack the problem by a brute force search through all possible districting arrangements.

Formally, the total number of distinct[171] plans that can be created using *n* population blocks to draw *r* districts is characterized by the function:[172]

$$S(n,r) = \frac{1}{r!}\sum_{i=0}^{r}(-1)^i\left(\frac{r!}{(r-i)!\,i!}\right)(r-i)^n$$

Even under the assumption that each district is composed of exactly *k* population blocks[173] (hence, $r = n/k$) the number of possible plans is still a rapidly growing function:

$$S'(n,r,k) = \frac{n!}{k!(r!)^k}.$$

The magnitude of this problem is often not fully recognized. For even a small number of census tracts and districts, the number of possible districting arrangements

---

[171] This number reflects districts that are distinct, ignoring the numbering order of districts. Merely renumbering the districts without changing the composition of at least one district does not result in a different plan.

[172] *S* is known as a "Stirling Number of the Second Kind." See Even (1973) for a good introduction.

[173] Which is not correct, but closer to the real situation than the formula above.

becomes enormous. As an example of this, Table 5-1 lists the number of possible plans that could be used to divide a small hypothetical state into two districts:

Type of Population Block	*Total Number of Blocks*	*Blocks per District*	*Number of Plans*
counties	10	5	945
census tracts	50	25	$5.8 * 10^{31}$
census blocks	250	125	$4 * 10^{245}$

Table 5-1. Number of plans available to divide a hypothetical area into two districts, by type of population block.

As Table 5-1 shows, the size of the redistricting problem grows rapidly as a function of the number of population units being used. In fact, this table understates the size of the problem, because it assumes all districts have an identical numbers of blocks.

The number of districts, *r*, is also an important factor in determining how many plans are possible. The number of plans possible will increase in *r* up to a point, and then decrease, as Table 5-2 shows:

Number of Districts	*Blocks per District*	*Number of Plans*
1	24	1
2	12	$3.2 * 10^{11}$
3	8	$9.2 * 10^{12}$
4	6	$4.5 * 10^{12}$
6	4	$9.6 * 10^{10}$
8	3	$1.6 * 10^{9}$
12	2	$1.3 * 10^{6}$
24	1	1

Table 5-2. Number of Plans possible when dividing 24 population blocks evenly into *N* districts.

An exhaustive search to find the optimal plan will be impractical for all but the coarsest population units and extreme number of districts per census block. Only by using a large population unit, such as a county, for our “indivisible” units, can we make exhaustive search manageable. Unfortunately, using such coarse granularity is likely to substantially decrease the quality of our solutions. Furthermore, use of such coarse population divisions is unlikely to lead to solutions where even rough population equality is maintained between districts. If we want to draw districts using accurate, fine-grained population units, such as census tracts or blocks, the number of plans involved makes exhaustive searches unmanageable.[174]

Several proponents of automated redistricting, when faced with a prohibitive number of possible plans, suggest that other, nonexhaustive procedures be used to generate districts (Nagel 1972; Papayanopoulos 1973). Certainly, exhaustive search is not necessarily the only method guaranteed to find optimal districts. However, in the remainder this section, we will show that *any method for finding optimal districts* is likely to be computationally hard, and thus impractical for all but the “smallest” redistricting problems.

[174] For a state such as California, where 100,000 census blocks must be assigned to 50 districts, if started at the creation of the universe, a computer that could examine a million districts a second would still not be finished.

5.3.2. Candidate Value Functions

The difficulty of solving the redistricting problem will depend upon the particular value function and constraints we use. In the next section, I summarize the most common candidates for value functions before analyzing the difficulty of the redistricting problem for each one.

While there are practically no political values that are not subject to debate, a number of criteria are commonly thought to be good candidates for redistricting goals. Grofman and Lijphart summarize these, and I list the five most common types below (Grofman 1985; Lijphart 1989):

1. *Population equality* between districts is believed by many scholars to be necessary for political fairness.

2. *Contiguity* has received much attention in combination with compactness.

3. *Compactness,* which attempts to capture the geographic regularity of districts also appears in many state constitutions. Compactness has been defined in many different ways. (See Niemi et al. (1991), for a survey of these.)

4. *Creating fair electoral contests* is another criteria that is sometimes found in state constitutions. Of course, there are many possible definitions of a "fair contest": including *maximal competitiveness* (maximizing the number of close elections),[175] *neutrality* (which

---

[175] Here I group together a number of different types of criteria including: "electoral

specifies that the electoral system should not be biased in favor of any political party in awarding seats for a certain percentage of the vote), and the goal of a *constant swing ratio* (seats/votes share) for each party.

5. The last set of common redistricting goals dealing directly might be termed *representational* goals, as they are difficult to formulate without referring to a concept of representation. These include *protection of communities of interest* and *nondilution* of minority representation.

As well as being theoretically and philosophically important, these values often carry the weight of law (Grofman 1985): The U.S. Supreme Court has found the constitution to require *de minimis* population deviations between Congressional districts and only somewhat larger deviations between state legislative or local government districts. Furthermore, 37 states require districts to be contiguous, 24 states require compactness[176] and 2 states require what might be loosely interpreted as an electoral response function.[177] Representational goals also have some legal force: Protection of communities

---

responsiveness," "neutrality," "competitiveness," and "constant swing ratio," which are often addressed separately in the literature.

[176] Only three states formally define "compactness."

[177] These are vaguely defined in the constitutions of these states as directives to "not unduly favor any person or political party (faction)."

of interest is required in five states, and nondilution of minority interests is required under the Voting Rights Act.

Despite the relative popularity of the four types of goals above, there is no political or academic consensus over them. Nor can formalization or automation somehow make the goals “objective” — the political consequences of redistricting goals will still exist. Although many have recognized this before, it cannot be emphasized too strongly that at best, automation can neutrally implement these goals, once they have been decided upon by a *political* process.[178]

5.3.3. What is a Computationally Hard Problem?

In this section I will show that for any of the aforementioned value functions (or combinations of them), the problem of finding an optimal districting plan is computationally complex - any attempt will probably be thwarted by the size and complexity of the redistricting problem. To prove this result I will have to introduce some formal definitions from computational complexity theory.[179]

---

[178] In this point I agree with two of the main proponents of automated redistricting (Browdy 1990a; Issacharoff 1993).

[179] In order to present the next set of results, it is necessary to define a number of terms and ideas referring to problems, solution methods, and solution complexity. The length of this chapter necessitates that this section be limited to what is essential for

Computational complexity (or "structural complexity") theory and the related field of computability theory are two branches of theoretical computer science. These disciplines are devoted to analyzing the difficulty of solving specified discrete problems using computers.[180]

Researchers in computer science and in operations research use computational complexity theory extensively when they analyze problems. While this type of analysis has been adopted only recently by political scientists, computational tractability is becoming recognized as a prerequisite for practical electoral rules[181]: Kelly (1988a, 1988b) analyzes the complexity of a number of voting rules, and he establishes some conditions for computable electoral rules (Kelly 1988a; Kelly 1988b). Bartholdi, Tovey and Trick analyze the complexity of manipulating elections; they argue that while almost

---

understanding the result. For a more lengthy and formal characterization, see Papadimitriou (1994).

[180] For a review of recent developments in this field, see Book (1994).

For problems that are (unlike partitioning) continuous, rather than discrete, the field of "information-based complexity" also has relevance. For an introduction to this latter field, see Traub and Wozniakowski (1992).

[181] Also see Deng and Papadimitriou (1994) on the complexity of different cooperative solution concepts that are used in some positive political theory models.

all electoral rules are theoretically open to manipulation, some rules may be practically impervious to manipulation because of the complexity of the calculations a manipulator would have to perform (Bartholdi, Tovey and Trick 1989; Bartholdi, Tovey and Trick 1992).

Basic to computational complexity theory is the definition of a problem. A *problem* is a general question to be answered. In the case of redistricting, the problem is to find the districting plan that maximizes our value function — formally, we must find the optimal partition.[182] I will use the term *redistricting sub-problem* to distinguish the case where we have pre-specified a particular value function, such as compactness, rather than taking the value function itself to be a parameter. Hence, redistricting to maximize (a particularly formally defined measure of) population equality is a sub-problem. A problem possesses several *parameters*, or free variables. For any redistricting sub-problem, the parameters consist of the population units from which we are to draw the plan and the vector of values assigned to those population units. An *instance* of a problem is created by assigning values to all parameters. Finding the arrangement of

[182] We can characterize redistricting either as a general problem where the value function itself is a parameter, or as a class of similar partitioning problems, each with a separate value function. Our choice of characterization does not affect the results found below.

Iowa’s 1980 census tracts that maximize population equality would then be an instance of a redistricting sub-problem.

The second set of terms refers to solutions to the preceding problems. An *algorithm* is a general set of instructions, in a formal computer language that, when executed, solves a specified problem. An algorithm is said to *solve* a problem if and only if it can be applied to *any* instance of that problem and is guaranteed to produce an exact solution to that instance. To continue the example above, an algorithm would be said to solve the population-equality redistricting problem only if it were guaranteed to find a population-equality-maximizing solution for any set of census blocks that we put into it.

The final set of terms refers to properties of problems and their solutions. A problem is said to be *computable* if and only if there exists[183] an algorithm which solves the problem.[184] For computable problems we define the *time-complexity* (hereafter

[183] Here I use the term *exists* in the formal, mathematical sense — we do not necessarily have to know which algorithm solves a problem to show that such an algorithm must exist.

[184] Turing (1937) (1937) showed that there are problems for which solutions exist that are not computable in the sense used above. I am not arguing that practical redistricting criteria are likely to be noncomputable, although this is a theoretical possibility.

abbreviated to "complexity") of an algorithm to be a function that represents the number of the instructions that algorithm must execute to reach a solution.[185] The complexity of an algorithm is expressed in terms of the *size* of the problem, roughly equivalent to the number of input parameters.[186] The size of redistricting is simply the number of population units that are used as input. An algorithm is said to take polynomial time if its time-complexity function is a polynomial and is said to take exponential time, otherwise.[187]

---

[185]This definition assumes a serial (single processor) computation model, but the results are not altered if we use parallel-processing: The sum of the time needed by a set of parallel-processors to solve a problem can be no less than the total required in the serial model.

[186] The time complexity of an algorithm is conventionally denoted as $O(f(n))$ where n is the size of the problem. Additive and multiplicative constants are omitted, as these vary with the computing model used. Thus the number of steps to solve an algorithm of $O(n)$ complexity is a linear function of the number of inputs.

[187] In addition to analyzing the time required to solve a problem, we can formulate analogous tractability criteria for the storage space requirements of a problem (or for practically any other of its resource requirements). It can be shown that problems that require exponential space will also require exponential time, but not vice-versa. Fortunately, none of the redistricting sub-problems discussed here needs exponential

A problem is often said to be *computationally tractable* if there exists an algorithm which is of polynomial complexity for all instances and which solves the problem. Conversely, a problem will be said to be *computationally intractable* (also “computationally complex,” or “computationally hard”) if the (provably) optimal algorithm for solving the problem cannot solve all instances in polynomial time.

Although we have defined complexity in terms of time, we may usefully think of it as a measure of *cost* as well. If time is costly, and if there are no exponential economies of scale associated with time, computationally intractable problems will be prohibitively expensive, since the cost to solve such problems will also grow at an exponential rate.[188] Obviously, the time-costs of a redistricting are unlikely to exhibit exponential economies of scale. If anything, wasted time will likely exhibit constant or negative economies of scale: if redistricting takes too long, it will start to disrupt elections seriously.

This characterization of problem difficulty has two main strengths. It is independent of any particular computer hardware design technology and it classifies the difficulty of the problems themselves, not of particular methods used to solve these problems.

---

space.

[188] Alternatively, if we were to use parallel-processors to solve the problem, the *number* of computers would grow exponentially (at least)— thus our costs still grow exponentially.

First, results under this characterization are *implementation independent.* Different computer languages (and encoding schemes for the parameters) may alter the time complexity of an algorithm, but no "reasonable"[189] language will be convert a

---

[189] All known, physically constructable, computer architectures are *"reasonable"* in this sense, and it is believed that all possible computers based on classical physical principles preserve this property (Papadimitriou 1994). However, there is some debate over whether this implementation independence applies to hypothetical computers designed to utilize unexplored properties of quantum physics. Two authors, in particular, assert that the above model does not accurately describe all potential problem solving devices.

Deutsch asserts that under the "many-universes" interpretation of quantum theory, one could design a device to exploit an infinite number of alternative universes for parallel calculation (Deutsch 1985). Under this controversial interpretation of quantum theory, devices may be built which would be able to compute some (but not all) problems in polynomial time that are computable on all conventional computers only in exponential time (Deutsch and Jozsa 1992).

Penrose (1989; 1994) makes a somewhat different argument, asserting that currently unresolved areas of quantum physics may provide fundamentally different ways of solving problem than is represented by the Turing model. Penrose argument, which is too rich and polynomial algorithm to an exponential algorithm. While a more powerful computer may be able to perform each atomic operation more quickly, it will not alter the time complexity function of the problem. Intractable problems cannot be made tractable through improvements in hardware technology.

Second, results under this characterization apply to the problem itself, not to a particular method used to solve this problem. Problems which are shown to be difficult under this characterization are difficult for *any* possible computer method. Since it is the problem, itself that requires exponential time, these problems cannot be made tractable through advances in software or algorithmic design.

This characterization is also subject to several important limitations. These limitations have caused its use as an absolute measure of problem difficulty, especially for social science problems, to be justly criticized (Page 1994). I will briefly summarize these limitations here, and in Section 5.5 I will extend the analysis to address the relevance of these limitations for the redistricting problem.

First, the distinction between tractable and intractable problems is most important for instances of large size - where the exponential factors in the time requirements of these problems become dominant. Consider the following two problems. Problem "A" is

---

detailed to be adequately summarized here, asserts not only that quantum physics allows mechanisms for problem solving which are fundamentally different from those used in today's computers, but that the human brain actually employs such mechanisms.

computationally intractable and takes $O(1.1^n)$ steps to solve. Problem "B" is computationally tractable and takes $O(n^{14})$ steps. Although the time needed to solve problem A will eventually become much greater than the time required for problem B, for problem sizes less than one thousand, we can actually solve problem A much more quickly.

Second, when we use this characterization we require that problems be solved exactly. Some problems that are computationally difficult to solve may be *approximated* much more quickly. If the approximation reached is (provably or empirically) close enough to the optimal solution to the problem, for practical purposes we may not need to find the exact best solution.

Third, when we use this characterization we require that our algorithms *always* reach a *correct* solution for every problem instance, requirements that make computational complexity a function of the worst-case problem instance. Since we base our analysis on the worst-case, we may overstate the complexity of the problem on average. Furthermore, since we require that our solution-algorithm neither make errors, nor give up on a problem, we will drop from our analysis some algorithms that are probabilistic. While such algorithms do not formally "solve" a computationally hard problem, they may be quite useful if their rates of error and of failure are sufficiently low.

The three caveats above offer to us possible escape routes around computationally intractable problems, but these are only *possible* routes. And as I will show in Section

5.5, in general the requirements of automated redistricting procedures make these avenues unlikely to be fruitful.

5.3.4. Redistricting is a Computationally Hard Problem

In the previous sections, I showed how redistricting is deeply connected to mathematical partitioning problems. Many researchers in computer science have examined partition problems and reached some conclusions about their computational complexity. In this section, I show that the redistricting problem in general, and even many simpler redistricting sub-problems are likely to be intractable.

Proving that a problem is intractable is difficult — researchers have been unable to determine whether most problems are tractable (Papadimitriou 1994). There are, however, a number of large classes of problems that computer scientists believe to be intractable. The oldest of these is called the class of *NP-complete* problems.

Cook defined the first NP-complete problem (Cook 1971), which has now been shown to belong to a large set, consisting of hundreds of problems in many fields. Karp characterized the most important property of NP-complete problems (Karp 1972): *polynomial-time reducibility*. Any NP-complete problem can be transformed into any other NP-complete problem in polynomial time.[190] Thus, if you could prove that any NP-

[190] Polynomial reductions are defined so as to preserve space complexity characteristics as well. Furthermore, there is an even deeper equivalence between all

complete problem is formally intractable, you would have proved all such problems intractable, and vice-versa.

Search for a proof of the intractability of NP-complete problems has been of the most famous open problems in computer science for over two decades. While no proof of intractability has been found, no polynomial algorithms have ever been found that solve any of these problems, and because of the breadth of the class of problems, it is widely believed that no such algorithms exist.

The class of NP-complete problems is not the only class that is believed to be intractable — there are many other classes of problems that are equivalent to each other but not to problems in the NP-complete class. For our purposes, however, we need consider only the NP-complete class and the related class of *NP-hard* problems: The *NP-hard* class is a superset containing the *NP-complete* class; this class is potentially harder to solve than NP-complete problems because although if any NP-complete problem is intractable, then all NP-hard problems are intractable, the reverse is not true.[191] The diagram below illustrates the probable relationship between the NP-complete, NP-hard and tractable classes of problems. (Figure 5-1)

---

known NP-complete problems — each problem can be transformed to any other by a simple functional mapping (technically a “bijection”).

[191] Any NP-Hard problem can be shown to be NP-complete for at least some instances, but not necessarily for all instances.

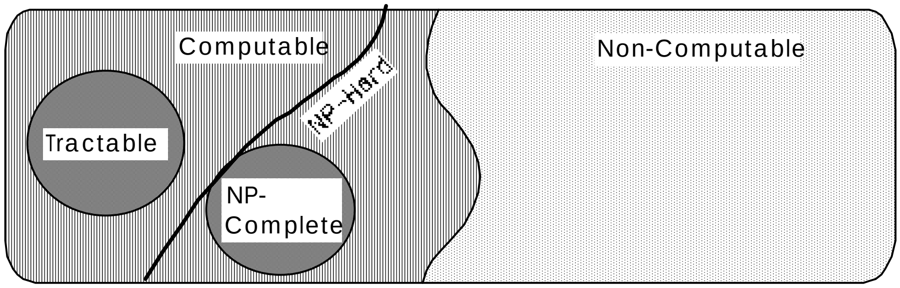


Figure 5-1. A diagram showing NP-Complete and Related Complexity Classes

In the appendix to this chapter, I show formally how the most common redistricting sub-problems, such as finding the optimal set of compact districts, are NP-complete or NP-hard.[192] Table 5-3 summarizes these formal results; for each sub-problem I give a short example of its formal characterization, and a reference to the corresponding problem in complexity theory.

---

[192] This finding contradicts Garfinkels's and Nemhauser's (1970) (Garfinkel and Nemhauser 1970) prediction that the time required for their integer programming technique should be expected to be approximately linear in population units.

Redistricting Sub-Problem	*Example Characterization*	*Reference Problems*
Equal population districts	*set partition*: divide members of set into subsets minimizing value function • population unit: represented by an integer-valued set member • district: each subset in partition is a district • value function: the largest difference in population between any two districts	*Weighted Set Partition* (Karp 1972) *3-Partition* (Garey and Johnson 1983) *Weak Partition* (Johnson 1982) *Integer Programming* (Garey and Johnson 1983)
Compact districts	*set partition:* as above • population unit: represented by a vector valued set member • district: as above • value function: (a) maximum distance between points in a district (b) total perimeter of districts	*Distance-d Partition of Points in the Plane* (Johnson 1982) *Minimum Perimeter Partition into Rectangles* (Johnson 1982)
Competitive districts	*set partition:* as above • population unit: represented by a weighted set member. Weight is Republican registration - Democratic registration in each districts. • district: as above, district weight = sum of weights of the population units it contains • value function: minimize sum of squared district weight	*Minimum Sum of Squares* (Garey and Johnson 1983)
Contiguous districts (with population equality)	*graph partition*: divide nodes of a graph into connected subsets minimizing value function • population unit: represented by a node of the graph • contiguity relationship: population units contiguous to each other are connected by edges in graph • district: as above • value function: the largest difference in population between any two districts	*Cut into Connected Components of Bounded Size* (Johnson 1982; Johnson 1984)

Table 5-3. Redistricting Problems that are at least NP-Complete

While I refer to a number of particular characterizations in Table 5-3, it is important to realize that none of my results in this chapter is dependent on the use of a particular characterization.[193] Since all of these characterizations are in the set of NP-complete problems, they are all formally equivalent — any algorithm that solves one problem can be simply reformulated to solve any other without a change in complexity class. The advantage of multiple characterizations is that each characterization may suggest various limitations that can be put on the problem that may make the problem *easier* to solve or approximate. (See Section 5.4 below.)

While I have focused on sub-problems, you should note, as well, that these complexity results apply to the general redistricting problem. Since complexity results describe worst case properties, my results demonstrate that the redistricting problem, *in its most general form,* is at least as difficult as any NP-complete problem.[194] A further

---

[193] Reformulating the redistricting problem in other ways, can be useful if it suggest natural restriction that can be put on the problem in order to simplify it. Restrictions that re natural in one context, such as planarity or graphs in the graph-partitioning problem, may not be obvious when the problem is formulated as an integer partition problem.

[194] Each of the redistricting sub-problems above is an instance of the redistricting problem. If some instances of the sub-problems require exponential time, then the general problem as well must also require this amount of time. Hence, the redistricting problem has a time complexity at least as large as its sub-problems. We have not shown that these sub-problems represent worst cases of the redistricting problem, however, so it is possible that some instances of the general redistricting problem may be worse than NP-complete,

implication of these results is that finding an optimal district under any *combination* of these "hard" goals is also hard. In sum, we should not expect automated optimal redistricting to be tractable for an arbitrary choice of population units, number of districts, and an arbitrary value function.[195]

---

or even unsolvable.

[195] When making the claim that automated redistricting is formally intractable, it is sometime objected that this claim cannot be true because political actors are able to gerrymander so well. If political actors can gerrymander optimally, the argument guess, then so should computers. I believe this objection is based on three false assumptions:.

• Claim one: *Humans already perform (near) optimal gerrymandering*. There is little evidence for this claim; in fact, there is much evidence to the contrary, and many examples of attempted gerrymanders that have had far from the intended results.

• Claim two: *Automated redistricting is no more difficult than gerrymandering.* Again, the available evidence seems to point toward the opposite conclusion. Automated redistricting is significantly more complicated than gerrymandering in three important respects. Gerrymandering usually involves the maximization of one simple goal. Optimal redistricting may involve many simultaneous, complicated, and conflicting goals. Gerrymandering is often limited to relatively small modifications of an existing plan. Automated redistricting processes must examine a much wider range of possible plans. Finally, gerrymanders need not be optimal to be politically effective. The social value

It is also important to note that the redistricting goals I have listed in Table 3 are common, straightforward, and relatively simple. For example, of all the redistricting goals specified in the literature, population equality is probably the most straightforward to quantify, and the easiest to evaluate.[196] Yet even the task of drawing a district plan to minimize population deviation alone is computationally hard.

---

function being optimized by an automated process may be much more sensitive to suboptimality.

• Claim three: *If human's can perform a (mathematical) task well, computers can (at least theoretically) perform the same task as well.* The lack of success in the field of artificial intelligence is evidence to the contrary. See Dreyfus 1992 (Dreyfus 1992) for a review of the successes and failures in this field, and a more detailed argument against claim three. Even advocates of artificial intelligence recognize that such tasks as recognizing human social and political relationships, and applying "common sense," are among the hardest problems for computers. (See Chapter 3 of Taylor 1989.)

[196] Even this goal can be defined in a number of ways, depending on how you wish to measure the inequality between two districts. For instance, we might focus on the maximum differences between the largest and smallest districts, or their ratios, or their average differences, etc.

Still, this does not demonstrate that the redistricting problem is hopeless. Section 5.5 explores the avenues available for managing intractable problems and the promise of these avenues for redistricting.

5.4. Attempts to Escape the Intractability of Automated Redistricting

In Section 5.4 of this chapter I demonstrate that the redistricting problem is NP-complete, and we briefly reviewed several caveats to our definition of intractability. Do these caveats allow automated redistricting to escape intractability?

Recall the definition of "tractability": A problem is tractable only if a polynomial-time algorithm exists which is *guaranteed* to *exactly* solve *all* instances of this problem. Are we being overly demanding? If we demand somewhat less of our solution algorithms, can we find practical solutions to automated redistricting problems?

NP-completeness is a limited form of intractability,[197] and there are a number of possible avenues for dealing with these types of problems. In Section 5.4 we did not

---

[197] NP-complete problems certainly do not encompass all intractable problems, nor are they the "worst" class of intractable problems. The intractability of NP-complete problems takes a limited form.

A problem may be intractable for a number of different reasons. I list three here, in order to illustrate several levels of problem difficulty. First it may be that the solution itself demonstrate that *all* possible redistricting sub-problems are intractable: If we restrict our redistricting values sufficiently, we can certainly find a tractable way of finding optimal

---

is unmanageably large or otherwise unmanageable. Second, it may be that the solution in itself is manageable, but is difficult to both to solve and verify. Third, in the least difficult case, only finding the solution is difficult — once found it can be easily verified and put to use.

An example of a problem of the first type, where the solution itself is unmanageable is: "enumerate the set of all possible district plans." Simply printing this solution for this type of problem is untenable for all but the most minuscule instances. This type of problem is provably intractable. Fortunately, problems of this type are in some ways uninteresting, in that, even should a solution for the problem eventually be obtained, one would be unlikely to be able to make use of it.

Problems of the second type are esoteric and too awkward to describe here — see Bellare and Goldwasser (1994) for a description.

NP-complete problems fall into the third category. An example of a problem in the third type is: "Does there exist a district plan where the maximum population difference between districts is less that B, and if so, specify that plan?" For this problem, as long as computation of the value function f is tractable, finding a satisfactory plan may be difficult — but given a satisfactory plan, it is easy to verify that the value of the plan exceeds B.

districts under them. Similarly, if we restrict the set of inputs to our problem enough, we may find it quite tractable.

Furthermore, we did not in eliminate all *practical* approaches to general redistricting problems. If the problem cannot be made tractable through restricting our values or data, we might be able to find a probabilistic method that solves the problem most of the time, or a deterministic algorithm that works well on most cases. Alternatively, we may be able to find a method that quickly finds an approximately-optimal solution.

We must, however, be cautious; although these escape routes are open in theory, they pose a number of technical and political difficulties. In the remainder of this section, I will explore these possible escape routes.

5.4.1. Solving NP-Complete Problems — An Example

While all NP-complete problems are reducible to each other, in a practical sense they are not all equally hard to solve: Some problems can be easily restricted, answered probabilistically, or closely approximated.

We can use a simple example to show the practical difference between two formally intractable problems. For this example, imagine that a rich acquaintance has died and that you have been asked to be executor of the estate. The rich acquaintance was a collector of art and antiques, and the estate is made up entirely of unique and valuable items. You are told that you must divide the estate between its inheritors in such a way that (1) you give away all the items (you cannot sell them and give away the money), (2) each item goes to a single inheritor, and (3) you must maximize the *subjectively equality* of the division — the monetary value each person assigns to her own share must be as close as possible to the value another person assigns to his own share.

Formally this is an NP-complete problem,[198] but if all inheritors share common values for each object (perhaps they are all antique dealers who know current market prices), this problem is not very "hard" for many practical cases: If there are only two inheritors, and no item is "priceless," we can find the optimal solution in polynomial time using dynamic programming (Garey and Johnson 1983). On the other hand, if there are more inheritors, but all the objects are of approximately equal value, we can easily minimize the inequality between shares.[199]

Now suppose that each inheritor has differing private values[200] for sets of objects; it is especially difficult when the relevant sets differ for each inheritor as well. (For

---

[198] It is an example of the weighted set partition problem (Garey and Johnson 1983).

[199] Even if the antiques differ greatly in value, we may be able to get relatively close to an even division. While I can find no approximability results for this partition problem, the following algorithm is quite successful for bin packing, a similar problem: Simply order all the objects from least to most valuable, then assign each object in this list to each inheritor in turn, until the list is exhausted (Garey and Johnson 1983).

[200] Here I disregard the additional difficult of eliciting the inheritor's private values for each set of items — the problem is difficult even if you are completely informed about

example, aunt Martha attaches great sentimental value to the sofa and end-table combination, whereas uncle Henry hates the end-table but has always coveted the sofa and matching wall-hanging.) If you have a lot of antiques to distribute, although the problem remains much the same in structure to our first problem, it has become much more difficult to deal with practically.

5.4.2. Problem Size and Computational Complexity

Even problems which are computationally hard may be solved easily for sufficiently small cases. Are the sizes of the redistricting problem which are typically encountered small enough that intractability is not a practical issue?

Unfortunately, the answer to this question is probably no. As Section 3 shows, the number of possible solutions to a problem grows both as a function of the number of districts being drawn (to a limited extent) and as a function of the number of population blocks that we use. As the appendix shows, the time necessary to solve redistricting sub-problems grows exponentially as a factor of both districts and population blocks.

While the number of districts is typically reasonably small, ranging from one to approximately fifty districts in the United States, the number of population blocks in practical cases is quite large — ranging from several thousand to approximately one hundred thousand. Although we can only approximate the time needed to solve these redistricting problems, the number of solutions that must be searched is large enough that

---

each inheritor's basic private values.

we can reasonably expect the exponential time-requirements of the algorithm to be dominant, even if the exponential growth is relatively small.[201]

5.4.3. Restricting the Redistricting Problem

It is well understood that any problem, when sufficiently generalized, becomes computationally hard, and conversely, any sufficiently restricted problem becomes trivial. Because of this common-sense principle, we must consider whether there are any natural restrictions that we can place on the redistricting problem that would make it tractable.[202]

There are three basic types of restrictions that we can place on the redistricting problem: First, we can restrict the redistricting goals that we will consider. Second, we can rule out some redistricting plans, such as those containing noncontiguous districts, as

---

[201] An exponential growth factor as low as 1.001^n is likely to make such large problems intractable.

[202] Note that here I use *restriction* to refer to how we limit the goals and inputs that we will consider for redistricting problem itself. These restrictions do not necessarily correspond to *constraints* on districting that limit electoral manipulation (i.e., “gerrymandering”). For example, restricting the automated redistricting problems to consider only county-level population data, rather than finer-grained population data, might actually make electoral manipulation easier.

illegal — we can throw them out of the analysis. Third, we can restrict the data, or population units, that we feed into our automated redistricting process.

For purposes of clarity, we will discuss the various types of restrictions separately. In practice, restrictions of all types are often combined to simplify problems and avoid computational complexity. But although we can reduce the practical difficulties of finding a plan by combining multiple restrictions, we cannot use these combinations to eliminate the political and normative problems we have already encountered.

Restrictions on Value Functions

We may make the computer's task easier by limiting the number and types of goal function that it has to optimize. By using such restrictions, we may be able to take advantage of specially tailored optimization algorithms, or we may be able to eliminate exponential time requirements for more general optimization algorithms.

While there must exist goal functions that are "easy" to optimize,[203] there are two difficulties with this approach. The first difficulty is a purely practical one — reasonable candidates for redistricting goals seem to be computationally difficult to optimize. As we saw in Section 4, even the simplest and most popular value functions are *all*

---

[203] For example, a function that assigns the same constant value, although completely uninteresting, is easy to optimize. Unfortunately, there seem to be no interesting value functions that are "easy" to optimize on unrestricted inputs.

computationally hard to optimize. Furthermore, any (simply weighted) combination of these functions with each other, or with any other function, will also be computationally hard. In sum, our present goals do not lend themselves to easy computation.

The second difficulty is normative and political. Proponents of automated redistricting argue that one of the strengths of automation is that it allows goals to be decided by a political process — automation should only affect the implementation of goals, not their choice. This separation of goal choice and implementation is violated by placing restrictions on allowable redistricting goals. Allowing the limitations of a computerized process to restrict, *ex ante*, the type of representational goals that society is allowed to pursue may be normatively and politically unacceptable.

Restrictions on Input

Restrictions on input can also, theoretically, allow us to escape intractability. In this chapter, as in most automated redistricting research, the basic inputs have been defined in terms of vector-valued *population blocks* such as census blocks. We can place restriction on census blocks directly or indirectly. Restrictions on these population blocks can be used directly and indirectly for this purpose.

We can use limitations on input *directly* to reduce the computational complexity of a problem by eliminating its worst cases — if we can predict what cases create problems for our redistricting algorithms. For example, if we want to create district plans solely to minimize population inequalities, we can simplify this problem by restricting all population blocks to be equal in size.[204] In this restricted case, any of the heuristics listed in Section 5.2 will quickly find an optimal solution to this problem. While not completely trivializing the problem, this eliminates the possibility of cases which might cause the algorithm to take exponential time to complete.

For more complicated problems, we can use input restricting indirectly, and in combination with restrictions on redistricting goals, to reduce the number of solutions that an optimization algorithm needs to consider. Methods such as "branch and bound" and "cutting planes methods" use different applications of this general principle (Balas and Toth 1985; Reinelt 1991). Balas gives a succinct description of these methods:

> "Enumerative (branch and bound, implicit enumeration) methods solve a discrete optimization problem by breaking up its feasible set into successively smaller subsets, calculating bounds on the objective

---

[204] A similar alternate strategy for drawing equal population districts is to arbitrarily split population-blocks in order to make them approximately equal. This method seems to be one of the most popular in the political arena, where it is not uncommon to find districts that are based on census blocks or tracts, but which splits these in order to maintain equality between districts. Since in this chapter "population unit" has been assumed to reflect the finest grained unit for which census information is available - splitting population-blocks must be regarded as an approximation method. The limits of approximations methods are discussed in Section 5.2 below.

> function value over each subset, and using them to discard certain subsets from further consideration. The bounds are obtained by replacing the problem over a given subset, with an easier (relaxed) problem, such that the solution value of the latter bounds that of the former. The procedure ends when each subset has either been reduced to a feasible solution, or has been shown to contain no better solution than the one already in hand. The best solution found during he procedure is a global optimum."

These methods are usually very sensitive to our choice of goal functions — we must know quite a lot about the goal function to derive the "relaxed" problem that we use to set bounds.[205] We must then restrict our input set if we want to guarantee that the branch-and-bound procedure does not itself take exponential time.

---

[205] There are many general sorts of "relaxations" that we can apply to redistricting. For example, if we can formulate our redistricting problem as in integer-linear program, we might pretend that census blocks are infinitely divisible — converting this into a linear program that is simple to solve. From the solution to the linear program we might be able to compute bounds on our linear program. In general, however, choosing a successful relaxation of a problem requires mathematical intuition and cleverness.

Although we can make the automated redistricting process easier through restricting inputs, many useful restrictions may not be practical or possible. To eliminate the worst case behavior of these programs, we must enforce regularity upon the population blocks that we use, but many of these characteristics are exogenous. For example, geographical regularities can help us to draw compact districts. If your state is a perfect square with people uniformly distributed across it, then drawing equal-sized compact districts is simple. We might even require that all districts be equal-sized rectangles. Unfortunately, given the irregular boundaries and irregular distributions of population in many states, it is impossible to draw such regular districts, and finding the best set of districts is much harder. In sum, problems may be serendipitously easy, but we cannot rely on good luck.

Furthermore, we can only rely upon input restriction to eliminate worst-case behavior when we also carefully restrict our value functions. Restrictions that make the satisfaction of one goal easier may exacerbate the difficulties involved in satisfying other goals. There are few restrictions that have been shown to be useful across a variety of value-functions.[206] We may not be able to politically or normatively justify such restrictions.

---

[206] There is a notable exception. First, restricting the maximum value (population, size, etc.) that any one population block can be assigned to can sometimes allow problems to be solved in polynomial time. NP-complete problems that have been shown easily

Restrictions on Plans

Instead of trying to restrict goals or data, we might, more reasonably, restrict the types of districts or plans we will consider. For example, if we only consider districts that are contiguous rectangles, if our state has no "holes" (e.g., lakes are not counted as part of the district), and we are interested only in maximizing one form of compactness (in this case, perimeter-minimizing), we can more easily find "optimal districts" (Johnson 1982). Most political planners, however, would consider such restrictions unrealistic and overly restrictive.

We should also remember that while stating restrictions in this way is mathematically convenient, politically and normatively such restrictions imply a set of restrictions on value functions or inputs. For example, if we require that our plans be

---

solvable under this restriction are referred to as "pseudopolynomial." NP-complete problems that have been shown to remain intractable under this restriction are termed "strongly NP-complete."

A number of redistricting sub-problems, when further restricted to draw only *two* districts are "pseudopolynomial" in this sense. That is, if we are trying to divide an area into exactly two districts, and we can set an upper bound on the value of each population block, we can compute the optimal plan in polynomial time for many, but not all, sub-problems. These same problems are strongly NP-complete for more than two districts. See Appendix A for more details.

contiguous we exclude any value functions that allows contiguity to be weighed against other goals. Because of this equivalence, we cannot escape the limitations of value and input restrictions by reformulating them as plan-restrictions. Although a plan restrictions may be valuable for formulating issues clearly, we must be careful that these are not used by political manipulators to obscure the other types of restrictions that they imply.

5.4.4. Can Sub-Optimal Redistricting Help?

When we defined theoretical computational complexity, our solution concept was very demanding. Political solutions, however, seldom require the precision that we demand of theory. If we relax our requirements for precision, can we find general, easy, practical methods of redistricting? Can we find optimal plans most of the time, or find approximations that are "close enough"?

Optimal Redistricting — Most of the Time

There are two possible ways in which we could quickly perform optimal redistricting "most" of the time: We might develop a method to get close to optimum for all cases, or to find the optimum for most cases.

Remember that when we defined computational-complexity, we required that our algorithms be guaranteed to generate a *correct* solution to a problem. We might relax this guarantee of correctness, and allow our methods to make occasional errors.[207] The

---

[207] Technically, we expand our solutions to algorithms which produce the result with

possibility of "solving" some NP-complete problems has not been theoretically foreclosed. Unfortunately, it is conjectured that NP-complete problems are not susceptible to this type of solution[208] (Johnson 1984; Papadimitriou 1994).

Instead of relying on probabilistic methods, we might hope for an easy draw of problems — it might be that an algorithm solves the redistricting problem with certainty either on average or in practice.[209] As long as there exists some case for which

---

probability ε — we are not here bounding the magnitude of the error, only the probability of its occurrence.

[208] An algorithm is said to be "boundedly probabilistic polynomial" (BPP) if it computes a "solution" to all instance of a problem in polynomial time, and the "solution" computed has a probability strictly greater than 50% of being correct, each time the algorithm is executed. Because the probability of error on *each* execution of the algorithm is independent, algorithms in this class can be executed repeatedly to obtain any desired probability of correctness. The conjecture referred to above is that no NP-complete problem is a member of the class BPP.

[209] Technically, we might say that an algorithm works well *on average* when we describe its behavior upon problems drawn from a theoretical, mathematically-defined probability distribution; whereas, we would claim an algorithm works well *in practice* when we have evaluated it against real empirical data.

exponential time is required, the problem is still, by definition, computationally intractable, but it may not be quite easy most of the time.

This is primarily an empirical question: How well do real methods work on real data? As Section 2 showed, however, the track-record of automated redistricting is not very good. Researchers' previous attempts to generate optimal districting plans suggest that this problem can be quite difficult "in practice," and the literature reveals no procedure that is both demonstrably optimal and that has been generally effective on data sets large enough to be useful for political redistricting.[210] Yet, the question of whether redistricting sub-problems can be solved in practice remains open. Analysis of average-case complexity requires, in most cases, that we specify a particular distribution function. My literature search reveals no specific average-case complexity results concerning the redistricting sub-problems discussed above for any distributions.[211] Because of these, A formal analysis of average case properties of the redistricting problem is beyond the scope of this chapter.

Finally, a recent theoretical result should give us pause before we place our hopes on the average case: It has been proved that for the class of simple computable distributions,

---

[210] See Gudgin and Taylor (1979) for a discussion of some early attempts.

[211]It is not obvious what mathematical function can be used to accurately describe the distribution of values over population blocks (though see Gudgin and Taylor (1979)).

the average-case complexity of all NP-complete problems is exponential (Li and Vitanyi 1992). For some distributions, average-case complexity cannot be an escape from intractability.

As the previous section shows, the prospects of discovering a tractable procedure for optimal redistricting are dim. In practice, automated procedures will almost certainly result in sub-optimal redistricting. This section discusses both practical and political implications concerns surrounding approximation methods.

Guaranteed Approximations

Computational complexity is a measure of the difficulty of obtaining *optimal* solutions, but it says little about the difficulty of approximation. In fact, NP-complete problems, while equivalent for optimal solutions, are not equivalent for approximation. Some problems are much easier to approximate than others. For some NP-complete optimization problems there are methods that will, in polynomial time, generate a solution that is *guaranteed* to be within a particular percentage of the optimal value.[212]

---

[212] One must be careful to distinguish between measurements of approximation based on the value of the solutions produced versus measures based on percentage of solutions which are excluded. For example, simple hand drawn districts are likely to be in the top 99% of all the possible districts - simply because there are so many possible districts with little value. However, these simply drawn districts may be much lower in value than the

Methods that obtain arbitrarily close "solutions" in polynomial time are known as *fully polynomial approximations.* Unfortunately, it can be proved that no fully polynomial approximations exist for many of the redistricting sub-problems discussed above.[213]

Although the redistricting problem does not allow arbitrarily close approximations, we have not excluded the possibility of approximating a solution to within some fixed percentage of the optimum. No such guaranteed approximation procedure for a redistricting sub-problem has yet been demonstrated, but the question remains open.[214]

Making an Educated Guess

---

optimal district.

[213] Epsilon approximations are methods that guarantee that the ratio of the value of the approximal solution to the value of the true optimum is no less than 1-ε. Fully polynomial approximations are epsilon approximations which with time requirements bounded by a polynomial function of ε, for all ε.

It has been shown that no fully polynomial approximations exist for problems which are *strongly* NP-complete (Papadimitriou 1994). As Appendix 1 shows, a number of redistricting sub-problems are at least strongly NP-complete.

[214] I conjecture that good guaranteed approximation limits can be obtained for the problem of minimizing population inequalities between districts.

By far the most common way of approaching automated redistricting is through the use of *heuristics*. A heuristic is any methods that is considered useful for problem solving, but for which no guarantees of optimality (or approximate optimality) apply. While heuristic methods give no guarantee of approximate optimality, they can reduce the set of plans that need to be considered.

As I noted in Section 2, all of the procedures that researchers have used for automated redistricting are in fact heuristics. The computer science and operations research literature also offers several heuristics that have been useful in solving mathematically similar (partition) problems:

- Simulated annealing, a process that is analogous to the slow cooling in metals, has been suggested for use in the redistricting problem (Browdy 1990b). Zissimopoulos (1991) uses one variation of this technique on a number of set-partition problems.

- Genetic algorithms use processes analogous to crossover, mutation, and evolutionary selection to generate solutions to optimization problems. Chandrasekharam(1993) examines the performance of some variants of genetic algorithms on selected graph-partitioning problems.

- (Hershberger 1991) develops a polynomial time algorithm that divides a polygon into two optimally compact subregions. He suggests a divide-and-conquer heuristic for the more general polygon-partitioning problem, where his algorithm would be used to repeatedly subdivide a polygon.

•(El-Farzi and Mitra 1992) transforms set-partition problems into integer-programming problems *Lagrangean relaxation* and other assignment relaxation heuristics as part of a *branch-and-bound* solution method.

The heuristics above have been shown to be useful in several sets of test cases and merit further investigation for redistricting. However, the size of the test problems in the papers above corresponded approximately to redistricting problems on the order of 10 districts and 100 population units — much smaller than the typical, practical, redistricting problem. Furthermore, for at least one common sub-problem, drawing compact districts according to one common definition of compactness, many of these methods have performed poorly. (See Chapter 2 and Chapter 4.)

Political Implications of Sub-Optimal Redistricting

While sub-optimal redistricting methods offer the only practical means of implementing automated redistricting, these methods have a number of disturbing normative and political implications.

In his argument for automated redistricting, Isaccharoff writes (emphasis added):

> If reviewing courts use the absence of a *verifiable computer algorithm* or some other clear ex-ante articulation of the bases for reapportionment decisions as presumptive evidence of constitutional infirmity, the logic of adjudication may - as with the development of fixed equipopulation rule after *Reynolds* - propel the states toward the use of verifiable criteria for reapportionment (Issacharoff 1993).

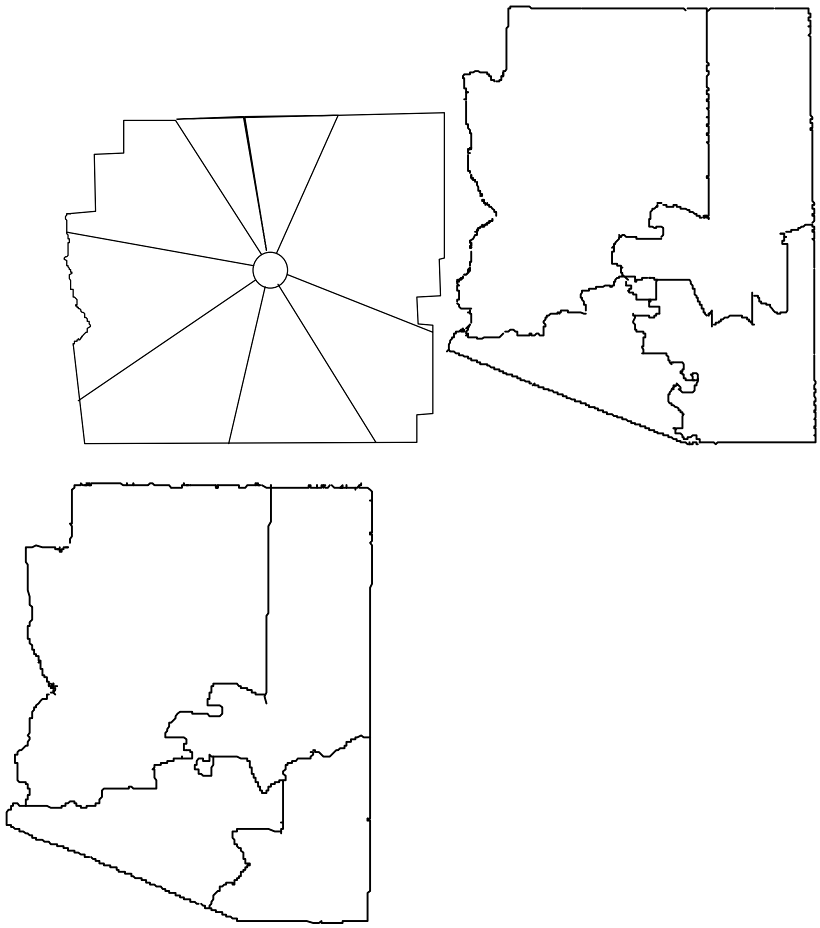
All current methods for automated redistricting are heuristic—they offer no guarantee that a solution is close to the optimum. In practice, we can attempt to gauge the effectiveness of a heuristic process thorough experimentation and simulation, but such analyses will depend on the value functions we use. If we cannot place guarantee their performance, automated redistricting methods lose much of their normative appeal.

Furthermore, judicial review becomes more difficult, as the quality of the plans that we generated becomes a very technical question, dependent both on the choice of value functions used and heuristic methods. It will be difficult to determine whether a method is biased, and also difficult to determine whether a method is *narrowly tailored* to pursue legitimate goals. If, for example, creation of minority-opportunity districts is one of our goals, how will we determine that it has not been weighed too heavily by the redistricting algorithm?

Second, guarantees of performance notwithstanding, heuristics and approximation techniques select from a large body of solutions, and may be biased in their selection. Even if we select only among methods which arrive at solutions in the top 99.99% of all possible district plans, the pool of possible solutions is staggeringly large, and the actual values of such solutions may differ dramatically. Solution methods can be biased toward certain classes of solutions.

Heuristics may leave plenty of room for manipulation. The *Ohio State University* (OSU) is a particularly clear example of a method that both embodies a particular bias and that can be manipulated for a variety of goals. The OSU method draws plans that are arbitrary in their particulars but follow the same general plan: they make one large central district, and then other districts that radiate out from this central district (Hale 1966) (Figure 1a). In contrast to the OSU method's ideal, Arizona's 1980 redistricting plan used the same type of arrangement for a racial gerrymander (Figure 1b). The final court approved plan is also based upon the same general lines (Figure 1c) (Congressional Quarterly Staff 1983; Light 1982).[215]

[215] I have removed all but the state boundary and district lines from the maps in Figure 1 in order to maximize the clarity of the illustration.



A *B* *C*

Figure 1: Will The Real Gerrymander Please Stand Up?

Bias can also enter the process in the form of initial conditions from which these heuristics start. Since most redistricting heuristics are *path-dependent*, the starting conditions for these methods can influence the types of plans that they generate. For example, the two most popular automated redistricting algorithms, the Weaver-Hess procedure, and the Nagel method, both described in Section 2, are sensitive to initial conditions. As Weaver writes (Weaver 1970):

> "CROND's districting technique does work by trial and error...Different starting points, whether by hand or mechanical techniques discussed below, will usually produce different sequences and different sets of districts."

Unless all district plans near the optimum are very similar in the *types* of plans they produce, how we choose starting points matters. These choices offer opportunities for hidden political manipulation of the outcomes. While individuals may not be able to use such manipulation to choose a precise plan, interest groups may be able to use methods that are biased toward those groups. Perversely, this could make partisan gerrymandering easier, because it would limit the gains that incumbents could make on their own — giving incumbents in the same party a stronger incentive to cooperate.

The intertwining effects of value functions, methods, and starting points weaken the distinction between intent and effect that the automation procedure is supposed to make clear. The veil of automation has been pierced. Judicial review of such intertwined and technical procedures may be significantly more difficult than judicial review of the individual districting plans. Furthermore, political groups with large computing resources will be much better equipped to discover the combinations of methods and functions that will work to their advantage.

If manipulation cannot be eliminated by automation, perhaps the opportunities to choose plans can be opened up equally to all parties. Why not make computer resources available to all parties, let interested parties submit plans, and select the plan with the highest value?

This scheme has two shortcomings. First, the incentives remain for strategic manipulation at the value-choice stage. Automation cannot be used as a “veil of ignorance” to force participants to bargain fairly. Second, this scheme discourages compromise. Since under this scheme the plan with the highest value function acts as a “trump” and is implemented *in toto*, there is a strong incentive for secrecy. If an opponent knows the details of your proposed plan, he will be able to know whether more work needs to be done on his plan. In addition, he may be able to use iterative improvements to submit a slightly higher valued plan that better suits his goals.

5.5. Automating the Redistricting Process Limits Conflicts with Representational Goals

Even if we manage to overcome the technical difficulties involved in finding optimal districts, we will still need to formalize the notion of an “optimal district.” In attempting this, we are bound to encounter two political and normative problems. First, technical details of our formulation may be used to manipulate the redistricting process — we will be able to manipulate by defining our goals in particular ways. Although automated redistricting may limit incumbent gerrymandering, it is likely to encourage partisan and interest-group gerrymandering. Second, the very act of formalizing our goals in a form suitable for computer processing will eliminate from consideration a number of important representational goals.

5.5.1. Limit on the Formalization of Redistricting Goals

To automate the redistricting process, the goals we use must first be quantified — the extent to which a plan meets (or does not meet) each of the goals must be representable by a number.[216] For example, population inequality might be quantified by subtracting the largest and smallest districts.

Not all goals may be easily quantified, however. In particular, many goals are *representational* – they refer not to a directly measurable aspect of a district, such as its population or shape, but to the role that district plays in creating political fairness. As Justice White wrote for the court in *Gaffney* v. *Cummings* (1973):

> The very essence of districting is to produce a different — a more "politically fair" — result than would be reached with elections at large, in which the winning party would take 100% of the legislative seats.

Representational goals, such as protection of communities of interest and nondilution of minority representation, are especially difficult to describe formally.

The Court has recognized a wide variety of valid goals for redistricting in a number of cases. In *Karcher vs. Daggett* (1983), the court's opinion stated that "any number of

---

[216] In theory this number might represent either an ordinal or cardinal ranking. In practice, all of the automated procedures described in the literature have used cardinal rankings.

consistently applied legislative policies" might be important enough to justify deviations from absolute population equality, including compactness, protection of municipal boundaries, preserving the core of previous districts, and protecting incumbents. In *Davis* v. *Bandemer* (1986), the court determined that "vote dilution" is a relevant and justiciable factor in redistricting. Most recently, in *Miller* v. *Johnson* (1995), the court chastised the defendents for ignoring "traditional" principles, including (but not limited to) contiguity, compactness, "respect for political subdivisions or communities defined by actual shared interests," and race neutrality[217]. Could we program a computer to evaluate and balance all of these factors?

---

[217] The complexity and subtlety of representational goals is also illustrated by the court's commentary on multi-member districts in *White* v. *Regester* (1973), and by the similar commentary found in the Senate report on 1982 amendments to Section 2 of the Voting Right's Act.

The high court, in *White* v. *Regester* recognized that the "totality of the circumstances" must be used in the evaluation process. While the court listed a number of important circumstantial factors including racial motivation for a plan, history of discrimination and voting patterns, and the number of majority-minority districts, it stressed that decisions were not limited to these factors.

After listing the seven well known "totality of the circumstances" factors, the Senate committee states:

To evaluate plans with respect to representational goals, we need to be able to recognize natural geographic, social, and historical patterns, and to use extensive knowledge about political and social relationships.[218] It is exactly such patterns that computers are least able to recognize, especially when these patterns are dynamic, and cannot be simply "hard-wired." In fact, humans perform better than any computer

---

> While these enumerated factors will often be the most relevant ones, in some cases other factors will be indicative of the alleged dilutions.
>
> The cases demonstrate, and the Committee intends that there is no requirement that *any particular number of factors be proved, or that a majority of them point one way or another* (U.S. Congress. Senate. Judiciary Committee 1982, emphasis added).

Although these factors were to be used when evaluating whether multi-member at-large districts were discriminatory and neither the court nor the Senate report specified that these factors should be used when drawing district lines, they show how complex the concept of fair representation can be.

[218] Even compactness in some of its forms is difficult to formalize. For example, Grofman (1993) argues that instead of using purely formal compactness criteria for evaluating district geography, we should turn to *cognizability,* the ability of a legislator to define, in commonsense language, based on geography, the characteristics of her district.

algorithms in just these sorts of *natural* recognition tasks (Dreyfus 1992). Evaluation of any one of the circumstantial factors listed above is beyond the reach of current computer technology, and it is unlikely that the "totality of the circumstances"[219] can be adequately formalized in the foreseeable future.[220] The legislature would be abdicating its responsibility if it allowed such goals to be jettisoned in favor of automation.

Even goals that may seem to be simple, easily quantified, and even generally agreed upon may be subject to several different, and unequal, quantifications. For example Niemi, et al. (1991) lists close to two dozen quantifications of the goal "compactness," most of which will differ (in at least some cases) in the values they assign to districts. For example, the controversy over apportionment methods (how to allocate representatives on the basis of population) in the U.S. shows that even seemingly small differences in the formalization of an informal standard may be of great practical importance.[221]

---

[219] Although it might be claimed that the three prongs of *Thornburg* v. *Gingles* has eliminated the need to look at the totality of the circumstances, this is not the case. *Growe* v. *Emison* and *Voinovich* v. *Quilter* and *Johnson* v. *DeGrandy* clearly assert that these three conditions are necessary but not sufficient.

[220] See also Karlan (1993), for further arguments concerning the limited ability of formalism to capture representational goals.

[221] See Balinski and Young's excellent book on the politics of choosing an

Furthermore, redistricting typically involves weight multiple social goals.[222] The number of possible weights and weighting functions is literally infinite. Even should

---

apportionment method (Balinski and Young 1982). Also see *U.S. Department of Commerce* v. *Montana* (1992) for an overview of this history and its legal significance.

[222] Many state constitutions contain multiple criteria for evaluating districts (Grofman 1985). An extreme example of what an automated procedure would have to accomplish is illustrated by the Hawaiian constitution, which requires the following:

a. No district shall extend beyond the boundaries of any basic island unit.

b. No district shall be so drawn as to unduly favor a person or political faction.

c. Except in the case of districts encompassing more than one island, districts shall be contiguous.

d. Insofar as practical, districts shall be compact.

e. Where possible district lines shall follow permanent and easily recognized features, such as streets, streams, and clear geographical features, and when practicable shall coincide with census boundaries.

f. Where practicable, representative districts shall be wholly included within senatorial districts.

overall redistricting goals and their formulations be agreed upon, the choice of weights may still be subject to contention. The complexity of weighing different goals should not be underestimated. It is unreasonable to expect that a dynamic and subtle weighing of social values can be captured by a simple fixed set of linear weights. A study by Sheth & Hess in which political scientists were asked to give fixed linear weights to two popular redistricting goals (Sheth and Hess 1971) provides an example of the dramatic differences in weighting that can occur even in a relatively homogenous group deciding among extremely restricted options.

5.5.2. Political Implications of the Formalization of Redistricting Goals

You might well argue that while these three limits on formalization should not be ignored, they are not limits on formalization *per se*, but only limits on the interpretation of a particular formalization as "objectively neutral." Some proponents of automated redistricting (Browdy 1990a; Issacharoff 1993; Vickrey 1961) do not claim that automation makes redistricting neutral, but instead specify that the decisions involved in choosing and formalizing redistricting goals must be made by a political process. Even

---

g. Not more than four members shall be elected from any district.

h. Where practicable, submergence of an area in a larger district wherein substantially different socio-economic interests predominate shall be avoided.

when the goals and formalizations are chosen by a political process, however, we should expect that this type of formalization will have its disadvantages.

Seemingly simple characterizations of formal criteria, such as geographical compactness, can have counterintuitive implications for individual cases.[223] Given this fact, and the multiplicity of possible formalizations that we can develop even for such seemingly intuitive goals as "geographic compactness," automated redistricting will have three serious defects:

• First, automation will shift the political debate from redistricting goals to technical details because there are so many differing characterizations of common goals each with potentially different political effects. Automation therefore seems likely to engender extensive and arcane battles over the mathematical details of goal functions. As Richard Nelson writes in *The Moon and the Ghetto* (Nelson 1977): "However, there surely is a tension between the language of optimization... and creative problem solving, which implicitly is understood to be the proper language for looking at the real decision process."

• Second, requiring that all redistricting criteria be formalized will disadvantage social values that are difficult to mathematize. Representational goals such as preserving communities of interest or preventing minority vote dilution are particularly difficult to quantify because they succinctly condense a host of dynamically changeable, interacting

---

[223] See Young (1988) and Chapter 2 for examples of nonintuitive districts resulting from seemingly simple compactness measurements.

social factors. Current sophistication in computer programming is simply insufficient to capture such goals.

• Proponents of automated redistricting claim that it eases judicial and public review by making political purposes clear. Instead, formal characterizations often mask, rather that clarify, political purposes. In cases where formal criteria have counterintuitive implications, debate over particular district plans may illuminate political motives much more clearly than debate over the technical characterizations of redistricting goals.[224]

5.6. Discussion

Computers are eminently useful tools for many purposes — including redistricting. Unfortunately, good tools cannot always eliminate political and social problems: Automated redistricting is neither as simple nor as objective as has been claimed.

In this chapter I have demonstrated that the general redistricting problem and common sub-problems belong to a class of problems that is widely believed to be computationally intractable. Practical methods for generating districts automatically will almost certainly be based on heuristics— procedures which make guesses at solutions.

---

[224] Since the consequences of particular formulations are often less clear than the consequences of plans, one might expect that *risk averse* bargainers would prefer to bargain over plans than to bargain over goals, *ceteris paribus*. The ceteris *paribus* assumption is a strong one, however, and a formal game model of the bargaining process is needed to investigate these implications further.

The current proposals for implementing automated redistricting make exactly such guesses. This guesswork is disturbing because researchers offer little in the way of empirical data or theoretical analysis to show that these methods are successful in implementing our goals, and objective in excluding other considerations.

Even if we develop heuristics with better empirical and theoretical support, or go beyond heuristics altogether, we will run into political and normative difficulties when we try to make every districting goal recognizable to a computer algorithm. Representational goals, which are based upon the analysis of social and political patterns, are well beyond the current state of computer technology to quantify. Many simpler goals, such as geographic compactness, can be easily quantified, but are open to a plethora of conflicting quantifications.

If districts are created in an open political process, we may require only that districts meet minimum standards, such as population equality, in order to provide a “fair playing field”; and we may let the political procss determine the rest. If, instead, we override the political process and substitute for it a computer program, we must have confidence that the computer can not only satisfy our minimum standards, but that it can affirmatively meet our representational goals. There is a vast difference between the two.

Proponents claim that automated redistricting promotes fairness and illuminates the redistricting process. They assume that social goals for redistricting are easily and simply characterized. In fact, as this chapter argues, even the most seemingly straightforward of redistricting goals may be subject to many conflicting and confusing technical characterizations. More subtle social goals, especially those goals which are explicitly representational, may be given short shrift or have to be disregarded altogether to accommodate the automation process. Thus, the automation process advantages nonrepresentational goals over representatational one — a telling vice for an attempt to design a system of representation.

Some proponents also claim that automated redistricting can operate neutrally — merely implementing social goals chosen in a previous political debate. In fact, representational values are difficult or impossible to adequately formalize using current techniques, and no feasible methods for finding optimal plans for arbitrarily specified value functions have been discovered. Heuristic methods further complicate this picture because they intertwine our choice of values with how we create districts.

Proponents claim that automated redistricting eases judicial review. In fact, the technical complexities of characterizing goals formally, and of finding plans to serve these goals, is a barrier to judicial and public review and to participation in the districting process.

Finally, proponents have claimed that redistricting will reduce political manipulation. In fact, automated redistricting may only serve to change the types political manipulation that occurs. Automation may reduce political manipulation that uses redistricting plans for the purpose of advantaging particular individuals. It may promote, however, political manipulation using formal characterization of goals and choice of automation methods so as to advantage particular political groups. It may disadvantage particular politicians, but strengthen partisanship.

Proponents hold up automation as a veil of ignorance that can be used to promote fairness and prevent manipulation in the redistricting process. Unfortunately, automation may not eliminate the opportunity to politically manipulate, but instead shift that opportunity toward those groups that have access to the most extensive computing facilities. At the same time, automation shrouds this manipulation under the illusion of neutrality. Even if resources are equal, the potential for political side effects stemming from the choice of different characterizations of values and different heuristics for implementing plans, pierces the veil of ignorance and subverts the fairness of the automated process.

Appendix: Proofs

The most common way to show that a problem is NP-complete is to show that it can be polynomially reduced to another problem that is already known to be NP-complete. Similarly, the most common way to show that a problem is NP-hard is to show that solving it requires solving a problem that is known to be NP-complete. I will use these two general methods to show that the redistricting sub-problems discussed in Section 3.5 are NP-hard.

First, we will need some notation:

- We will use $\mathbf{x}_i$ to refer to the *i*th census block. These blocks are vector-valued, and we will assume that blocks comprise population values, partisan registration percentages, and geographic locations, among other features.
- We will use $\mathbf{d}_i$ to refer to the *i*th district. A district is a set of census blocks: $\mathbf{d}_i = \{\mathbf{x}_j, \mathbf{x}_k, \ldots, \mathbf{x}_n\}$.
- We will use $\mathbf{p}_i$ to refer to a particular plan. A plan is a partition, which has been defined in the Section 3.1 (footnote 20), of the set of all census blocks into a set of districts: $\mathbf{p}_i = \{\mathbf{d}_1, \ldots, \mathbf{d}_n\}$.

We can now move to the proofs. As the proofs are repetitive, I will show the first in more detail, and sketch the remaining three.

Claim 1: *Creating Equal Population Districts is NP-Hard*

Definitions: First we need to define a measure of population inequality. We can compute the population in a district by summing over its census blocks:[225]

$\mathrm{pop}(\mathbf{d}_i) = \sum_{\mathbf{x}_i \in \mathbf{d}_i} \mathrm{pop}(\mathbf{x}_i)$. We will use a simple measure of the inequality of a plan, V($\mathbf{p}_i$), the difference between the population of its largest and its smallest districts:[226]

$\mathrm{V}(\mathbf{p}_i) = \sup_{\mathbf{d}_i \in \mathbf{p}_i} \mathrm{pop}(\mathbf{d}_i) - \inf_{\mathbf{d}_i \in \mathbf{p}_i} \mathrm{pop}(\mathbf{d}_i)$.

Proof: Given a particular set of census blocks that we must district, let us label the set of all possible plans **P**. To draw districts that are as close as possible to this goal, we must find a plan, $\mathbf{p}^*$ that minimizes *V* over all possible plans using those census blocks, and having a fixed number of districts, *k*:

$$\mathbf{p}^* = \underset{\{\mathbf{p}_i \in \mathbf{P} \mid |\mathbf{p}_i| = k\}}{\arg\inf} \mathrm{V}(\mathbf{p}_i).$$

Suppose we produce an algorithm that finds $\mathbf{p}^*$, for any set of census blocks. Given $\mathbf{p}^*$ it is trivial to decide the question: “Is there a plan such that V($\mathbf{p}^*$)=0?”

---

225 I assume that the population in each census block is fixed and positive.

226 The redistricting problem remains NP-hard as long as *V*($\mathbf{p}_i$) equals 0 if and only if we have absolute population equality — my particular measurement of population equality is only for convenience.

But if we can answer this question we can solve an NP-complete problem. For k=2, answering the question is equivalent to solving the following NP-complete problem (listed in (Garey and Johnson 1983)), simply relabel our census blocks a, relabel our two plans A' and (A-A'), and relabel our population meaurement s:[227]

Set Partition

Instance: Finite set A and a size $s(a) \in Z^+$ for each $a \in A$.

Problem: Is there a subset $A' \subseteq A$ such that $\sum_{a \in A'} s(a) = \sum_{a \in A - A'} s(a)$?

This problem is NP-complete, but not strongly NP-complete. If we need to create a plan with more districts, the problem is strongly NP-complete. If we can decide whether V($\mathbf{p}^*$)=0 for k>2 then we can answer the following strongly NP-complete question (listed in (Garey and Johnson 1983)), by performing a similar relabeling:

3-Partition

[227] For an alternative formulation of partitioning as an integer program, see El-Farzi and Mitra (1992).

Instance:[228] Set A of $3m$ elements, a bound $B \in Z^{+}$ and a size $s(a) \in Z^{+}$ for each $a \in A$ such that $B/4 < s(a) < B/2$ and such that $\sum_{a \in A} s(a) = mB$.

Question: Can A be partitioned into m disjoint subsets such that for

$$1 \leq i \leq m, \ \sum_{a \in A_i} s(a) = B \ ?$$

Claim 2: *Creating A Maximally Compact District Plan is NP-Hard*

Definition: We will measure the compactness of a district as maximum distance between the centers of any pair of census blocks in that districts,[229] and we will measure the compactness of a plan by measuring its least compact district.[230]

---

[228] Since the 3-Partition problem does assume more restrictions for its set-members than we put on our census blocks, this shows our redistricting problem to be NP-Hard, but is not sufficient to show that it is NP-complete.

[229] Using this measurement, large values of V indicate the district is ill-compact. Papayanopoulos (1973) uses a similar definition of compactness. Furthermore, the maximum distance between census blocks in a plan is used as part of calculating a large number of other indices, see Niemi et al. (1991).

[230] It is a common practice to use the worst district to measure the compactness of an entire plan (Papayanopoulos 1973).

$$V(\mathbf{d}) = \sup_{\mathbf{x}_i,\mathbf{x}_j \in \mathbf{d}} \left(\text{distance}\left(\mathbf{x}_i, \mathbf{x}_j\right)\right)$$

$$V(\mathbf{p}) = \sup_{\mathbf{d}_i,\mathbf{d}_j \in \mathbf{p}} \left(V\left(\mathbf{d}_i, \mathbf{d}_j\right)\right)$$

Proof: Again, suppose we produce an algorithm that finds a $\mathbf{p}^*$ that solves our problem for any set of census blocks. Given $\mathbf{p}^*$ it is then trivial to decide the question (for some given constant D): "Is there a plan such that $V(\mathbf{p}^*)<D$?"

But if we can answer this question, we can solve an NP-complete problem. Answering the question is equivalent to solving the following problem, which is (strongly) NP-complete for any k>2 (discussed in (Johnson 1982)):

Distance-d Partition of Points in the Plane

Instance: Finite set $\mathbf{P} \subset Z \times Z$ of integer-coordinate points in the plane, positive integers D and k.

Problem: Is there a partition $\mathbf{P} = \mathbf{P}_1 \cup \ldots \cup \mathbf{P}_k$ such that if u and v are distinct points in some set P_i of the partition, then the Euclidean distance $d(u,v)<D$?

Claim 3: *Creating A Plan with Maximally Competitive Districts is NP-Hard*

Definition: Define a maximally competitive plan as one that minimizes the overall expected difference between Republican (R) and Democratic (D) registration[231] in each district:

---

[231] Instead of registration, we can use expected Republican and Democratic turnout, as long as the expected turnout in a census block is independent of the expected turnout in

$$\mathrm{V}(\mathbf{d}) = \sum_{\mathbf{x} \in \mathbf{d}} \mathrm{R}(\mathbf{x}) - \mathrm{D}(\mathbf{x})$$
$$\mathrm{V}(\mathbf{p}) = \sum_{\mathbf{d} \in \mathbf{p}} \left(\mathrm{V}(\mathbf{d})\right)^2$$

Proof: Again, suppose we produce an algorithm that finds a $\mathbf{p}^*$ that solves our problem for any set of census blocks. Given $\mathbf{p}^*$ it is then trivial to decide the question (for some given constant D): "Is there a plan such that $\mathrm{V}(\mathbf{p}^*)<D$?"

Again, if we can answer this question, we can solve an NP-complete problem. Answering the question is equivalent to solving the following problem, which is NP-complete (discussed in (Garey and Johnson 1983)):

Minimum Sum of Squares

Instance:[232] Finite set A, a size $s(a) \in Z^+$ for each a, positive integers $k \leq |A|$ and D.

Problem: Can A be partitioned into k disjoint subsets such that $\sum_{i=1}^{k}\left(\sum_{a \in A_i} s(a)\right)^2 \leq D$?

---

the district.

[232] Note that the minimum sum of squares problem assumes that each set member is positively valued. This makes it a subset of the competitive redistricting problem, because Democrats may outnumber Republicans in a census block. Hence this demonstration shows that the redistrict is NP-hard but is not sufficient to show that it is NP-complete.

This problem is NP-complete, but is not strongly NP-complete for fixed k (hence this is strongly NP-complete in the number of districts not in the number of census blocks). NP-completeness is preserved if the exponent 2 is replaced by any fixed rational number $\alpha>1$.

Claim 4: *Creating A Contiguous Equal-Population Redistricting Plan is NP-Hard*

Definition: We will use the same measurement of population as we did in the first problem. In addition, if any two census blocks i,j are contiguous, we will connect them with a unique edge $e_{i,j}$. Our problem in this case is to find a partition into k districts, such that the edges fully contained in that district form a connected graph, and such that we minimize $V(\boldsymbol{p})$, as in problem 1.

Proof: Again, suppose we produce an algorithm that finds a $\mathbf{p}^*$ that solves our problem for any set of census blocks. Given $\mathbf{p}^*$ it is then trivial to decide the question (for some given constant D): "Is there a plan such that all districts are contiguous and $V(\mathbf{p}^*)<D$?"

If we can answer this question, we can solve an NP-complete problem. Answering the question is equivalent to solving the following problem, which is (strongly) NP-complete (discussed in (Johnson 1982; Johnson 1984)):

Cut into Connected Components of Bounded Weight

Instance:[233] Graph **G**=(**V**,*E*), a size $s(v) \in Z^{+}$ for each *a*.

Problem: Is there a partition of **V** into disjoint sets $\mathbf{V}_1$ and $\mathbf{V}_2$ such that $\sum_{v \in \mathbf{V}_1} s(v) \le K$ and $\sum_{v \in V_2} s(v) \le K$ and both $\mathbf{V}_1$ and $\mathbf{V}_2$ induce connected subgraphs of **G**?

---

[233] Note that the minimum sum of squares problem assumes that each set member is positively valued. This makes it a subset of the competitive redistricting problem, because Democrats may outnumber Republicans in a census block. Hence this demonstration shows that the redistrict is NP-hard but is not sufficient to show that it is NP-complete.

Chapter 6. Do Traditional Districting Principles Matter?

> So geographers, in Afric maps,
>
> With savage pictures fill their gaps,
>
> And o'er uninhabitable downs
>
> Place elephants for want of towns.
>
> from *Poetry, a Rhapsody* by Jonathan Swift

6.1. Why **might** *traditional districting principles matter?*

Since its beginning, drawing maps has been an integral part of political life in America. Traditionally, it was politicians (sometimes with scholarly assistance) who shaped the lines that would, in turn, shape their own elections. The courts seldom intervened. Even when called upon, they would do no more than invalidate a single election (Cortner 1970). The current Court, in a spate of untraditional activism, has invalidated congressional redistricting plans, ironically claiming t.d.p.'s as its compass. In a series of cases, starting with *Shaw* v. *Reno* (1993, henceforth "Shaw I") and continuing in *Abrams* v. *Miller* (1997), it has all but required that states follow t.d.p.'s such as contiguity, compactness, respect for political boundaries, and population equality — if they create majority-minority districts. In a reversal of roles, scholars and politicians trail the court asking: Do traditional districting principles matter?

Is there justification for concern with gerrymanders? Will traditional districting principles "solve" the "problem" of racial and partisan gerrymandering? Early research failed to find much of a connection between intentional gerrymandering and electoral outcomes. Recently, racial gerrymandering has been blamed for many of the heavy Democratic losses in the 1992 and 1994 congressional elections (Hill 1995; Swain 1994). Furthermore, new studies have shown that redistricting has significant influence on representation. (Cameron, Epstein and O'Halloran 1996; Gelman and King 1994; Kousser 1995; Cain, 1985 #96 but see O'Rourke 1980; Rush 1993 for an opposing view.; Squire 1995) We know less about how to effectively control gerrymanders. While some political scientists have studied the effects of redistricting, other political and legal scholars have proposed to limit its abuses by imposing rules on redistricting (Polsby and Popper 1991; Stern 1974). These two bodies of work, one studying rules for redistricting, the other redistricting's consequences, have remained separate, and the efficacy of the latter's proposals remains largely untested.

In this study, I bridge the gap between the (primarily) empirical work on how redistricting affects elections and the (primarily) theoretical work on redistricting principles. Many redistricting principles (See Lijphart 1989, for a survey of both traditional and non traditional principles.) have been proposed on the basis of theory, under the supposition that they would control gerrymandering, reduce bias, and improve elections; few, however, have been subjected to rigorous empirical analysis. For example, even the most recent and detailed studies of district compactness (Niemi et al. 1990; Niemi and Wilkerson 1991) have not attempted to analyze the connections between this principle and electoral outcomes. Many empirical questions about redistricting principles remain unanswered: How much do redistricting principles affect gerrymandering? Do compact districts change the results of elections by increasing partisan fairness, or by making them more competitive? Do “ugly” districts decrease turnout, or change the attitudes of voters in those districts?

In this chapter I answer these questions by building on previous statistical maximum-likelihood analyses of redistricting and by applying these statistical models to a novel set of data. I measure the compactness, contiguity, “respect for political boundaries,” and malapportionment of all United States congressional districts from 1789 through 1993[234] — creating a set of data that has not previously, to my knowledge, appeared in print.

6.1.1. How Do T.D.P.’S Matter In The Court?

In the courts, many types of districts have been under attack, but congressional districts have undergone particularly close recent scrutiny by the Supreme Court. In all the recent cases in which the Court has particularly emphasized compactness and other t.d.p.’s, the Court has been looking at congressional districts.

Compactness, contiguity and malapportionment have long been factors in redistricting jurisprudence. “Traditional districting principles,” however, first entered the Court’s opinions in *Shaw I* and the Court continued to emphasize them in *Miller* v.

---

[234] At the time of writing, this data series is incomplete. The current series omits compactness scores for districts from 1913-1960 and omits intra-decadal redistrictings for compactness and malapportionment scores after 1913.

Johnson (1995) and in *Bush* v. *Vera* (1996). The way that the Court uses traditional districting principles, however, has changed even through these recent cases. (For the remainder of this paper, I will refer to “traditional districting principles” as t.d.p.'s and to districts that violate them as “ugly” districts, whether or not the districts actually appear irregular on a map.)

The Court, legal scholars, and political scientists, have long claimed that lack of compactness, contiguity and population inequality were the evidence and/or instruments of partisan, incumbent and minority vote dilution. (See Chapter 2.) In Chapter 1, I argued that the current Court has departed from previous jurisprudence and now claims two additional roles for t.d.p.’s: First, in *Shaw I* and in *Bush* v. *Vera* (1996), the Court claims that violations of t.d.p.’s cause symbolic and “expressive” harm by sending “pernicious messages” to both voters and politicians. Second, in *Miller* v. *Johnson* (1995) and in *Shaw II*, the court treats violation of traditional districting principles in majority-minority districts as circumstantial evidence of “racial classification.”

There is no need here to recreate the arguments of previous chapters, but a brief review of the Court’s use of t.d.p.’s is in order. In *Shaw I* the Court stated that redistricting legislation that is “so bizarre on its face that it is unexplainable on grounds other than race” demands close scrutiny. Although the Court stressed that compactness, contiguity and other criteria were not constitutionally required, they indicated that these are “objective factors that may serve to defeat a claim that a district has been gerrymandered on racial lines.” For the majority in this case, reapportionment was an area where “appearances do matter” because, in the majority’s view, districts that separate people by race while disregarding political and geographic boundaries reinforce the *perception* that members of the same race necessarily share political views. Such districts also send the “pernicious” message to politicians that they should only represent the majority voting group in the district. In *Shaw*, violation of t.d.p.’s is an integral part of the harm perceived by the Court — violation of these principles *actively* cause harm by sending a pernicious message to politicians. (Pildes and Niemi**,** in their oft-cited 1993 article, both give shape to the implicit and inchoate theory of the Court and name it *expressive harm* (Pildes and Niemi 1993). Adopting this terminology, O’Connor explicitly refers to “expressive harm,” in the plurality decision for *Vera*.)

In *Miller*, t.d.p.’s are still important, but they are no longer an integral part of the harm caused by redistricting. Instead, violations of these principles act as merely circumstantial evidence of intent:

> The plaintiff's burden is to show, either through circumstantial evidence of a district's shape and demographics *or more direct evidence going to legislative purpose*, that race was the predominant factor motivating the legislature's decision to place a significant number of voters within or without a particular district (emphasis added).

Similarly, in *ShawII* Justice Rehnquist maintains this use of traditional criteria as circumstantial evidence, writing that “The plaintiff bears the burden of proving the race-based motive and may do so either through circumstantial evidence of a district’s shape and demographics or through more direct evidence going to legislative purpose.” (116 S.

Ct. 1900) Justice Rehnquist explicitly denies Justice Stevens's dissenting claim that adherence to traditional districting principle isolates a case from strict scrutiny:

> In his dissent, Justice Stevens argues that strict scrutiny does not apply where a State respects or compl(ies) with traditional districting principles … That, however, is not the standard announced and applied in *Miller*, where we held that strict scrutiny applies when race is the Apredominant≅ consideration in drawing the district lines… such that Athe legislature subordinate(s) race-neutral districting principles... to racial considerations.≅

In *Vera*, which was delivered concurrently with *Shaw II*, the role of t.d.p.'s changed yet again — compliance with t.d.p.'s is not merely one piece in a body of circumstantial evidence, but is used as a threshold requirement for strict scrutiny. We can avoid strict scrutiny altogether, even if we are motivated by race, if we pay reasonable[235] attention to "traditional districting criteria": "We do not hold that any one of these factors is independently sufficient to require strict scrutiny. The Constitution does not mandate

[235] The court does not require that districts be drawn strictly to follow these criteria: "We thus reject, as impossibly stringent, the District Court's view of the narrow tailoring requirement, that a district must have the least possible amount of irregularity in shape, making allowances for traditional districting criteria."

regularity of district shape... and *the neglect of traditional districting criteria is merely necessary*, not sufficient" (116 S. Ct. 1953, emphasis added).[236] O'Connor, delivering the judgment of the court, stresses this point repeatedly, the same point that Rehnquist disavows :"Under our cases, the States retain a flexibility that federal courts enforcing 2 lack, both insofar as they *may avoid strict scrutiny altogether* by respecting their own t.d.p.'s"

6.1.2. Why Might Traditional Districting Principles Matter (To Political Scientists)?

Tradition may sometimes bear its own weight in the courts, but political scientists, as a rule, are interested in results. The Court posits two connections between t.d.p.'s and politics: They claim that violation of t.d.p.'s cause expressive harm that changes the behavior of voters and politicians, and that these violations are good proxies of gerrymanders. Not surprisingly, many political scientists and other scholars have made additional claims.

Many advocates of the control of gerrymandering through formal principles (traditional and otherwise) have singled out the "traditional" principle of compactness as

---

[236] Again, in *Abrams* v. *Johnson* (1997), the Court stressed the importance of traditional districting principles. Although it dealt with t.d.p.'s in less detail than the other cases citted, it weighed t.d.p's and especially the principle of following traditional county boundaries, in upholding the districts drawn to replace those invalidated by *Miller*.

especially potent. Many different compactness standards have been developed for this purpose; Niemi (Niemi, et al. 1990) provides a thorough survey. Advocates of compactness commonly claim that compactness will reduce or prevent partisan gerrymandering, and some claim that compactness standards will end gerrymandering, without harming representation (Stern 1974). Other advocates of compactness have made similar, but weaker, claims.

The effects that are claimed for various t.d.p.'s include, but are not limited to, the prevention of gerrymandering: Stern (1974) argues that compact districts, by limiting gerrymandering, will increase district competitiveness. Polsby and Popper, in an article in the *Yale Law & Policy Review* in 1991, claim that compactness, contiguity and population equality will together make "the gerrymanderer's life a living hell" (Polsby and Popper 1991). In a later article, they claim in addition that compactness will benefit geographically concentrated minorities (Polsby and Popper 1993).

O'Rourke argues that, in practice, districts that split county boundaries raise campaign costs and that districts that split precincts confuse both election officials and voters (O'Rourke 1995). Related to this is a hypothesis, originally put forward by Bernard Grofman, that violations of traditional districting criteria indicate severe violations of "cognizability" (Grofman 1985; Grofman 1993)[237]. Grofman's claims that

---

[237] Cognizability is, roughly, the ability to easily describe and identify the district. "Egregious violations of the cognizability principle can be identified by making use of

violations of cognizability can weaken voters' attachment to districts and legislators, strengthen incumbency advantage, and decrease turnout. Opponents of compactness and other t.d.p.'s dispute the claim that these restrictions will reduce gerrymandering without negative consequences. Cain shows that "esthetic" geographical criteria can interfere with "good government" criteria such as unbiasedness and competitiveness, and argues against the former on those grounds (Cain 1984, ch. 3). Lowenstein contends that compactness is not a fair standard because he believes that geographically compact districts systematically bias district plans in favor of Republicans (Lowenstein and Steinberg 1985). Karlan, in a pre-*Shaw* article argues in general that compactness is a limited principle and that violations of compactness may be necessary to prevent minority vote dilution (Karlan 1989); post-*Shaw* she specifically disputes the Court's claims that violation of t.d.p.'s cause "expressive" harms (Karlan 1993).

Pro and con, much has been claimed about the relationships between t.d.p.'s and political harms, but little has been empirically tested. Erikson searched for, but did not find, a relationship between electoral bias and malapportionment in the 1960's

---

standard criteria of districting, such as violation of natural geographic boundaries, grossly unnecessary divisions of local sub-unit boundaries (such as city and county lines), and sundering of proximate and contiguous natural communities of interest." (Grofman 1993, 1263) congressional elections[238] (Erikson 1972), but few other attempts have been made to connect districting principles with electoral effects. If any proponent of, or even some opponents of, t.d.p.'s are right, then districting plans that follow t.d.p.'s should be politically different from districting plans that violate t.d.p.'s. In general, are districting plans that are compact, contiguous, and well-apportioned better? Do such plans produce more competitive races? Are such plans, overall, more responsive to swings in partisan voting? Are they less biased? Do voters in such districts find their districts more "cognizable" — do such voters turn out to vote at a higher rate, vote less frequently for the incumbent, or have more trust in government or their representative? Do Representatives from such districts act differently?

To answer these questions required that I confront several hurdles. In previous chapters, I have developed ways of measuring t.d.p.'s quantitatively, obtained data sources that would allow me to assess t.d.p.'s, and use simulation to make prediction about the effects of compactness on partisan outcomes. In this final chapter, I put these measurements together with electoral and demographic data and use statistical models to analyze the effects of t.d.p.'s on electoral outcomes.

[238] Erikson found that reapportionment after the *Baker* v. *Carr* unintentionally advantaged the Republicans, but found no connection between this advantage and t.d.p.'s.

6.2. Data Sources

In Chapter 3, I measured the extent to which all U.S. decadal redistricting plans followed t.d.p.'s. This data is necessary, but not sufficient. In this chapter, I have combined my data on t.d.p.'s with numerous sets of data describing the other half of the relationship, political outcomes. Fortunately, these sets of data are more widely available and more readily accessible than data on t.d.p.'s.

For the political data used in this paper, I turned to a number of standard publicly available sources, in addition to those documented in Chapter 3. For records of voter turnout, and to compute the competitiveness, bias and responsiveness of district plans, I used election returns from I.C.P.S.R. [Inter-university Consortium for Political and Social Research, 1972 #82; King, 1994 #81. I relied on the latter source for electoral data after 1900, as it has undergone more extensive checking and error correction than the former source. The disadvantage of this source, however, is that it does not include third-party turnout in races which were primarily bipartisan. To evaluate voter trust I used survey data from the *American National Election Studies* (Miller and Studies. 1995). For detailed demographic and electoral data in the 1990's, I relied on U.S. Census STF3A data and *Politics in America* (Congressional Quarterly Staff 1996; U.S. Dept. of Commerce 1992).

6.3. Geographic and Population Measures

In this section I review the measures of t.d.p.'s used in the main analysis. These measures are the same as were used in the historical analysis of Chapter 3.[239]

For this data, most of the methods for measuring malapportionment give very similar evaluations of districts. In the main part of this analysis, I focus on the common and understandable measure, the population coefficient of variation.[240]

I adopt the same methods used to categorize contiguity and violations of traditional districting principles, as I have used in previous chapters. Roughly, contiguity refers to mathematical non-single-point contiguity, and "traditional boundaries" are equated with town, county and city lines. As I showed in Chapter 3, however measuring contiguity and breaches of "traditional boundaries" is more complicated in practice than in theory – Chapter 3 documents the (relatively small) number of instances where classification was questionable.

---

[239] In addition to these measures, a number of other less popular measures are used in the section comparing geographic and population measures (Section 4.1), so that the comparison between types of measures is broader.

[240] This is the standard deviation divided by the mean. Think of it as a measure of the average deviation.

6.4. Geography, Population and Politics

6.4.1. Can Compactness and Malapportionment be Measured Consistently?

As I showed in Chapter 3, trends in the numbers of non-contiguous districts and in the number of districts that split traditional boundaries are dramatic, even when difficulties in classifying such difficulties are accounted for. This section compares measurements of malapportionment and compactness. In this section we will see that malapportionment can be measured by a number of different methods, yet still yield the same rankings over districts. Compactness, however, is different. Compactness measures do not seem to be measuring the same thing.

For the purposes of this section, in order to make the comparisons more broadly applicable, I have added a number of additional popular measures of compactness and several popular measures of malapportionment. In order to compare the compactness and malapportionment scores over the same set of districts, this section focuses on the 1st through 62nd congresses, as both compactness and malapportionment data were available for this period.

There are a number of different methods to measure malapportionment used in the scholarly literature and by the courts. The most popular early measures of population variation were the difference between the largest and smallest districts (divided by the mean), the population variance ratio, which is the ratio of the largest to the smallest district, the maximum (or average) percent deviation from the mean, and the “electoral percentage,” which is the minimum percentage of the population represented by a bare majority of seats.[241] Later court cases have tended to stress the difference between the largest and smallest districts, and this measure was emphasized in *Karcher* v. *Daggett* (1983). These measures have a number of theoretical faults, and Foster (1985) argues cogently that such measures as the Gini index, Theil's measure of entropy, and the coefficient of variation have more desirable theoretical properties.

I included the "electoral percentage" for the sake of completeness, not because it is a particularly good measure. There are a number of problems that result from applying the "electoral percentage" to congressional district plans of individual states. First, as Dixon (1968) argues, this measure runs into numerical problems when a plan has few districts — in a state with two congressional districts the "electoral percentage" is necessarily 100 percent. Dixon also argues (See Chapter 7, Section 7.a of Dixon.) that this measure does not adequately reflect the realities of controlling an election, and I agree with him. It is also true that the other population measures fail to directly assess political power; but unlike the electoral percentage, the other measures can be argued to capture the formal, intrinsic value of an individual vote (Lowenstein 1990). Last, the electoral percentage was defined with legislative districts in mind, and when we apply it instead to congressional districts it loses much of its political meaning, since capturing a bare majority of a state's congressional seats does not have the same direct political implications as does capturing a legislature. Consequently, I expected that the electoral

[241] See Dixon (1968) Chapter 17, Section 4 and Chapter 18, Section 2.

percentage would perform poorly in these circumstances and would not be consistent with the other measures of malapportionment.

In the rest of this study, I have used three measures of compactness. For comparison in this section, I include in this section a fourth, and very common, measure, the direct area-perimeter (AP) comparison, defined as A/0.282P (Flaherty and Crumplin 1992).

	(Max-Min) /Mean	*Pop. Variance Ratio*	*Coeff. of Variation*	*Gini Coeff*	*Thiel's Entropy*
Pop. Variance Ratio	0.96				
Coeff. of Variation	0.77	0.76			
Gini Coeff.	0.81	0.80	0.89		
Theil's Entropy	0.80	0.79	0.91	0.94	
Electoral Percentage	-0.28	-0.28	-0.14	-0.18	-0.16

Table 6-1. Kendall's τ_B between population equality scores for multi-district plans: 1-62nd Congresses (349 observations).

Surprisingly, despite theoretical qualms about how best to measure malapportionment, all but one of these measures seem to be measuring the same underlying concept (Table 6-1). The courts can use any popular measure of malapportionment and they will come to the same conclusions in the vast majority of cases. This should give us confidence that these standards are being applied consistently, and that these standards are judicially manageable.

Unfortunately, we cannot be so confident about measuring compactness. Compactness measures do not seem to be measuring the same thing. Many of the measures of compactness in this study disagree more often than not. Table 6-2 and Table

6-3 show the level of agreement (τ_B) over scores for different compactness plans, while Table 6-4 shows the level of agreement among scores for individual districts; neither shows the same level of consistency that was shown by malapportionment measures.[242]

	AC	*AP*	*LW*
AP	0.08		
LW	0.56	0.09	
NORM	0.53	0.07	0.27

Table 6-2. Kendall's τ_B between mean compactness scores for multi-district plans: 1-62nd Congresses (349 observations).

[242] Here, my results disagree with those in (Niemi and Wilkerson 1991). There are five reasons for this disagreement: Niemi, et al. compare some other compactness measures that come from the same class of general measures, they use a much more limited range of district data, their district plans have more districts (15–100 districts) than the average congressional plan, and they use two measures of correlation, Pearson's p and Spearman's rho, that measure agreement less directly than Kendall's Tau. See Liebtrau (1983) for a discussion of the relative merits of these measures.

	AC	*AP*	*LW*
AP	0.23		
LW	0.51	0.29	
NORM	0.45	0.26	0.16

Table 6-3. Kendall's τ_B between minimum compactness scores for multi-district plans: 1-62nd Congresses (349 observations). Here, the compactness of a plan is defined by its least compact district.

	AC	*AP*	*LW*
AP	0.11		
LW	0.54	0.08	
NORM	0.51	0.09	0.28

Table 6-4. Kendall's τ_B between compactness scores for individual districts: 1-62nd Congresses (3390 observations).

In Chapter 2, I had noted a number of theoretical objections to the AP measure. We see, in practice, that the AP measure has little in common with any other measure, as might be expected from a standard that is entirely arbitrary.

Furthermore, although in theory sufficiently compact districts will be contiguous,[243] in practice compactness and contiguity are often at odds. Table 6-5 shows two different

[243] Compactness implies contiguity only for "well-behaved" measures of compactness,

measures of concordance between contiguity and compactness scores.[244] While I have included Kendall’s measure for consistency in presentation, all but 58 districts are contiguous, which causes an extreme percentage of ties across the contiguity category and distorts Kendall’s measure. A more appropriate measure in this case is Somer’s d which is very similar to Kendall’s measure except that it looks only at pairs of districts which differ in their levels of contiguity.

	τ_B	*Somer’s d*
AC	0.06	0.32
AP	0.08	0.44
LW	0.05	0.28
NORM	0.06	0.34

Table 6-5. Comparison between level of contiguity (contiguous, questionable, and non-contiguous) and compactness scores for individual districts: 1-62nd Congresses (3390 observations). All but 58 of these districts were practically contiguous.

Since both Somer’s d and Kendall’s Tau are positive, the more compact a set of districts, the more likely they are to be contiguous. This relationship is weak, however.

---

and even then only at some threshold level. See Altman 1995.

[244] I divided districts into three categories of contiguity, as described in Chapter 3.

Using Somer's d, we can see that when we compare two districts with different levels of contiguity, the least contiguous district will often be the more compact of the two.[245]

6.4.2. Relationships

Clearly, things have changed since the Court imposed, and strengthened, equal-population requirements. Fewer districts are contiguous, more districts split traditional boundaries, and districts are, at least by perimeter scores, less compact than they once were. Does it matter? Do violations of t.d.p.'s, which have clearly become more common, commonly affect politics? In this section, I test several of the most popular hypotheses regarding the political consequences of t.d.p.'s.[246]

---

[245] Both the Tau B measure and Somer's d use all 3390 observations to generate comparison pairs. Somer's d, however, effectively discards *pairs* where both members of the pair are contiguous (or both are equally non-contiguous).

[246] I have not tested all hypotheses related to the electoral effects of traditional district criteria. I would like to note three untested hypotheses here: First, I have not tested the effects of t.d.p.'s on incumbency advantage. Although I find that t.d.p.'s do not affect electoral responsiveness on the whole, it is possible that they affect the propensity for a voter to choose the incumbent. Second, I have not tested the hypothesis that "ugly" districts cause members of Congress to behave differently toward their electoral constituency. Third, the court claims that "ugliness" is of particular importance in

As there was a change in the rules of the game and in the appearance of districts in the 1960's, I focus on the period from 1962-1994. All districts were included in the analyses, except multi-member and at-large districts and a few others noted in the Appendix. Compactness scores, which are the focus of the most attention of all t.d.p.'s, are recorded for all four decadal redistrictings in this period.[247] Since the measurement of malapportionment becomes problematic after equal population standards were applied, as noted earlier, malapportionment scores are omitted from the 1970's decadal redistricting onward. Because of limits in the data sources, intra-decadal redistrictings are omitted, and violations of "traditional" boundaries are omitted from the 1980's decadal redistricting.

T.D.P.'s and Voter Attitudes

---

majority-minority districts. I have included, but not singled, out these districts in my analysis.

[247] I report scores for the normalized perimeter and the area measure, the most popular measures of compactness. All tests were duplicated with the length-width measure and with logged compactness scores on all measures, but the results were not substantially different.

Justice O'Connor, writing for the court,[248] contends that violations of t.d.p.'s, especially in majority-minority districts, sends "pernicious messages" to representatives and voters in such districts. Is there evidence for this contention? Do voters in ugly districts act or feel differently than voters in other districts? Do they trust government less or feel that Congress is less representative? Do they turn out to vote in fewer numbers in congressional elections?

The Court claims that members of Congress from ugly districts will represent only (or to a relatively greater extent) those who elected them. Do members of Congress from ugly districts represent only the winning part of the constituency? Do members from other districts represent everyone in the district, as the Court implies?

Given a particular group of voters with strong and easily identifiable interests, it is possible to directly measure the extent to which each member of Congress votes for the interests of these voters (Cameron, et al. 1996). In general, however, I cannot assume that I know the interest of a district's various constituents *a priori*, and survey data that could be used to estimate the interests of district constituents is incomplete or unavailable. We might expect, however, that the partisan competitiveness of the district should be a rough indicator of the "moderateness" of the district's constituents.

---

[248] In addition to Justice O'Connor, a number of legal scholars argue that redistricting can cause symbolic and expressive harms. (See Chapter 1.)

Accordingly, I constructed a measure of the partisan competitiveness of each district by taking the absolute value of the difference between partisan vote percentages for president in 1992. I used Poole-Rosenthal W-nominate scores to calculate the extremeness of each representative's voting pattern in the 103rd congress.[249] I regressed this measure, a dummy variable for the party of the representative, and compactness and contiguity measures, on the extremeness of the representative.

---

[249] Poole-Rosenthal scores measure the relative location of politicians in a multi-dimensional issue space. They tend to measure ideology better than interest group scores, and are the best currently known predictor of roll-call voting.

I defined the extremity of the representative as the absolute difference between that representatives W-nominate$_1$ score and the median score of the congress (Poole and Rosenthal 1991).

Poole-Rosenthal score data can be freely obtained from the authors through their web-site.

Intercept	0.52*	(0.03)
Democrat	-0.43*	(0.02)
Competitiveness	0.01*	(0.00)
Hispanic %	0.01	(0.06)
Black %	-0.14	(0.07)
Questionable Contiguity	0.05	(0.03)
Normalized Perimeter Compactness	0.00	(0.08)
Area Compactness	0.13	(0.09)
	Adj. R-squared	.66
	N. Obs.	426

Table 6-6. Effect of compactness on extremeness of representative.

None of the compactness parameters is significant at the p<0.05 level (two-tailed), although area compactness is significant at the p<0.10 level. The sign on the area compactness variable is the reverse of what advocates of compactness expect. In other words, more compact districts, if they have an effect, tend to increase the extremeness of voting behavior.

A regression of the W-nominate scores using demographic variables, and geographic terms *interacted* with party shows that the effect of compactness is more significant for Republicans than for Democrats, and that perimeter and area compactness may have opposite effects. This table shows that for Republicans, perimeter compactness is correlated with more centrist opinions, but area compactness is correlated with more extremist opinions (Table 6-7).

Intercept	0.47*	0.047
Democrat	-0.81*	0.058
Competitiveness	0.00*	0.001
Hispanic %	-0.08	0.068
Black%	-0.09	0.082
Area Compactness (Republican)	0.32*	0.153
Normalized Perimeter Compactness(Republican)	-0.33*	0.132
Questionable Contiguity (Republican)	-0.00	0.096
c_ac (Democrat)	-0.21	0.206
Normalized Perimeter Compactness(Democrat)	0.24	0.177
Questionable Contiguity (Democrat)	-0.07	0.109
	Adj. R-squared	0.85

Table 6-7 Effect of compactness interacted with party on issue position of representative.

The Court contends that "ugly" districts send a pernicious message to voters that political identity should be racial, a message that they claim threatens to deeply divide American society. It is natural to assume that such virulent messages should have some affect on voters' attitudes toward government and Congress. As a test of whether voters in "ugly" districts felt differently about government or Congress, I used individual survey-response data on trust in government and feelings toward Congress from the American National Election Survey. I estimated the effect of t.d.p.'s on the responses of potential voters to three questions[250] about congressional responsiveness, and general trust in government:

---

[250] Surveys of the size necessary to detect district-level effects are rare, and the range of available questions is limited: These questions do not ask about each voter's particular

• Question 610, which reads: "Generally speaking, those we elect to Congress in Washington lose touch with the people pretty quickly. (Agree/Disagree)" Higher values indicate disagreement.

• Question 625, which reads: "How much attention do you think most congressmen pay to the people who elect them when they decide what to do in Congress, a good deal, some, or not much?" Higher values represent more attention.

• Question 626, which is not an individual question, but a five-point index derived from five other questions. Higher values represent more trust in government.

To evaluate the influence of t.d.p.'s on voter's attitudes towards government, I used an ordered probit models with the survey responses as dependent variables.[251] (See, for example, Maddala 1983.) Where possible, I analyzed the first two surveys after each decadal redistricting. Some questions were, however, asked only during certain years. The results of this analysis are presented below (Table 6-8, Table 6-9).

---

representative, nor were all of them asked over the entire period of the N.E.S. Furthermore, even though the National Election Survey is the largest and most detailed survey to address congressional representation, the number of survey participants from each district is small.

[251] O.L.S. is often used for this purpose as well. O.L.S. analysis of ordered categorical variables can be biased if the interval between each pair of categories is not equal. The ordered probit model allows for unequal intervals among categories.

	Trust					
	1964	*1972*	*1974*	*1984*	*1992*	*1994*
Intercept	2.47 (0.10)	0.70 (0.08)	0.42 (0.098)	0.65 (0.082)	0.44 (0.071)	0.25 (0.08)
Percent Deviation from Mean District Population	0.29 (0.11)	--------	--------	--------	--------	--------
Split Sub-Unit Boundaries	0.02 (0.11)	-0.066 (0.055)	0.035 (0.066)	--------	--------	--------
Normalized Perimeter Compactness	-0.46 (0.27)	0.28 (0.25)	0.07 (0.31)	-0.34 (0.24)	-0.39 (0.27)	0.048 (0.25)
Area Compactness	-0.40 (0.30)	-0.27 (0.28)	-0.37 (0.35)	0.52 (0.30)	-0.037 (0.23)	-0.28 (0.30)
Questionable Contiguity or Discontiguous	-0.14 (0.14)	-0.067 (0.71)	-0.64 (0.81)	none in Sample	-0.03 (0.10)	0.13 (0.10)
Number of Observation	1356	2204	1532	1814	2204	1741
Correctly Predicted /% pred. with intercept only	43% /43%	24% /25%	38% /38%	24% /24%	37% /37%	44% /42%
Likelihood	1942	3500	2300	2900	3300	2500

Table 6-8. Relationship between T.D.P.'s and perceived congressional responsiveness and trust in government. The dependent variable in each model is an individual categorical survey response for a question on the ANES survey. The independent variables are measured district-level characteristics. Dashed lines indicate that the independent variable was missing for that year. Standard errors are reported in parentheses. Starred variables are significant at 0.05% (two-tailed test).

Attention				**Lose Touch**	
	1964	*1972*	*1974*	*1972*	*1974*
Intercept	1.1 (0.10)	0.91 (0.088)	0.60 (0.10)	-0.39 (0.09)	-0.44 (0.12)
Percent Deviation from Mean District Population	0.41 (0.13)	--------	--------	--------	--------
Split Sub-Unit Boundaries	0.072 (0.11)	-0.021 (0.61)	0.013 (0.08)	-0.12 (0.063)	-0.03 (0.082)
Normalized Perimeter Compactness	0.31 (0.32)	0.067 (0.27)	0.06 (0.33)	0.64 (0.28)	0.32 (0.39)
Area Compactness	0.58 (0.29)	-0.21 (0.30)	0.02 (0.36)	-0.78 (0.32)	-0.45 (0.43)
Questionable Contiguity or Discontiguous	-0.015 (0.16)	0.74 (0.83)	0.98 (0.84)	-2.8 (12.0)	-3.0 (15.0)
Number of Observation	1356	2701	1476	1452	1452
Correctly Predicted /% pred. with intercept only	43% /43%	52% /52%	53% /53%	69% /69%	69% /69%
Likelihood	1942	2100	1500	900	900

Table 6-9. Relationship between T.D.P.'s and perceived congressional responsiveness and trust in government. The dependent variable in each model is an individual categorical survey response for a question on the ANES survey. The independent variables are measured district-level characteristics. Dashed lines indicate that the independent variable was missing for that year. Standard errors are reported in parentheses. Starred variables are significant at 0.05% (two-tailed test).

This table lends no support to the hypothesis that t.d.p's have a significant affect on voters' attitudes toward Congress or on voters' trust of government. Although I have omitted the additional tables for brevity, I ran additional analyses adding race and other demographic variables from the N.E.S. survey as additional independent variables, and found the same results. Although the addition of demographic variables improves the predictive ability of the model, t.d.p.'s remain insignificant (at any conventional level).

Although this data does not support the conclusion, it is still of, course, statistically possible that t.d.p.'s have "expressive" effects. The burden of proof hould, however, be considered. Both legal and scientific tradition holds that we should require evidence before accepting a new claim. Quite simply, no evidence has previously been offered to show that "expressive harms" exist, and this is the first study to attempt to empirically detect them. The lack of evidence found here for expressive harms should signal the court that its attention should be focussed on the well known and easily detectable political effects of redistricting.

T.D.P.'s and Voter Turnout

Although people in "ugly" districts do not admit to feeling different about their member of Congress, they might still behave differently. One obvious way of showing disaffection with the electoral process would be to avoid the voting booth, and previous authors have hypothesized that ugly districts would decrease turnout.[252] Is voter turnout lower in ugly districts?

[252] More precisely, previous authors have argued that ugly districts proxy "cognizability," which affects turnout. (See Section 6.1.)

Voter turnout is known to be influenced by demographic factors, regional factors, and electoral factors. Previous research indicates that turnout is influenced by such diverse and interconnected variables as race, region, residential mobility, education, age, income, registration laws, campaign spending and campaign advertising (Ansolabehere et al. 1994; Cox and Munger 1989; Wolfinger and Rosenstone 1980).

Because of the interconnections between turnout variables, choosing “control” variables is difficult. In order to show how well t.d.p.’s alone can proxy political behavior, I estimate a model of turnout with those variables alone, as well as with demographic variables drawn from the 1990 census.

Data on turnout is available for most districts and election periods only in aggregate form.[253] Since a linear probability function is most commonly used for estimating changes in percent turnout, I show the results of this model (using weighted least squares to implement the minimum chi square method (Maddala 1983, p. 28)) for comparison. Both methods lead to the same substantive conclusions (Table 6-10).

---

[253] Rather than assume a model of turnout based entirely on aggregate variables, I also reanalyzed turnout with an individual-level logit model of the decision to vote, and estimate its values from the aggregate data using a log-odds model of the percent of voting-age-population that turned out for Republican and Democratic candidates in the congressional election (King 1989, pgs. 119 & 139). The substantive implications of both models are the same.

	Turnout as a % of VAP (1992)			*(1994)*		
	Shape and Demo-graphic	*Shape*	*Demo-graphic*	*Shape and Demo-graphic*	*Shape*	*Demo-graphic*
Intercept	0.71* (0.21)	0.45* (0.01)	0.79* (0.21)	0.45* (0.186)	0.34* (0.012)	0.46* (0.190)
Normalized Perimeter Compactness	0.087* (0.03)	0.11* (0.047)	------	0.12* (0.029)	0.16* (0.038)	------
Area Compactness	0.038 (0.038)	0.09 (0.052)	------	-0.02 (0.034)	0.03 (0.044)	------
Bad or Questionable Contiguity	-0.042* (0.013)	-0.066* (0.018)	------	0.00 (0.015)	-0.04 (0.020)	------
Percent Black	-0.21* (0.024)	------	-0.25* (0.024)	-0.21* (0.021)	------	-0.24* (0.021)
Percent Hispanic	-0.41* (0.026)	------	-0.43* (0.026)	-0.27* (0.022)	------	-0.27* (0.023)
Percent College+	0.19* (0.06)	------	0.17* (0.065)	0.18* (0.058)	------	0.16* (0.059)
log(Median Income)	-0.020 (0.021)	------	-0.024 (0.022)	-0.01 (0.019)	------	-0.01 (0.019)
Open Seat	0.009* (0.0084)	------	-0.01 (0.0089)	0.02* (0.012)	------	0.03* (0.012)
Adjusted R^2	0.53	0.11	0.49	0.47	0.09	0.44

Table 6-10. Relationship between t.d.p.'s and percent turnout in 1992 and 1994 elections. The independent variables are measured district-level characteristics. Dashed lines indicate that the independent variable was missing for that year. Standard errors (approximate) are reported in parentheses. At-large districts and uncontested districts in which elections were not held are omitted (417 observations).Starred variables are significant at 0.05% (two-tailed test).

These results tend to support O'Rourke's and Grofman's hypotheses, at least strictly: Ill-compact and non-contiguous districts tend to have lower turnout, even when other demographic factors are included in the model. The effects of compactness and contiguity are, however, dwarfed by demographic factors.

Partisan Bias and Electoral Responsiveness

Regardless of the non-effect of t.d.p.'s on individual voter behavior, do t.d.p.'s affect elections in the aggregate? Would elections be different if districts were "perfect"? In the final part of this section, I use Gelman's and King's method of evaluating electoral plans (Gelman and King 1994) and their *Judgeit* program to estimate the change in bias and electoral responsiveness[254] that would have resulted had districts been contiguous, evenly-apportioned (in 1962), compact, and aligned along traditional boundaries (this in 1962 and 1972 elections only).

I use Judgeit to estimate bias and responsiveness in each election using election returns from that election and the following election. I include, as "predictive" variables,

---

[254] Bias and responsiveness have a long history as measures of electoral results (Niemi and Deegan 1978). Responsiveness is, in essence, the elasticity of the seats-votes curve around its midpoint, and bias is a measure of the asymmetry of that curve.

all of the t.d.p. variables that were available for that decade. I then use *Judgeit* to evaluate a hypothetical election in which all districts are perfectly ‘pretty’.[255]

Obviously, it is not possible, in real life, for all districts to be perfect on all scores (notably compactness scores). Nor do the resulting hypothetical election results capture what would happen if districts were really perfect. What the hypothetical election results show is the statistical predictive relationship between t.d.p.’s and electoral results.

[255] More precise results could be obtained with additional data. In particular, because of limits in the present data-set, it is not possible to do analyses of compactness criteria between 1913-1961, or of intra-decadal redistrictings after 1913. Since the electoral effects of redistricting are strongest immediately after redistricting (Gelman and King 1994), however, the omission of intra-decadal data should tend to increase our estimates of the effect of compactness criteria, if anything. (Without the intra-decadal data, we cannot know if changes in compactness were greater than decadal changes. Nevertheless, the range of the independent variables in the decadal offers a rich range of compactness and malapportionment scores.)

	1962 (Observed)	(Ideal)	(I-O)	**1972** (Observed)	(Ideal)	(I-O)
Responsiveness at Vbar=0.45 to0.55	1.8 (0.063)	1.8 (0.07)	-0.03	1.4 (0.058)	1.4 (0.07)	0.017
Responsiveness at Vbar=Observed	1.8 (0.12)	1.8 (0.12)	-0.01	1.4 (0.11)	1.4 (0.11)	0.0065
Bias at Vbar=0.45 to0.55	-0.21 (0.0052)	-0.013 (0.0057)	0.20*	0.0036 (0.0040)	0.0032 (0.0056)	-0.0004
Bias at Vbar=050	-0.22 (0.0058)	-0.013 (0.0063)	0.20*	0.0039 (0.0044)	0.0036 (0.0060)	-0.0003

	1982 (Observed)	(Ideal)	(I-O)	**1992** (Observed)	(Ideal)	(I-O)
Responsiveness at Vbar=0.45 to0.55	1.7 (0.076)	1.6 (0.071)	-0.021	2.0 (0.067)	2.0 (0.085)	0.038
Responsiveness at Vbar=Observed	1.6 (0.13)	1.6 (0.12)	-0.01	1.8 (0.11)	1.8 (0.14)	0.019
Bias at Vbar=0.45 to0.55	-0.0059 (0.0052)	-0.0031 (0.0066)	0.0028	0.014 (0.0053)	0.014 (0.0054)	0.0006
Bias at Vbar=050	-0.0049 (0.0059)	-0.0023 (0.0072)	0.0026	0.015 (0.0058)	0.016 (0.006)	0.0003

Table 6-11. Comparison of observed congressional bias (positive values indicate a Democratic bias) and responsiveness in first elections following decadal redistrict, compared to "ideal" districts that always follow t.d.p.'s "At-large" districts and uncontested districts that did not hold elections were excluded. Significance is calculated for the columns of differences between ideal and observed, where starred variables are significant at 0.05% (two-tailed test).

The results in Table 6-11 indicate that in the elections immediately following the 1960, 1970, and 1980 decadal redistrictings, t.d.p.'s did not have a large effect. If districts had been ideally shaped, followed all traditional boundaries, and been evenly apportioned, then the 1960's elections would have been somewhat less biased against the Democrats, but little else would have changed. Districting plans, on the whole, would be no more responsive to partisan shifts, and bias in 1972 and in 1982 would not have been significantly altered.

6.5. Discussion

Despite theoretical misgivings about some popular measures of population inequality, all but one of the malapportionment measures that I tested produced very similar evaluations over a large set of real districts. In effect, the courts can use almost any convenient measure of population equality and obtain consistent results. Because of this, the courts were free to pick easily calculable and manageable population measures. This choice of measures was relatively uncontroversial, and the courts' resulting measurements of malapportionment have been predictable.

Unfortunately, this cannot be said of compactness measures. The measures of compactness in this study, which are typical examples of the three most popular types of geographical compactness measures, disagree more often than not, and compactness measures can disagree with assessments of practical contiguity.

Population equality is often held to be intrinsically valuable. The principle of "one person one vote" is enforced by the courts on its own merit, as well as in its instrumental role in creating fair outcomes. This intrinsic value lends strength to the concept of population equality and seems to make it easier to measure. On the other hand, if there is some intrinsic value to geographic compactness, the majority of proponents for compactness leave it unarticulated, and this study shows that numeric measures fail to agree on it.

This disagreement among measures raises concerns about the judicial manageability of compactness standards. The disagreement that I find among compactness evaluations contradicts Polsby's and Popper's (1991) assertion that compactness measures generally lead to the same results, and it supports Young's (1988), conclusion that compactness is still a "hazy and ill-defined concept" (pg. 114), and Lowenstein's and Steinbergs's (1985) contention that measuring compactness is neither simple nor straightforward.

Districts have changed in recent decades. Although in practically every decadal apportionment from the first through fifty-eighth Congress, some districts violated practical criteria of contiguity;violations of contiguity were few, and were concentrated in a small number of states (New York, Massachusetts, and the Carolinas). The violations of contiguity in the latest round of redistricting far exceeded traditional baselines.

Modern districting plans exceed their predecessors to an even greater extent in the frequency with which their districts violate town, counties and other sub-unit boundaries. Except in dense urban areas, early districts followed county and town boundaries exclusively. Districts split larger boundaries with increasing frequency after the 43rd

Congress, but the frequency of violations of "traditional boundaries" skyrockets following the population requirements imposed by the Court in *Reynolds* and *Wesberry*.

Of the four formal criteria examined in this paper, compactness has been the most controversial. How to define compactness and whether it can be usefully defined are questions that have been debated since the first compactness requirement passed by Congress. Unlike patterns of population equality and contiguity, patterns of compactness depend on which definition of "compactness" that we use. Whether or not there were traditional norms of compactness, and whether the current round of "ugly" districts violate these norms, depends crucially on the precise method that we choose to measure this property. Still, under a common definition of compactness, the normalized-perimeter score, district compactness has dropped steadily since *Reynolds*.

While districts have changed, these changes seem not to have had a large effect on elections. No one would deny that odd shapes and other violations of "traditional districting principles" can sometimes signal odd intents, but this research indicates that they do so only in the larger context of our political knowledge. Odd lines can indicate gerrymanders to the experienced researcher, politician, or judge, but only because the researcher also knows the political and demographic composition of the area on which the lines were imposed. Formal measurements and formal principles, are not, in general, sufficient to accomplish the task of judging politics.

The Court, and some legal scholars, believe that ugly districts change voter behavior. If ugly districts have an effect, it is small. I find no support for the hypothesis that "ugly" districts send pernicious messages to voters that affect their attitudes toward government or Congress. And although there is some evidence that, as O'Rourke and Grofman argue, violations of "traditional districting principles" reduce turnout, the effect is dwarfed by other well-known influences on turnout.

It has also been claimed that violations of "traditional districting principles" are proxies for gerrymandering. This research shows that, overall, "traditional districting principles" are not good proxies. A proxy is most useful when it provides us with a cheap way of learning about something that is otherwise expensive or impossible to study directly. While "traditional districting principles" are sometimes slightly predictive of bias and turnout, there are much more effective methods to determine the political effects of redistricting (Gelman and King 1994; Kousser 1995).

Cases

Abrams v. *Johnson*, 117 S. Ct. 1925 (1997)

Baker v. *Carr*, 369 U.S. 186 (1962)

Brown v. *Board of Education*, 347 U.S. 483 (1954)

Bush v. *Vera*, 116 U.S. 1941 (1996)

Colegrove v. *Green*, 328 U.S. 549 (1946)

Davis v. *Bandemer*, 478 U.S. 109 (1986)

Fortson v. *Dorsey*, 379 U.S. 433 (1965)

Gaffney v. *Cummings*, 412 U.S. 735 (1973)

Northeastern Florida Chapter Of The Associated General Contractors Of America, Petitioner (General Contractors) v. *City Of Jacksonville, Florida, Et Al. 508 U.S. 656* (1993)

Gomillion v. *Lightfoot*, 364 U.S. 339 (1960)

Growe v. *Emison*, 278 U.S. 109 (1986)

Hayes v. *Louisiana*, 114 S. Ct. 2731 (1994)

Holder v. *Hall*, 62. U.S.L.W. 4738 (1994)

Johnson v. *DeGrandy*, 62 U.S.L.W. 4755 (1994)

Karcher v. *Daggett* , 103 S. Ct. 2653 (1983)

Lujan v. *Defenders of Wildlife*, 504 U.S. 555 (1992)

Miller v. *Johnson*, 63 U.S.L.W. 4726 (1995)

Powers v. *Ohio*, 499 U. S. 400 (1991)

Reynolds v. *Sims*, 84 S.Ct. 1362 (1964)

Shaw v. *Hunt*, 517 U.S. 899 (1996)

Shaw v. *Reno*, 113 S. Ct. 2816. 92-357 (1993)

Thornburg v. *Gingles*, 478 U.S. 30 (1986)

United Jewish Organizations of Williamsburgh (UJO) v. *Carey*, 430 U.S. 144 (1977)

U.S. Department of Commerce v. *Montana*, 503 U.S. 442 (1992)

Wells v. *Rockefeller*, 290 FSupp 878 (1968)

Wesberry v. *Sanders*, 376 U.S. 1 (1964)

White v. *Regester*, 412 U.S. 765 (1973)

Bibliography

Adams, Bruce. 1977. “A Model State Reapportionment Process: the Continuing Quest for "Fair and Effective Representation".” *Harvard Journal on Legislation* 14: 825-904.

Adams, John. 1776. “Thoughts on Government.” In *The Works of John Adams*, ed. Charles Frances Adams. Boston:

Anderson, Arthur J. and William S. Dahlstrom. 1990. “Technological Gerrymandering: How Computers Can Be Used in the Redistricting Process to Comply with Judicial Criteria.” *The Urban Lawyer* 22(1): 59-77.

Ansolabehere, Stephen, Shanto Iyengar, Adam Simon and Nicholas Valentino. 1994. “Does Attack Advertising Demobilize the Electorate.” *American Political Science Review* 88(4): 829-38.

Appel, J.S. 1965. “A Note Concerning Apportionment by Computer.” *American Behavioral Scientist* 7: 36.

Avins, Alfred. 1966. “Literacy Tests and the Fourteenth Amendment: The Contemporary Understanding.” *Albany law Review* 30(1): 229-60.

Backstrom, Charles. 1982. “Problems of Implementing Redistricting.” In *Representation and Redistricting Issues*, ed. Bernard Grofman, Arend Lijphart, Robert S. McKay and Howard A. Scarrow. Lexington: Lexington Books.

Balas, E. and P. Toth. 1985. "Branch and Bound Methods." In *The Traveling Saleman Problem*, ed. E.L. Lawler, J.K. Lenstra, A.G.G. Rinnooy Kan and D.B. Shmoys. New York: John Wiley & Sons.

Balinski, Michel and H. Peyton Young. 1982. *Fair Representation.* New Haven: Yale University Press.

Bartholdi, John J., III, Craig A. Tovey and Micheal A. Trick. 1992. "How Hard is it to Control an Election?" *Mathematical and Computer Modelling* 16: 27-40.

Bellare, Mihir and Shaft Goldwasser. 1994. "The Complexity of Decision Versus Search." *SIAM Journal of Computing* 23:

Blair, B.J. and T.H. Bliss. 1967. "The Measurement of Shape in Geography: An Appraisal of Methods and Techniques." *Bulletin of Quantitative Data for Geographers* 11: 1-44.

Blumstein, James F. 1995. "Racial Gerrymandering and Vote Dilution: *Shaw* v. *Reno* in Doctrinal Context." *Rutgers Law Journal* 26: 517-95.

Blumstein, James F. 1996. "Shaw v. Reno and Miller v. Johnson: Where we are and where we are headed." *Cumberland Law Review* 26: 503-12.

Bodin, Lawrence D. 1973. "A Districting Experiment with a Clustering Algorithm." In Democratic Representation and Apportionment: Quantitative Methods, Measures, and Criteria New York: Annals of the New York Academy of Sciences

Book, Ronald V. 1994. "Relativizations of the P=?NP and Other Problems: Developments in Structural Complexity Theory." *SIAM Review* 36(2): 157-175.

Boyce, R.R. and W.A.V. Clark. 1964. "The Concept of Shape in Geography." *Geographical Review* 55: 561-72.

Browdy, Michelle H. 1990a. "Computer Models and Post-*Bandemer* Redistricting." *The Yale Law Journal* 99: 1379-1398.

Browdy, Michelle H. 1990b. "Simulated Annealing: An Improved Computer Model for Political Redistricting." *Yale Law & Policy Review* 8: 163-179.

Butler, Katharine Inglis. 1995. "Affirmative Racial Gerrymandering: Fair Representation for Minorities or a Dangerous Recognition of Group Rights?" *Rutgers Law Journal* 26(3): 595-624.

Butler, Katharine Inglis. 1996. "Affirmative Racial Gerrymandering: Rhetoric and Reality." *Cumberland Law Review* 26: 316-353.

Cain, Bruce. 1984. *The Reapportionment Puzzle.* California: University of California Press.

Cain, Bruce. 1984. *The Reapportionment Puzzle.* California: University of California Press.

Cain, Bruce. 1990. “Perspectives on *Davis* v. *Bandemer* : Views of the Practitioner, Theorist and Reformer.” In *Political Gerrymandering and the Courts*, ed. Bernard Grofman. New York: Agathon Press.

Cameron, Charles, David Epstein and Sharyn O'Halloran. 1996. “Do Majority-Minority Districts Maximize Substantive Black Representation in Congress.” *American Political Science Review* 90(4): 794-812.

Chandrasekharam, R., S. Subhramanian and S. Chaudhury. 1993. “Genetic Algorithm for Node Partitioning problm and application in VLSI design.” *IEE Proceedings-Series E* 140(5): 255-260.

Congressional Globe. 1842. 27th Cong., 2d Sess., vol. 11, pgs.

Congressional Globe. 1871. 42d Cong., 2d Sess., vol. 11, pgs.

Congressional Information Service. 1980. *CIS US congressional committee prints index : from the earliest publications through 1969*. Washington: Congressional Information Service.

Congressional Quarterly Staff. 1983. *Congressional Quarterly’s Guide to 1980 Redistricting*. Washington, D.C.: Congressional Quarterly, Inc.

Congressional Quarterly Staff. 1995. “Court Considers Question of Minority Districts.” *Congressional Quarterly Weekly Report* 53(April 22): 1133-4.

Congressional Quarterly, Inc. 1964. *CQ Census Analysis: Congressional Districts in the United States.* Washington, D.C.: Congressional Quarterly.

Congressional Quarterly, Inc. 1964. *CQ Census Analysis: Congressional Districts in the United States.* Washington, D.C.: Congressional Quarterly.

Congressional Quarterly, Inc. 1983. *Congressional Districts in the 1980's.* Washington, D.C.: Congressional Quarterly.

Congressional Quarterly, Inc. 1993. *Congressional Districts in the 1990's.* Washington, D.C.: Congressional Quarterly.

Congressional Quaterly Staff. 1996. *Congressional Quarterly's politics in America [1996 edition].* Washington, D.C.: CQ Press.

Congressional Record. 1882. 47th Cong., 1st Sess., vol. 12, pgs.

Congressional Record. 1901. 56th Cong., 2d Sess., vol. 7, pgs.

Cook, S.A. 1971. "The Complexity of Theorem-Proving Procedures." In Proceedings of the 3rd Annual ACM Symposium on Theory of Computing New York: Association for Computing Machinery

Cortner, Richard C. 1970. *The Apportionment Cases.* Knoxville: The University of Tennessee Press.

Cox, Gary W. and Micheal C. Munger. 1989. “Closeness, Expenditures and Turnout in the 1982 U.S. House Elections.” *American Political Science Review* 83(1): 217-31.

Cramer, Harald. 1946. *Mathematical Methods of Statistic*. Princeton: Princeton University Press.

Dahl, Robert A. 1989. *Democracy and Its Critics.* New haven: Yale University Press.

Dahl, Robert. 1956. *A Preface to Democratic Theory*. Chicago: University of Chicago Press.

Deng, Xiaotie and Christos H. Papadimitriou. 1994. “The Complexity of Cooperative Solution Concepts.” *Mathematics of Operation Research* 19: 267-266.

Deutsch, David and Richard Jozsa. 1992. “Rapid Solution of Problems by Quantum Computation.” *Proceedings of the Royal Society of London* 439: 553-558.

Deutsch, David. 1985. “Quantum Theory, the Church-Turing Principle and the Universal Quantum Computer.” *Proceedings of the Royal Society of London* 400: 97-117.

Dixon, Robert G., Jr. 1968. *Democratic Representation.* New York: Oxford University press.

Dixon, Robert. 1968. *Democratic Representation: Reapportionment in Law and Politics.* New York: Oxford University Press.

Downs, Anthony. 1957. *An economic theory of democracy.* New York: Harper.

Dreyfus, Hubert L. 1992. *What Computers Still Can't Do.* Cambridge: M.I.T. Press.

Eig, Larry M. and Micheal V. Seitzinger. 1981. *State Constitutional And Statutory Provisions Concerning Congressional and State Legislative Redistricting*. 81-143A.

El-Farzi, Elia and Gautam Mitra. 1992. "Solution of Set-Covering and Set-Partitioning Problems Using Assignment Relaxations." *Journal of the Operational Research Society* 43(5): 483-93.

Ely, John Hart. 1197. "Standing to Challenge Pro-Minority Gerrymanders." *Harvard Law Review* 111: 576-592.

Erikson, Robert S. 1972. "Malapportionment, Gerrymandering, and Party Fortunes in Congressional

Even, Shimon. 1973. *Algorithmic Combinatorics.* New York: The Macmillon Company.

Fenno, Richard F. 1979. *Home Style: House Members in Their Districts.* USA: Harper Collins.

Flaherty, Mark S. and Wiliam W. Crumplin. 1992. "Compactness and Electoral Boundary Adjustment: An Assessment of Alternative Measures." *The Canadian Geographer* 36: 159-71.

Flaherty, Mark S. and Wiliam W. Crumplin. 1992. “Compactness and Electoral Boundary Adjustment: An Assessment of Alternative Measures.” *The Canadian Geographer* 36: 159-71.

Forrest, Edward. 1964. “Apportionment by Computer.” *American Behavioral Scientist* 7: 23-25.

Foster, James e. 1985. “Inequality Measurement.” *Proceedings of Symposia in Applied Mathematics* 33: 31-67.

Frolov, Yu S. 1974. “Measuring the Shape of Geographical Phenomena: a History of the Issue.” *Izvestuya Vsesoyuznogo Geogficheskogo Obshchestva* 4: 281-91.

Garey, Micheal R. and David S. Johnson. 1983. *Computers and Intractability: A Guide to the Theory of NP-Completeness.* New York: W.H. Freeman and Company.

Garfinkel, R.S. and G.L. Nemhauser. 1970. “Optimal Political Districting by Implicit Enumeration Techniques.” *Management Science* 16(8): B-495-508.

Gelman, Andrew and Gary King. 1994. “A Unified Method of Evaluating Electoral Systems and Redistricting Plans.” *American Journal of Political Science* 38(2): 514-554.

Gibbs, Jack P. 1961. “A Method for Comparing the Spatial Shapes of Urban Units.” In *Urban Research Methods*, ed. Jack P Gibbs. Princeton: D. Van Nostrand Company, Inc.

Goff, Tom, “Reagan, Reinecke Denounce Court; Legislative Leaders Praise Action,” *Los Angeles Times*, 19 Jan. 1972, sec. A.

Griffith, Elmer C. 1974 (1907). *Rise and Development of the Gerrymander*. New York: Arno Press.

Grofman, Bernard, Lisa Handley and Richard G. Niemi. 1992. *Minority Representation and the Quest for Voting Equality*. Cambridge: Cambridge University Press.

Grofman, Bernard. 1985. “Criteria for Districting: A Social Science Perspective.” *UCLA Law Review* 33: 77-184.

Grofman, Bernard. 1993. “Would Vince Lombardi Have Been Right if He Had Said, ‘When it Comed to Redistricting, Race isn’t Everything, It’s the Only Thing’.” *Cardozo Law Review* 14: 1237-76.

Grofman, Bernard. 1993. “Would Vince Lombardi Have Been right If He Had Said: "When it Comes to Redistricting, Race Isn't Everything, It's the *Only* Thing?” *Cardozo Law Review* 14: 1237-1276.

Gudgin, G. and P.J. Taylor. 1979. *Seats, Votes and the Spatial Organization of Elections*. London: Pion Limited.

Guinier, Lani. 1994. “[E]Racing Democracy: The Voting Rights Cases.” *Harvard Law Review* 109-137:

Hale, Myron Q. 1966. “Computer Methods of Districting.” In *Reapportioning Legislatures*, ed. Howard D. Hamilton. Columbus: Charles E. Merrill Books, Inc.

Harris, Curtis C., Jr. 1964. “A Scientific Method of Districting.” *Behavioral Science* 9: 219-225.

Held, David. 1987. *Models of Democracy*. Stanford, California: Stanford University Press.

Hershberger, John. 1991. “Finding Tailored Partitions.” *Journal of Algorithms* 12: 431-463.

Hill, Kevin A. 1995. “Does the Creation of Majority Black Districts Aid Republicans? An Analysis of the 1992 Congressional Elections in Eight Southern States.” *Journal of Politics* 57(2): 384-401.

Hofeller, Thomas and Bernard Grofman. 1990. “Comparing the Compactness of California Congressional Districts Under Three Different Plans: 1980, 1982, and 1984.” In *Political Gerrymandering and the Courts*, ed. Bernard Grofman. New York: Agathon Press.

Idelson, Holly. 1995. “It's Back to Drawing Board on Minority Districts.” *Congressional Quarterly Weekly Report* 53(October 7): 3065-3066.

Infosynthesis, Inc. 1998. "USSC+". St. Paul: Infosynthesis, Inc.

Inter-university Consortium for Political and Social Research. 1972. “UNITED STATES HISTORICAL ELECTION RETURNS, 1824-1968 [Computer file].” In ed. Ann Arbor, MI:: Inter-university Consortium for Political and Social Research [producer and distributor],.

Inter-University Consortium for Political and Social Research. 1994. *CANDIDATE AND CONSITUENCY STATISTICS OF ELECTIONS IN THE UNITED STATES, 1788-1990 [COMPUTER FILE] 5th. ICPSR ed.* Ann Arbor, MI: Inter-University Consortium for Political and Social Research [Producer and distributor].

Issacharoff, Samuel and Thomas C. Goldstein. 1996. "Identifying the Harm in Racial Gerrymandering Claims." *Michigan Journalof Race and Law* 1: 47-65.

Issacharoff, Samuel. 1993. "Judging Politics: The Elusive Quest for Judicial Review of Political Fairness." *Texas Law Review* 71: 1643-1703.

Issacharoff, Samuel. 1996. "The Constitutional Contours of Race and Politics." *Supreme Court Review* 1995-2: 45-70.

Jewell, Malcom. 1962. "Political Patterns in Apportionment." In *The Politics of Apportionment*, ed. Malcom Jewell. New York: Atherton Press.

Johnson, David S. 1982. "The NP-Completeness Column: An Ongoing Guide." *Journal of Algorithms* 3: 182-195.

Johnson, David S. 1984. "The NP-Completeness Column: An Ongoing Guide." *Journal of Algorithms* 5: 433-447.

Joint Committee On Printing. 1913. *Congressional Directory*. 63rd Congress, 1st Session Washington, D.C.: U.S. Government Printing Office.

Joint Committee On Printing. 1923. *Congressional Directory*. 68th Cong., 1st Sess. Washington, D.C.: U.S. Government Printing Office.

Kaiser, Henry. 1966. “An Objective Method for Establishing Legislative Districts.” *Midwest Journal of Political Science* 10: 200-213.

Karlan, Pamela S. 1989. “Maps and Misreadings.” *Harvard Civil Rights and Civil Liberties Law Review* 24: 174-245.

Karlan, Pamela S. 1993. “All Over the Map: The Supreme Court's Voting Rights Trilogy.” *Supreme Court Review* : 245-87.

Karlan, Pamela S. and Darly J. Levinson. 1996. “The Importance of Political Deliberation and Race Conscious Redistricting: Why Voting is Different.” *California Law Review* 84: 1201-1232.

Karlan, Pamela, S. 1993. “The Rights to Vote: Some Pessimism About Formalism.” *Texas Law Review* 71: 1705-1740.

Karlan. 1996. “Still Hazy After all These Years: Voting Rights in the Post-*Shaw* Years.” *Cumberland Law Review* 26: 287-309.

Karp, R.M. 1972. “Reducibility Among Combinatorial Problems.” In *Complexity of Computer Computations*, ed. R.E. Miller and J.W. Thatcher. New York: Plenum Press.

Kelly, Alfred H. 1965. “Clio and the Court.” *Supreme Court Review* 1965: 119-58.

Kelly, J.S. 1988a. “Computational Complexity.” *Social Choice and Welfare* 5: 313-317.

Kelly, Jerry S. 1988b. “Social Choice and Computational Complexity.” *Journal of Mathematical Economics* 17: 1-8.

Kilgore, Sue T. 1997. “Between the Devil and the Deep Blue Sea: Courts, Legislatures and Majority Minority Districts.” *Catholic University Law Review* 46: 1299-1348.

King, Gary, John Bruce and Andrew Gelman. 1995. “Racial Fairness in Legislative Redistricting.” In *Classifying by Race*, ed. Paul Peterson. Princeton: Princeton University Press.

King, Gary. 1989. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference.* Cambridge, Mass.: Cambridge University Press.

King, Gary. 1994. “ELECTIONS TO THE UNITED STATES HOUSE OF REPRESENTATIVES 1898-1992 [Computer file].” In ed. MA: Gary King [producer] Cambridge. Ann Arbor, MI:: Inter-university Consortium for Political and Social Research [distributor], 1995.

Kousser, J. Morgan. 1996. "Estimating the Partisan Consequences of Redistricting Plans — Simply." *Legislative Studies Quarterly 21*

Kousser, J. Morgan. 1998. *Colorblind Injustice.* Forthcoming:

Kousser, Morgan 1992, "The Voting Rights Act and the Two Reconstructions" in Grofman, Bernard and Chandler, D. (eds.), *Controversies in Minority Voting*, Washington, D.C.: Brookings.

Kousser, Morgan J. 1995. "Shaw v. Reno and the Real World of Representation and Redistricting." *Rutgers Law Journal* 26(3): 626-710.

Lee, David R. and G. Thomas Sallee. 1970. "A Method of Measuring Shape." *Geographical Review* 60: 555-63.

Li, Ming and Paul M.B. Vitanyi. 1992. "Average case complexity under the universal distribution equals worst-case complexity." *Information Processing Letters* 42: 145-9.

Light, Larry, "New Arizona Districts: A Quandary for Udall," *Congressional Quarterly*, Jan 30, sec.

Liittschwager, John M. 1973. "The Iowa Redistricting System." In Democratic Representation and Apportionment: Quantitative Methods, Measures, and Criteria New York: Annals of the New York Acadamy of Sciences 221-234.

Lijphart, Arend. 1989. “Comparative Perspectives on Fair Representation: The Plurality-Majority Rule, Geographical Districting, and Alternative Electoral Arrangements.” In *Representation and Redistricting Issues*, ed. Bernard Grofman, Arend Lijphard, Robert B. McKay and Howard A. Scarrow. Lexington: Lexington Books.

Lively, Donald. 1992. *The Constitution and Race.* New york: Praeger.

Lowenstein, Daniel H. and Jonathan Steinberg. 1985. “The Quest for Legislative Districting in the Public Interest: Elusive or Illusory?” *UCLA Law Review* 33: 1-75.

Lublin, David. 1997. *The paradox of representation : racial gerrymandering and minority interests in Congress.* Princeton: Princeton University Press.

Maddala, G.S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics.* Cambridge, Mass.: M.I.T. Press.

Martis, Kenneth C. and Ruth Anderson Rowles. 1982. *The Historical Atlas of United States Congressional Districts, 1789-1983.* New York: Free Press.

Martis, Kenneth C. and Ruth Anderson Rowles. 1982. *The Historical Atlas of United States Congressional Districts, 1789-1983.* New York: Free Press.

Maveety, Nancy. 1991. *Representation Rights and the Burger Years.* Ann Arbor: University of Michigan Press.

Mayhew, David R. 1971. “Congressional Representation: Theory and Practice in Drawing the Districts.” In *Reapportionment in the 1970’s*, ed. Nelson W. Polsby. Berkeley: University of California Press.

McKay. 1965. *Apportionment: The Law and Politics of Equal Representation.* New York: The Twentieth Century Fund.

Miller, Warren E. and the National Election Studies. 1995. “ AMERICAN NATIONAL ELECTION STUDIES CUMULATIVE DATA FILE, 1952-1994 [Computer file] 8th ICPSR version.” In ed. Inter-university Consortium for Political and Social Research [distributor]. Ann Arbor, MI: University of Michigan, Center for Political Studies [producer].

Mitrinovic, D.S., Je. Pecaric and V. Volenec. 1989. *Recent Advances in Geometric Inequalities.* U.S.A.: Kluwer.

Morill, Richard. 1990. “A Geographer's Perspective.” In *Political Gerrymandering and the Courts*, ed. Bernard Grofman. New York: Agathon Press.

Moshman, Jack and Elaine M. Kokiko. 1973. “A Redistricting Algorithm Applied to Geographic Reorganization of Circuit Courts.” In Democratic Representation and Apportionment: Quantitative Methods, Measures, and Criteria New York: Annals of the New York Acadamy of Sciences

Moxley, Warden., Wayne. Walker and Robert Healy. 1974. *Congressional Districts in the 1970's.* Washington, D.C.: Congressional Quarterly.

Musgrove, Phillip. 1977. *The General Theory of Gerrymandering.* U.S.A.: Sage Publications, Inc.

Nagel, Stuart S. 1965. “Simplified Bipartisan Computer Redistricting.” *The Stanford Law Review* 17: 863-869.

Nagel, Stuart. 1972. “Computers and the Law and Politics of Redistricting.” *Polity* 5: 77-93.

Nelson, Richard R. 1977. *The Moon and the Ghetto.* New York: W.W. Norton & Co., Inc.

Niemi, Richard G. and John Deegan, Jr. 1978. “A Theory of Political Districting.” *American Political Science Review* 78: 1302-23.

Niemi, Richard G. and John Wilkerson. 1990. “Compactness and the 1980’s Districts in the Indiana State House: Evidence of Political Gerrymandering?” In *Political Gerrymandering and the Courts*, ed. Bernard Grofman. New York: Agathon Press.

Niemi, Richard G. and John Wilkerson. 1991. “Compactness and the 1980’s Districts in the Indiana State House: Evidence of Political Gerrymandering?” In *Political Gerrymandering and the Courts*, ed. Bernard Grofman. New York: Agathon Press.

Niemi, Richard, Bernard Grofman, Carl Carlucci and Thomas Hofeller. 1991. “Measuring Compactness and the Role of a Compactness Standard in a Test for Partisan and Racial Gerrymandering.” *Journal of Politics* 53: 1155-1179.

Niemi, Richard, Bernard Grofman, Carl Carlucci and Thomas Hofeller. 1990. “Measuring Compactness and the Role of a Compactness Standard in a Test for Partisan and Racial Gerrymandering.” *Journal of Politics* 53: 1155-1179.

Ordeshook, Peter. 1986. *Game Theory and Political Theory*. Cambridge: Cambridge University Press.

O'Rourke, Terry. 1980. *The Impact of Reapportionment.* Brunswick, N.J.: Transaction Books.

O'Rourke, Timothy G. 1995. “*Shaw* v. *Reno*: The Shape of Things to Come.” *Rutgers Law Journal* 26(3): 723-773.

Owen, Guillermo and Bernard Grofman. 1988. “Optimal Partisan Gerrymandering.” *Political Geography Quarterly* 7: 5-22.

Page, Scott E. 1994. “Two Measures of Difficulty.” *Social Science Working Paper, California Institute of Technology* 929:

Papadimitriou, Christos M. 1994. *Computational Complexity*. U.S.A.: Addison Wesley Publishing Co., Inc.

Papayanopoulos, L. 1973. “Quantitative Principles Underlying Apportionment Methods.” In Democratic Representation and Apportionment: Quantitative Methods, Measures, and Criteria New York: Annals of the New York Academy of Sciences

Parker, Frank R. 1990. *Black votes count : political empowerment in Mississippi after 1965.* Chapel Hill: U. North Carolina.

Parsons, Stanley B., Micheal J. Dubin and Karen Toombs Parsons. 1990. *United States Congressional Districts, 1883-1913.* New York: Greenwood Press.

Parsons, Stanley B., William W. Beach and Dan Hermann. 1978. *United States Congressional Districts, 1788-1841.* Westport: Greenwood Press.

Parsons, Stanley J., William W. Beach and Micheal J. Dubin. 1986. *Congressional Districts and Data, 1843-1883.* New York: Greenwood Press.

Parsons, Stanley J., William W. Beach and Micheal J. Dubin. 1986. *Congressional Districts and Data, 1843-1883.* New York: Greenwood Press.

Penrose, Roger. 1989. *The Emperor’s New Mind: Concerning Computers, Minds and the Laws of Physics.* New York: Penguin Books.

Penrose, Roger. 1994. *Shadows of the Mind.* Oxford: Oxford University Press.

Pildes, Richard H. 1997. “Principled Limitations on Racial and Partisan Redistricting.” *The Yale Law Journal* 106: 2505-2558

Pildes, Richard H. and Richard G. Niemi. 1993. "Expressive Harms, 'Bizarre Districts,' and voting rights: evaluating election-district appearances after Shaw." *Michigan Law Review* 92: 101-205.

Pitkin, Hannah. 1967. *The Concept of Representation.* Berkeley: University of California Press.

Polsby, Daniel D. and Robert D. Popper. 1993. "Ugly: An Inquiry into the Problem of Racial Gerrymandering Under the Voting Rights Act." *Michigan Law Review* 92: 652-582.

Polsby, David and Robert Popper. 1991. "The Third Criterion: Compactness as a Procedural Safeguard Against Partisan Gerrymandering." *Yale Law & Policy Review* 9: 301-353.

Pounds, Norman J.G. 1972. *Political Geography*. New York: McGraw Hill.

Rawls, John. 1971. *A Theory of Justice.* Cambridge: Belknap Press.

Reinelt, Gerhard. 1991. *The Traveling Salesman: Compuational Solutions for TSP Applications.* New York: Springer-Verlag.

Reock, Earnest C., Jr. 1961. "Measuring Compactness as a Requirement of Legislative Apportionment." *Midwest Journal of Political Science* 5(70-74):

Rose Institute of State and Local Government. 1980. "Reapportionment and Computers." In *California Redistricting*, ed. Claremont: Claremont Men's College.

Rush, Mark E. 1993. *Does redistricting make a difference? : partisan representation and electoral behavior.* Baltimore, MD: Johns Hopkins University Press.

Ryden, David K. 1997. *Representation in Crisis.* New York: State University of New York Press.

Schmeckebier, Laurence F. 1941. *Congressional Apportionment.* Washington, D.C.: The Brookings Institution.

Schwartzberg, J. 1966. "Reapportionment, Gerrymandering and the Notion of Compactness." *Minnesota Law Review* 50: 443-50.

Shaw v. Reno. 1993. 113 S. Ct. 2816.

Shepherd, J.W. and M.A. Jenkins. 1970. "Decentralizing High School Administration in Detroit: A Computer Evaluation of Alternative Strategies of Political Control." In Proceedings: Conference on Inter-Disciplinary research on Computer Science Computer Science Association of Canada:

Sherstyuk, Katerina. 1993. "How To Gerrymander: A Formal Analysis." *Social Science Working Paper, California Institute of Technology* 855:

Sheth, Pankaj J. and Sidney W. Hess. 1971. "Multiple Criteria in Political Redistricting: Development of Relative Values." *Rutger's Journal of Computers, Technology and Law* 6: 44-70.

Snyder, J. 1989. “Political Geography and Interest-Group Power.” *Social Choice and Welfare* 6: 103-125.

Squire, Peverill. 1995. “The Partisan Consequences of Congressional Redistricting.” *American Politics Quarterly* 23(2): 229-240.

Stanton, Dennis and Dennis White. 1986. *Constructive Combinatorics.* New York: Springer-Verlag.

Stern, Robert S. 1974. “Political Gerrymandering: A Statutory Compactness Standard as an Antidote for Judicial Impotence.” *The University of Chicago Law Review* 41(398-416):

Stern, Robert S. 1974. “Political Gerrymandering: A Statutory Compactness Standard as an Antidote for Judicial Impotence.” *The University of Chicago Law Review* 41(398-416):

Stern, Robert S. 1974. “Political Gerrymandering: A Statutory Compactness Standard as an Antidote for Judicial Impotence.” *The University of Chicago Law Review* 41(398-416):

Sunstein, Cass R. 1993. “Standing Injuries.” *The Supreme Court Review* : 37-64.

Taylor, A.J. 1989. *What Every Engineer Should now About Artificial Intelligence.* Cambridge: M.I.T. Press.

Taylor, P.J. 1973. “Some Implications of the Spatial Organizations of Elections.” *Transactions, Institute of British Geographers* 60: 121-36.

Taylor, Peter J. 1973. “A New Shape Measure for Evaluating Electoral District Patterns.” *The American Political Science Review* 67: 947-950.

Tobler, W.R. 1973. "A Continuous Transformation Useful for Districting." In Democratic Representation and Apportionment: Quantitative Methods, Measures, and Criteria New York: Annals of the New York Academy of Sciences

Torricelli, Robert G. and John I. Porter. 1979. “Toward the 1980 Census: The Reapportionment of New Jersey’s Congressional Districts.” *Rutger’s Journal of Computers, Technology and Law* 7: 135-156.

Traub, J.F. and H Wozniakowski. 1992. “Perspectives on Information-Based Complexity.” *Bulletin of he American Mathematical Society (New Series)* 26(2): 29-52.

Turing, A.M. 1937. “On computable numbers, with an application to the Entscheidungsproblem.” *Proceedings of the London Mathematical Society* 2(42): 230-265.

Tyler, Gus and David I. Wells. 1962. “New York: 'Constitutionally Republican'.” In *The Politics of Apportionment*, ed. Malcom Jewell. New York: Atherton Press.

U.S. Congress. Senate. Judiciary Committee. 1982. *Voting Rights Act Extension: Report of the Committee on the Judiciary on S. 1992*. 97th Cong., 2d sess., Rept. 97-417.

U.S. Government Printing Office. 1933. *Congressional Directory*. 73rd Cong., 1st Sess. Washington, D.C.: U.S. Government Printing Office.

U.S. Government Printing Office. 1943. *Congressional Directory*. 78th Cong., 1st Sess. Washington, D.C.: U.S. Government Printing Office.

U.S. Government Printing Office. 1953. *Congressional Directory*. 83rd Cong., 1st Sess. Washington, D.C.: U.S. Government Printing Office.

Van Alstyne, William W. 1965. "The Fourteenth Amendment, The 'Right' to Vote and the 39th Congress." *Supreme Court revew* 1965: 33-86.

Vickrey, William. 1961. "On The Prevention of Gerrymandering." *Political Science Quarterly* 76: 105-10.

Weaver, James B. 1970. *Fair and Equal Districts: A How-to-do-it Manual on Computer Use.* New York: National Municipal League.

Weaver, James B. and Sidney W. Hess. 1963. "A Procedure for Nonpartisan Districting: Development of Computer Techniques." *The Yale Law Journal* 72(2): 288-308.

Weaver, James B. and Sidney W. Hess. 1963. "A Procedure for Nonpartisan Districting: Development of Computer Techniques." *The Yale Law Journal* 72(2): 288-308.

Wells, David. 1982. "Against Affirmative Gerrymandering." In *Representation and Redistricting Issues*, ed. Bernard Grofman, Arend Lijphard, Robert B. McKay and Howard A. Scarrow. Lexington: Lexington Books.

Wells, David. 1982. “Against Affirmative Gerrymandering.” In *Representation and Redistricting Issues*, ed. Bernard Grofman, Arend Lijphard, Robert B. McKay and Howard A. Scarrow. Lexington: Lexington Books.

Young, H.P. 1988. “Measuring the Compactness of Legislative Districts.” *Legislative Studies Quarterly* 8: 105-115.

Young, H.P. 1988. “Measuring the Compactness of Legislative Districts.” *Legislative Studies Quarterly* 8: 105-115.

Zagarri, Rosemarie. 1987. *The Politics of Size: Representation in the United States, 1776-1850.* Ithaca: Cornell University Press

Zissimopoulos, V., V. Th. Paschos and F. Pekergin. 1991. “On the Approximation of NP-Complete Problems by Using Boltzmann Machine Method: The Cases of Some Covering and Packing Problems.” *IEEE Transactions on Computers* 40(12): 1413-1418.