THE PROMISE AND PERILS OF COMPUTERS IN REDISTRICTING

MICAH ALTMAN*
MICHAEL MCDONALD**

Since the 1960's, computers have been regularly promoted as both the salvation of redistricting and as a strong corrupting force. On the one hand, computing has been proposed as a way to remove human bias from the process of drawing electoral lines through automation and to detect gerrymanders through geographical and statistical analysis. On the other hand, computers have been accused of enabling redistricting authorities to effortlessly achieve any nefarious goal. The reality is more complex: fully automated redistricting is constrained by deep mathematical, computational, and philosophical limits; sophisticated analysis of redistricting plans has yielded better predictions of districts' electoral characteristics, but cannot serve as convincing "gerrymandering detectors." Although these advanced map-drawing tools have undoubtedly made the process faster and cheaper, they have not led to any fundamental changes in redistricting outcomes.

In the last decade, another application of computing to redistricting has emerged—the use of computing infrastructure to increase public participation. This use has tremendous potential to improve the redistricting process because there are no fundamental technical challenges to its success. Establishing standards for accessibility and transparency, however, will be critical.

* Micah Altman, Senior Research Scientist, Institute for Quantitative Social Science, Harvard University. Dr Altman has developed award winning research software, and authored numerous scholarly publications on redistricting and in the fields of informatics, statistics, computer science, political science, and other social-science disciplines.

** Michael P. McDonald, Associate Professor, Department of Public and International Affairs, George Mason University. He has been an expert or consulting witness in redistricting litigation in five states, has consulted for redistricting authorities in two states, and has written numerous scholarly publications on redistricting.

I. INTRODUCTION: COMPUTERS AND THE PROMISE OF IMPARTIAL REPRESENTATION

> "A computer does not substitute for judgment any more than a pencil substitutes for literacy. But writing ability without a pencil is no particular advantage."
>
> —Robert S. McNamara[1]

All electoral systems aggregate voters' choices into a collective outcome. Because no electoral system satisfies all values one might desire, the choice of electoral system ultimately embodies a preference among favored outcomes.[2] District-based systems are distinct, however, in that they incorporate the judgment of professional political actors into the vote aggregation mechanism *vis-a-vis* how individuals are partitioned within district boundaries. Districts created by these professionals can be defined to ensure representation for diverse communities: racial, geographical, economical, and political.

In theory, allowing political actors to periodically redefine district boundaries improves the quality of representation. In practice, it is often the case that, as the old saw goes, in districting, politicians pick voters instead of the other way around. Thus, the inherent tension of district-based systems is that if the role of the professional is not constrained, representation suffers, but if it is constrained too much the system loses its distinctive characteristic—the ability to dynamically incorporate human professional judgments into the quality of representation.

Over the six and a half decades since the creation of the first general purpose digital computer, computers successfully have conquered many problems that were once considered the exclusive domain of human experience despite the failure to create systems that are recognized as generally "intelligent."[3] Thus, it is perhaps unsurprising that many pose this question as a way to resolve the

---

1. ROBERT S. MCNAMARA, THE ESSENCE OF SECURITY: REFLECTIONS IN OFFICE 115 (1968).

2. Kenneth J. Arrow, *A Difficulty in the Concept of Social Welfare*, 58 J. POL. ECON. 328, 328–31 (1950).

3. *See generally* A.J. TAYLOR, WHAT EVERY ENGINEER SHOULD KNOW ABOUT ARTIFICIAL INTELLIGENCE (1989).

inherent tension in district-based electoral systems: *When it comes to redistricting, why not just let a computer do it?*

This question dates back to the 1960's. William Vickrey was perhaps the first to propose taming gerrymandering through automation.[4] Vickrey described an abstract algorithm for creating districts based only on population and proximity.[5] Variants of this algorithm remain in use for redistricting research. For example, it is quite similar to an algorithm used by Carmon Cirincione, Thomas Darling, and Timothy O'Rourke four decades later.[6] Later that decade, Stuart Nagel produced the first computer program for redistricting.[7] Unlike Vickrey, Nagel did not claim a gerrymandering solution, rather he sought to measure value judgments in such a way to increase transparency and thereby facilitate productive debate.[8]

Since these pioneering studies, many distinguished commentators have advocated computer automated redistricting, including:

- President Ronald Reagan: "There is only one way to do reapportionment—feed into the computer all the factors except political registration."[9]
- The Supreme Court: "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."[10]
- Multiple journalists: "Use a computer program to draw congressional district boundaries—one with no input for political party, race, affluence, urbanization or any other parameter now used to stack the deck."[11]

---

4. William Vickrey, *On The Prevention of Gerrymandering*, 76 POL. SCI. Q. 105 (1961).

5. Vickrey's proposal never mentioned the word "computer," but clearly implied the use of one. The algorithm described in the proposal was one that computers of the time could implement, and which would have been impractical in their absence. *Id.*

6. Carmen Cirincione, Thomas A. Darling & Timothy G. O'Rourke, *Assessing South Carolina's 1990s Congressional Districting*, 19 POL. GEOGRAPHY 189 (2000).

7. For a description of the program, *see* Stuart Nagel, *Simplified Bipartisan Computer Redistricting*, 17 STAN. L. REV. 863 (1965).

8. *Id.*

9. Tom Goff, *Governor Urges Redistricting Plan Without Partisan Politics*, L.A. TIMES, Jan. 21, 1972, at A3 (quoting then-Governor Ronald Reagan of California).

10. Karcher v. Daggett, 462 U.S. 725, 733 (1983).

11. *See, e.g.*, Joseph J. David, Jr., *Let a Computer Do It*, WASH. POST, May 21, 2003, at A32.

The idea of automating redistricting to reduce or eliminate gerrymandering is intuitively appealing, and is a recurrent theme in political science, geography, and law. Some notable examples include *Evaluation and Optimization of Electoral Systems*,[12] an exploration of the subject by mathematicians, and Michelle Browdy's student note on the subject in the Yale Law Journal.[13] Computer programmers outside of academia also find this an appealing approach, such as George Clark and Brian Olsen, who created software that draws compact districts.[14]

Each decadal redistricting since 1960 brought with it tremendous advances in computing technology and repeated promises of electoral salvation by computer. The practical reality of what computers can do in redistricting, and how they have been used, has been somewhat different but, the potential for computers to introduce greater transparency and public participation into redistricting does appear to be a realistic, if more modest, goal.

II. A TYPOLOGY OF COMPUTER USE IN REDISTRICTING

The first proposed application of computing to redistricting was aimed entirely at removing humans from the redistricting process by introducing quantitative data to create ostensibly "impartial" redistricting plans. The initial applications tended to produce redistricting plans that did not make representational sense and thus were not seriously considered. These innovations, nevertheless, demonstrated that computers were useful tools to process large amounts of data necessary to conduct redistricting under measurable constraints such as—and most notably at the time—equal population. Rapidly, applications of computers in the redistricting process evolved into different forms.

---

12. *See generally* Pietro GRILLI DI CORTONA ET AL., EVALUATION AND OPTIMIZATION OF ELECTORAL SYSTEMS (1999).

13. *See generally* Michelle H. Browdy, Note, *Computer Models and Post-*Banademer *Redistricting*, 99 YALE L.J. 1379 (1990).

14. *See, e.g.*, GEORGE L. CLARK, STEALING OUR VOTES: HOW POLITICIANS CONSPIRE TO CONTROL ELECTIONS AND HOW TO STOP THEM (2004); *Redistricter: A Non-Gerrymandered Impartial Redistricting Program*, http:// code.google.com/p/redistricter/ (last visited Apr. 14, 2010)

Table 1

A typology of computer use in redistricting.

Type of use	***Description***	***Goal***	***Early Studies***	***Current Maturity***
Geographic Information Systems	Systems that aid in capturing, managing, visualizing and analyzing spatial information	Aid in the efficient creation of maps associated with data.	Tomlinson[15]	Reached maturity in the 2000 round of redistricting.[16]
Fully Automated redistricting	Creation of legal, impartial redistricting plan entirely by computer.	Eliminate unfairness in redistricting.	Vickrey[17]	Not yet reached maturity,[18] and faces fundamental challenges.[19]
Semi-automated redistricting	Creation of legal plans by automated systems, based on criteria provided by the user.	Increase transparency in the redistricting process.	Nagel[20]	Available as a research prototype,[21] however still faces inherent limitations.[22]

15. *See generally* Roger F. Tomlinson, Computer Mapping: An Introduction to the Use of Electronic Computers in the Storage, Compilation and Assessment of Natural and Economic Data for the Evaluation of Marginal Lands, Proceedings of the National Land Capability Inventory Seminar (1962).

16. *See generally* Micah Altman, Karin Mac Donald & Michael P. McDonald, *From Crayons to Computers: The Evolution of Computer Use in Redistricting*, 23 SOC. SCI. COMPUTER REV. 334 (2005).

17. Vickrey, *supra* note 4, at 105–10.

18. Altman et al., *supra* note 16.

19. Micah Altman, *Is Automation the Answer: The Computational Complexity of Automated Redistricting,* 23 RUTGERS COMPUTER & TECH. L.J. 81 (1997).

20. Stuart S. Nagel, *Simplified Bipartisan Computer Redistricting*, 17 STAN. L. REV. 863, 863–69 (1965).

21. Micah Altman & Michael P. McDonald, *BARD: Better Automated Redistricting*, 34 J. STAT. SOFTWARE (forthcoming 2010).

Quantitative Indicia (such as geographic compactness measures, and measures based on the predicted seats-votes curve)	Automatically identify gerrymanders through geographic or statistical analysis of proposed redistricting plans.	Detect egregious gerrymanders.	Harris;[23] Reock,[24] Tufte[25] interpreting a measure by Edgeworth[26]	Prediction of electoral characteristics of proposed plans has reached maturity.[27] These predictive methods do not provide statistical evidence that gerrymandering caused a particular outcome,[28] and these models have not been widely accepted as reliable identifiers of impermissible gerrymanders.
Open Access	Computerized systems to offer access to plans, data for constructing plans, and tools to create plans.	Enhance transparency of redistricting facilitate public participation.	Altman, Mac Donald, McDonald.[29]	Just beginning to emerge in 2000 round of redistricting.[30]

22. Altman, *supra* note 19.
23. Curtis C. Harris, Jr., *A Scientific Method of Redistricting*, 9 BEHAV. SCI. 219 (1964).
24. Ernest C. Reock, *Measuring Compactness as a Requirement of Legislative Apportionment*, 5 MIDWEST J. POL. SCI. 70 (1961).
25. Edward R. Tufte, *The Relationship Between Seats and Votes in Two-Party Systems*, 67 AM. POL. SCI. REV. 540 (1973).
26. Frances Y. Edgeworth, *Miscellaneous Applications of the Calculus of Probabilities*, 51 J. ROYAL STAT. SOC'Y, 534, 534 (1898).
27. Bernard Grofman & Gary King, *The Future of Partisan Symmetry as a Judicial Test for Partisan Gerrymandering after* LULAC v. Perry, 6 ELECTION L.J. 2 (2007).
28. Micah Altman, *A Bayesian Approach to Detecting Electoral Manipulation*, 21 POL. GEOGRAPHY 39 (2002).
29. Altman et al., *supra* note 16.
30. *Id.*

The uses of computers in the redistricting process are summarized in Table 1. The goals most clearly distinguish each use, since the technologies and algorithms used to accomplish the goals overlap considerably. In the remainder of this section, we describe each of these goals in more detail and in the sections that follow we shall assess the state of the maturity of each: the development of geographic information systems, algorithms that provide for fully-automated redistricting by a computer, algorithms that provide for semi-automated redistricting given some inputs from a user, methods to use computers to detect the presence of gerrymandering through violation of quantitative indicia, and how computers may be used to improve open access to the redistricting process.

As applied to redistricting, the goal of geographic information systems (GIS) is to aid in the efficient creation of maps associated with data. In other words, GIS aims to support people in making decisions about districting. By itself, this is not controversial, but many have raised the concern that the manipulation of election data could make gerrymanders unprecedentedly easy to create and robust in effect. Thus, proponents of reform in this area of computing argue that the data available to professionals who use these systems should be artificially restricted to limit the ability to gerrymander by exploiting computer inputs. Iowa, for example, implements this reform strategy by preventing the legislature's advisory redistricting commission from considering election data or the residence of incumbents while drawing districts.[31]

The goal of automated redistricting is to eliminate unfairness. Proponents of automated redistricting argue that human judgments should be replaced by a set of neutral criteria such as equal populations, contiguity, and compactness within and across districts.[32] Many proponents for automated redistricting emphasize a particular neutral criterion, such as compactness,[33] rather than a particular computer program or algorithm. Regardless, the core argument is the same—by automatically creating lines to optimize a particular pre-specified set of criteria, we can retain a district-based system and

31. IOWA CODE § 42.4(5) (2008).

32. *See generally* Daniel D. Polsby & Robert D. Popper, *The Third Criterion: Compactness as a Procedural Safeguard Against Partisan Gerrymandering*, 9 YALE L. & POL'Y REV. 301 (1991).

33. *Id.* at 301–02.

eliminate gerrymandering. A difficulty with this approach is that seemingly neutral criteria may lead to biased outcomes; indeed, criteria may be chosen with a specific political outcome in mind.

The goal of semi-automated redistricting is to increase the transparency of the redistricting process.[34] In contrast to fully automated redistricting, this approach leaves the selection of redistricting criteria to people, and delegates the creation of a plan that optimizes that chosen criteria to computers.[35] Although both automatic and semi-automatic redistricting may sometimes use the same software and algorithms, fully automated redistricting algorithms and software are often highly-tailored to the particular criteria being advocated, whereas semi-automated algorithms must be sufficiently flexible to allow for a wide range of criteria to be effectively optimized. Thus, ironically, although the goals of semi-automated redistricting are "lower" than those of automated redistricting, the costs are considerably higher due to the difficulty of designing and implementing this type of redistricting.

The goal of quantitative indicia is to detect severe (if not all) gerrymanders based on districts' geographic and demographic characteristics. Proponents have argued that various indicia are associated with the presence of gerrymandering: Most commonly, these indicia are based on some measure of geographic compactness[36] or on an output from the hypothetical seats-vote curve based on results from the proposed districts.[37] Although some of these geographic criteria are theoretically calculable without a computer—and were first calculated in the mid-1800's—advances in computing have rendered a wider range of politically-plausible criteria to be calculated. It is easy, however, to conflate the prediction of electoral characteristics of plans with the detection of gerrymanders based on such predictions. The former has become fairly easy, whereas the latter remains deeply challenged.

---

34. *See* Nagel, *supra* note 20; Browdy, *supra* note 13, at 1387 (expanding upon Nagel's argument).

35. *Id.*

36. *See* Micah Altman, Districting Principles and Democratic Representation (Mar. 31, 1998) (unpublished Ph.D. dissertation, California Institute of Technology), *available at* http://thesis.library.caltech.edu/1871/.

37. For a typology of these approaches, *see generally* Richard G. Niemi & John Deegan, Jr., *A Theory of Political Districting*, 72 AM. POL. SCI. REV. 1304 (1978).

Finally, advocates of using computers to facilitate public access to redistricting stress the advantages of transparency and public engagement. In an article written with Karin Mac Donald,[38] we noted that in the 2000 round of redistricting, many states widely disseminated redistricting data and various proposed plans during the process for the first time. Some redistricting authorities considered publicly-drawn plans and encouraged public comments. Those supporting public access argued: "This change is tantalizing, since it suggests the potential for communities to examine the effects of plans on them, and to propose alternatives."[39] This public access aspect has a widely unrecognized potential to change the process of deliberation over districts by opening the door to wide public and interest group participation.

III. GIS—UBIQUITOUS AND UNJUSTLY FEARED

As computers have become more powerful, they have undoubtedly made it easier for redistricting authorities to create and evaluate more maps faster. At the same time, advances in communication technology have made it possible to gather fine-grained data to micro-targeting district boundaries. Many editorials have decried that this acceleration of GIS technology and data collection caused unprecedentedly sophisticated gerrymanders. For example:

- "Using powerful computers, line-drawers can now determine, with nearly scientific precision, how many loyal party voters need to be stuffed into any given district to make it impregnable."[40]
- "Mappers were able to specify a desired outcome or outcomes—the number of people in a district, say, or the percentage of Democrats in it—and have the program design a potential new district instantly. These systems allow redistricters to create hundreds of rough drafts easily

38. Altman et al., *supra* note 16.
39. *Id.* at 12.
40. Editorial, *Elections With No Meaning*, N.Y. TIMES, Feb. 21, 2004, at A14.

and quickly, and to choose from among them maps that are both politically and aesthetically appealing."[41]

- "Gerrymandering is not self-regulating anymore . . . the software has become too good."[42]
- "With sophisticated computer programs, politicians can draw lines to maximize precisely their party's representation and minimize the other's. The result is sham legislative elections in which fewer and fewer seats are competitive and moderates of both parties get squeezed out of office."[43]

Despite these repeated observations, computer technology does not appear to be at fault.[44] Gerrymandering existed even before Founding Father and namesake Governor Elbridge Gerry created a salamander-shaped Massachusetts state legislative district. Computers were primarily adopted as a redistricting tool much later, during the 1990s round of redistricting. Examining early adopter states that could either afford the expensive computer systems or were not barred by state law from analyzing election data during redistricting shows there to be little correlation between gerrymandering and redistricting outcomes.

Redistricting computer systems developed in the 1960s were innovative for their time, but did not provide a tool to quickly analyze calibrated changes in district boundaries with scientific precision. The units of geography analyzed by these programs were census tracts typically containing thousands of people, with data being entered on unwieldy punchcards. Resource-rich states began developing geographic information systems for redistricting purposes in the 1980s. By the 1990s, some states reported that their in-house developed redistricting GIS programs running on mainframe computers or state-of-the-art workstations. The software development and the compilation of the election and census data to power the software cost an average of $500,000, with some of the larger states

---

41. Don Peck & Caitlin Casey, *The Nation in Numbers: Packing, Cracking, and Kidnapping*, ATLANTIC MONTHLY, Jan./Feb. 2004, at 50, 50–51.

42. Jeffrey Toobin, *The Great Election Grab in* THE BEST AMERICAN POLITICAL WRITING 2004 165, 165–76 (Royce Flippin ed., 2004).

43. Editorial, *The Court Punts*, WASH. POST, May 2, 2004, at B6.

44. Altman et al., *supra* note 16 (analyzing the rise of computer assisted redistricting).

reporting spending several millions of dollars on redistricting systems.[45]

Ten years later, all states used redistricting software. The vast majority of states used commercially available programs capable of running on modestly-priced desktop computers, two orders of magnitude less expensive than a decade earlier.[46] However, the core capabilities of these systems did not change radically. Like their predecessors, the 2000 systems simply allow redistricting authorities to draw districts at the block level using point-and-click technology and evaluate basic measures of redistricting plans like population balance, party registration balance, and a few measures of compactness. They did not provide access to advanced statistical models, or multi-criteria optimization.

Another study we conducted with Karin Mac Donald did not find any statistical correlation between computer use, computer capabilities, or use of electoral data, and gerrymandered districts.[47] Moreover, the most dramatic changes in district competitiveness and compactness that occurred over the last forty years and preceded the widespread use of computers in the 1990s. Instead, the introduction of equal-population standards, which lead to the corrosion of use of whole counties, and the widespread use of block-level databases in the 1980s, appears to have been much more closely related to the increased ability of redistricting authorities to fine-tune districts.[48] As Bruce Cain relates in his case study of California's redistricting, a political guru like Phil Burton maintained a sophisticated database of California's political landscape in his head and was able to sit down with pen and paper to gerrymander a complex state like California.[49] Computers are a useful gerrymandering tool, but they do not necessarily replace a skilled person in this instance.

45. *Id.* at 337.

46. *Id.*

47. Micah Altman, Karin Mac Donald & Michael P. McDonald, *Pushbutton Gerrymanders*, *in* PARTY LINES: COMPETITION, PARTISANSHIP, AND CONGRESSIONAL REDISTRICTING (Thomas E. Mann & Bruce E. Cain eds., 2005).

48. *Id.* at 61–63.

49. Bruce E. Cain, *Assessing the Partisan Effects of Redistricting*, 79 AM. POL. SCI. REV. 320, 323–24 (1985).

IV. FULLY AUTOMATIC REDISTRICTING—PROBABLY POSSIBLE, PHILOSOPHICALLY PRESUMPTUOUS

> *There is a story about a very senior political scientist and a world-renowned scholar in the field of representation who traveled to Russia shortly after the fall of communism to lecture to the newly formed Duma. After speaking, a newly-minted member of the Duma approached him and asked him a question with great earnestness.*
>
> *"I have been elected as a representative," the Duma member asked, "so when I vote, should I vote the way I think the electors want me to, or should I vote the way I think is right?"*
>
> *"That's a good question," said the sage. "Scholars have been studying this for two thousand years. And, let me just say, there are many opinions."*[50]

Although fully automated redistricting has been propounded for forty-five years, until very recently it has been an impractical process to pursue, except in the most limited circumstances. Even in the last round of redistricting, there was no publicly available computer software capable of automatically creating congressional districts meeting minimal legal standards for equal population and contiguity, and no state claims to have generated their districting plan through a fully automated process.[51]

Times have changed. Although no commercial program is capable of automatic congressional redistricting, open source software projects are coming close to being able to provide this service for many states. The BARD system, for example, is being used experimentally to create contiguous, compact, equal-population districts using VTD's and census tracts, yielding results that are suggestive, but not yet acceptable by *de minimus* population standards. And the "redistricter" system[52] is able to create districts using census blocs that are within one percent of equal population. Neither reliably can create districts with legal population deviance, nor would equipopulous, contiguous, and compact districts be sufficient in the small number of states, such as Ohio, that have strict

50. This is a true story.
51. *See generally* Altman et al., *supra* note 16.
52. *Redistricter*, *supra* note 14.

requirements to minimize splits of county and other political boundaries. Notwithstanding this, the legality gap is narrowing rapidly for automated redistricting.

Figure 1

Redistricting plan for California, produced with "redistricter" package.[53]

There are, however, two important limitations to automated redistricting. The first limitation is mathematical: the problem of creating optimally compact, contiguous, equal-population districts is provably "NP-hard."[54] NP-hard partitioning problems are a class of problems generally considered by computer scientists and mathematicians to be computationally intractable[55] and probably impossible to create a computer program that solves these problems optimally and reliably except in very small or limited cases.

53. BDistricting: What Districting is and What it Could be, http://bdistricting.com (last visited May 27, 2010).

54. Altman, *supra* note 19, at 137–48.

55. Lance Fortnow, *The Status of the P versus NP Problem*, 52 COMMUNICATIONS OF THE ACM 9, 80–81 (2009). For a formal treatment of NP-Completeness *see* CHRISTOS M. PAPADIMITRIOU, COMPUTATIONAL COMPLEXITY 181–219 (1994) .

Redistricting is unlike the "hill-climbing" optimizers used to solve many statistical problems where there is clear path to the top of "hill," indicating the best solution for a mathematical function. For even a modest-sized jurisdiction there are an infinite number of ways census blocks can be assigned to districts. Across this infinite terrain, there are many peaks and valleys that define a redistricting plan that has equal population and contiguous districts minimizing the number of split counties. Further, the objective function to optimize becomes more complex when additional federal and state criteria are introduced. In this situation with multiple local optima there is no simple way to ensure that a local optimum, obtained by rearranging census blocks into districts, is indeed the global optimum.

The computational complexity of the redistricting problem is not limited to compact districts. Micah Altman showed that redistricting is computationally difficult even without optimal compactness.[56] Similarly, Clemens Puppe and Attlia Tasnadi showed that both *optimal* partisan gerrymanders and optimal partisan-unbiased redistricting are computationally hard as well.[57]

As might be expected from the theoretical computational complexity of redistricting, scholars have been unsuccessful at creating methods that yield solutions to significant problems. The most successful of the exact solution methods—those based on integer-programming formulations—have advanced considerably over the last decade but are still limited to relatively small problems. In a recent study, a group of scholars solved a redistricting problem for a fixed set of redistricting criteria on a 30x30 grid.[58] These authors speculated these methods could be extended up to a 50x50 grid. A typical state may have the equivalent of a 700x700 grid in absolute numbers of census blocks, but the blocks themselves typically do not align on a regular grid even in abstraction. Exact solution methods are additionally difficult or impossible to extend to arbitrary redistricting criteria. Integer programming, which is among the most general

56. Altman, *supra* note 19.

57. *See generally* Clemens Puppe & Attlia Tasnadi, *Optimal Redistricting Under Geographical Constraints: Why "Pack and Crack" Does Not Work*, 105 ECON. LETTERS 93 (2009); Clemens Puppe & Attlia Tasnadi, *A Computational Approach to Unbiased Districting*, MATHEMATICAL & COMPUTER MODELING 1455 (2008).

58. *See generally* Jeroen C.J.H. Aerts, Erwin Eisinger, Gerard B.M. Heuvelink & Theodor J. Stewart, *Using Linear Integer Programming for Multi-Site Land-Use Allocation*, 35 GEOGRAPHICAL ANALYSIS 148 (2003).

formulations for which exact solution methods are commonly used, requires extensive expertise to reformulate complex redistricting criteria—such as that necessitated by the Voting Rights Act—into integer partition constraints.

Because these problems are computationally intractable to general solutions, they are approached heuristically. Heuristics (or heuristic algorithms) are problem-solving procedures that, while they may yield acceptable results in practice, provide no guarantees of yielding "good" solutions in general. Specifically, an algorithm is heuristic if it cannot be shown to yield a correct result, or correct within a known error of approximation, or having a known probability of correctness. BARD, Redistricter, and every other program that has had even limited success in automated redistricting, use heuristics.

Computational intractability is a fundamental theoretical limitation of automated redistricting and is problematic for semi-automated approaches as well. The computational intractability might be overcome if the redistricting problem could be simplified greatly. For example, an optimal solution would not be necessary *if* (1) legitimate redistricting goals were limited to contiguity, equal population, and (some idiosyncratically-defined and approximate version of) compactness; (2) a system were developed yielded legal plans in practice; and (3) the plans by that system yielded were better than others proposed as measured by these agreed-upon criteria. If all of these elements come to fruition, *then* there might be widespread agreement that heuristics solutions were "good enough."

Unfortunately, no common agreement exists on the primacy of these goals or on the nature and measurement of representation. Even the political neutrality of these goals is contested, as there is evidence that compactness standards and others that emphasize geographic criteria can have distinct partisan effects where geographic patterns of support for each party differs.[59] Justice Kennedy reflected on this reality in *Vieth v. Jubelirer*[60] at the beginning of his concurrence: "The object of districting is to establish

59. *See generally* MICHAEL P. MCDONALD, THE MIDWEST MAPPING PROJECT (2009); Jason Barabas & Jennifer Jerit, *Redistricting Principles and Racial Representation*, 4 ST. & POL. Q. 415 (2004).

60. Vieth v. Jubelirer, 541 U.S. 267, 269 (2004) (Kennedy, J., concurring).

'fair and effective representation for all citizens.' . . . The lack, however, of any agreed upon model of fair and effective representation makes this analysis difficult to pursue."[61]

Political theory does not offer a clear solution either. Hanna Pitkin's foundational scholarship, which identifies formalistic, symbolic, descriptive, and substantive aspects of representation, has influenced political theory,[62] but the meaning, measurement, and current understanding of political representation remains complex, contested, and incomplete.[63]

Legal and academic scholars suggest many plausible criteria for evaluating the quality of districts, none of which are commonly implemented in fully-automated redistricting systems. For example, social scientists have suggested that the following criteria, among others, should be incorporated in redistricting:[64]

- Neutrality or symmetry of the projected seats-vote curve.[65]
- Range of responsiveness or the range of possible vote shares across which electoral results would change.[66]
- Competitiveness, maximizing the number of districts with competitive margins.[67]
- Consumer surplus or minimize the number of votes for a losing candidate.[68]
- Clustering, *per se*.[69]
- Continuity of representative relationship, (implying some degree of incumbency protection).[70]

---

61. *Id.*

62. *See generally* HANNA F. PITKIN, THE CONCEPT OF REPRESENTATION (1967).

63. *See generally* Nadia Urbanati & Mark E. Warren, *The Concept of Representation in Contemporary Democratic Theory*, 11 ANN. REV. POL. SCI. 387 (2008).

64. *See* Jeanne C. Fromer, *An Exercise in Line-Drawing: Deriving and Measuring Fairness in Redistricting*, 93 GEO. L.J., 1547, 1576–86 (2005) (outlining the various criteria suggested for evaluating districts).

65. *See generally* Richard G. Niemi & John Deegan, Jr., *A Theory of Political Districting*, 72 AM. POL. SCI. REV. 1304 (1978).

66. *Id.*

67. *Id.*

68. THOMAS L. BRUNELL, REDISTRICTING AND REPRESENTATION: WHY COMPETITIVE ELECTIONS ARE BAD FOR AMERICA (2008).

69. *See generally* Roland G. Fryer, Jr. & Richard T. Holden, *Measuring the Compactness of Political Redistricting Plans* (Nat'l Bureau of Econ. Research, Working Paper No. 13,456, 2007).

- Non-quantitatively defined communities of interest.[71]
- Further, states have these criteria:[72]
  - Coincidence with "major roads, streams, or other natural boundaries."
  - Coincidence with census tract boundaries.
  - Being "square, rectangular or hexagonal in shape to the extent permitted by natural or political boundaries."
  - Being "easily identifiable and understandable by voters".
  - Facilitating "communication between a representative and his constituents."
  - Preserving "media markets."
  - Enhancing "opportunity for voters to know their representative and the other voters he represents."
  - Aligning with "prior legislative boundaries."
  - Consistency with "political subdivisions."
  - Utilizing "vernacularly insular regions so as to allow for the representation of common interest."

If there was universal agreement on what "fair and effective representation" and how it should be measured, at least approximately, redistricting could be 'automated' even in the absence of computers. We would allow anyone to propose a plan, evaluate its representational quality using *a priori* agreed-upon measures, and pick the obvious winner at the close of an appointed period of time. A 2010 Ohio House redistricting reform proposal illustrates this approach.[73] Under the Ohio proposal, anyone can propose a plan, all submitted plans are evaluated for their representational quality (using a codified list of measures), and the best scoring plan will then be adopted.

---

70. *See generally* Nathaniel Persily, *In Defense of Foxes Guarding Henhouses: The Case for Judicial Acquiescence to Incumbent-Protecting Gerrymanders*, 116 HARV. L. REV. 649 (2002).

71. *See generally* Benjamin Forest, *Information Sovereignty and GIS: The Evolution of "Communities of Interest" in Political Redistricting*, 23 POL. GEOGRAPHY 425 (2004).

72. Roberto Casati, *Cognitive Aspects of Gerrymandering*, 20 TOPOI 203, 206–08 (2001).

73. H.J. Res. 15, 128th Gen. Assem. (Ohio 2009-10).

Practically, automated redistricted systems are not driven by any recognized measure of district quality but by the inability to calculate measures of chosen district qualities. Calculating any measure of districts or a redistricting plan as a whole may require sifting through a large amount of data. These calculations can quickly become expensive, which limits the speed by which an algorithm can search for a solution. The commonly recognized compactness measures that are used in automated systems are typically modified ad-hoc or new measures are created without rigorous peer-review concerning their strengths and weaknesses so as to increase computational speed. Examples of the compromises and adaptations inherent in the more recent systems and algorithms include the following:

> **Automatic redistricting to maximize voter homogeneity**.[74] This definition relies on maximizing the homogeneity of voters within districts. It is incompatible with many redistricting criteria, such as competitiveness, responsiveness, respecting the Voting Rights Act, and lack of bias. This algorithm assumes that voters can be assigned to districts individually (instead of only within census blocks), and may result in geographically non-contiguous districts that cannot guarantee optimality as self-defined.
>
> **A q-state Pott's model**. This algorithm, used by Chung-I Chou and S.P. Li, treats redistricting as a physics problem—a q-state Pott's model.[75] This solution heuristically attempts to yield districts that are contiguous, compact, and of equal population. Inherent in this approach is the assumption that no other criteria are relevant and that compactness can be defined exclusively as total plan boundary minimization—a definition rarely used in practice. In addition, optimality is not guaranteed, and Chou and Li report successfully applying the problem to less than 450 population units
>
> **Weighted Voronoi Diagrams**. This heuristic algorithm for location districting (which is similar to political districting) is based on weighted Voronoi diagrams. It also implicitly defines compactness

---

74. *See generally* Gregory B. Lush, Esteban Gamez & Vladik Kreinovich, How to Avoid Gerrymandering: A New Algorithmic Solution, presented at the Eighth Annual Conference on Intelligent Technologies (Dec. 13, 2009), *available at* http://www.cs.utep.edu/vladik/2007/tr07-51a.pdf.

75. *See generally* Chung-I Chou & S. P. Li, *Taming the Gerrymander—Statistical Physics Approach to Political Districting Problem*, 369 PHYSICA A: STAT. MECHANICS & ITS APPLICATIONS 799 (2006).

as boundary minimization.[76]

A shortest split-line algorithm. This algorithm splits area jurisdiction with alternating horizontal and vertical lines until districts approximately equal in population are developed. This yields districts that are roughly rectangular but not necessarily compact by any recognized measure. Nor does it allow for any other criteria to be incorporated. Even with these limitations, the districts produced still vary by five percent of the population, which is larger than allowed for congressional districts.[77]

An ad-hoc heuristic method that allowed a two-fold population difference between the largest and smallest methods. When Toshihiro Sakaguchi and Junichiro Wada tried to use this algorithm to automatically redistrict the Japanese Diet, it failed to find an optimal solution once the number of population units exceeded twenty.[78]

A system using mixed-integer programming heuristics. Although this family of heuristics is flexible enough to incorporate multiple criteria, only population equality, contiguity, and compactness were incorporated. This approach sought to achieve compactness by minimizing the maximum distance between two points in a district. Plans were generated for the nation as a whole and divided into less than a thousand population units for four multi-state districts. Even under these theoretical (and unrealistic) conditions, the system was unable to produce district plans that were well-balanced in population when the number of districts in a plan exceeded four.[79]

An ad-hoc randomized heuristic algorithm that did not explicitly incorporate population or compactness. When used to partially redistrict the North Carolina, it could only produce partial district

76. *Id. See generally* Antonio G.N. Novaes, J.E. Souza de Cursi, Arinei C.L. da Silva & João C. Souza, *Solving Continuous Location-Districting Problems with Voronoi Diagrams*, 36 COMPUTERS & OPERATIONS RES. 40 (2009) (providing a heuristic algorithm).

77. *See generally* Pan Kai, Tan Yue & Jiang Sheng, The Study of a New Gerrymandering Methodology (Aug. 16, 2007) (unpublished manuscript, on file with University of Science & Technology of China).

78. Toshihiro Sakaguchi & Junichiro Wada, *Automating the Districting Process: An Experiment Using a Japanese Case Study in* REDISTRICTING IN COMPARATIVE PERSPECTIVE 237, 242–50 (Lisa Handley & Bernard Grofman eds., 2008).

79. *See generally* Takeshi Shirabe, *District Modeling with Exact Contiguity Constraints*, 35 ENV'T & PLAN. 1 (2009).

> plans using grouped counties as population units.[80]
>
> **A *k*-means minimization heuristic**. "Redistricter" is a heuristic algorithm that essentially attempts to create districts where the component population units are closest to one another. There is no guarantee that any *k*-means heuristic algorithm will find the optimal configuration of population units into districts within a finite amount of time. Redistricter has undocumented modifications, and uses an ad-hoc combination of heuristics to "solve" the problem.[81]

In theory, fully-automated redistricting merely implements unambiguous redistricting criteria in the service of accepted representational goals. In practice, fully-automated redistricting criteria are modified for the sake of computational speed and to encode the representational goals of the system designer. Thus, fully-automated redistricting solutions put the proverbial cart before the proverbial horse.

V. SEMI-AUTOMATED REDISTRICTING—PROBABLY PROBLEMATIC, POSSIBLY USEFUL

Semi-automatic redistricting avoids the deep philosophical problem faced by fully-automated redistricting. Rather than hard-wire contested representational goals into the software, semi-automatic redistricting aims to be agnostic as to the choice of goals. The redistricting authority chooses criteria, and the software merely needs to find the optimal arrangement of districting division according to the chosen criteria. In theory, semi-automated redistricting increases transparency by moving the debate away from boundary line disputes to debates about the representational and legal constraints posed by such boundary divisions. And, in theory, the boundary lines are simply a consequence of these higher-level decisions.

Semi-automated redistricting is not so easy for two reasons. The first reason is mathematical—optimization of almost any interesting representational criteria is likely to be NP-hard. There is currently no

80. *See generally* David W. Peterson, *Putting Chance to Work: Reducing the Politics in Political Redistricting,* 21 CHANCE 1 (2008).

81. *Redistricter*, *supra* note 14.

known way of systematically determining the extent to which plans generated by any particular NP-hard heuristic are "close" to the optimum achievable. Further, many heuristics require *a priori* choices for starting points (*e.g.* an existing districting plan), and the results are sensitive to that choice. Since no heuristics work equally well for all problems, any requirement to use a particular software package or set of algorithms may create a hidden bias in favor of specific criteria.[82]

This bias is not merely theoretical. For example, Frederica Ricci and Bruno Simeone used several different general heuristics on the same set of optimization problems and found that one of these heuristics was much better than the others in achieving the goal of compactness.[83] Similarly, a study using three general heuristics for the same school-districting plan[84] found that the choice of starting points (seed values) dramatically affected the quality of the plans.[85]

Currently, the only openly available system for semi-automated redistricting is BARD, a system we developed. BARD provides methods to create, display, compare, edit, automatically refine, evaluate, and profile political districting plans. BARD aims to provide a framework for scientific analysis of redistricting plans and to facilitate wider public participation in the creation of new plans. Since redistricting is a computationally complex partitioning problem not amenable to an exact optimization solution, BARD implements a variety of selectable meta-heuristics that can be used to refine existing or randomly generated redistricting plans based on user-determined criteria. By assigning different weights to various criteria, such as district compactness or equality of population, BARD supports automated generation of redistricting plans and plan-profiling. These functions permit exploration of trade-offs among criteria. The intent of a redistricting authority may be explored by examining these trade-offs and examining why some reasonably observable plans were not adopted. Although we are enthusiastic about BARD, it has not solved the computational tractability problem. The use of multiple meta-

82. *See generally* D.H. Wolpert & W.G. Macready, *No Free Lunch Theorems for Optimization*, 1 IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION 67 (1997).

83. *See generally* Federica Ricci & Bruno Simeone, *Local Search Algorithms for Political Districting*, 189 EUR. J. OPERATIONAL R. 1409 (2008).

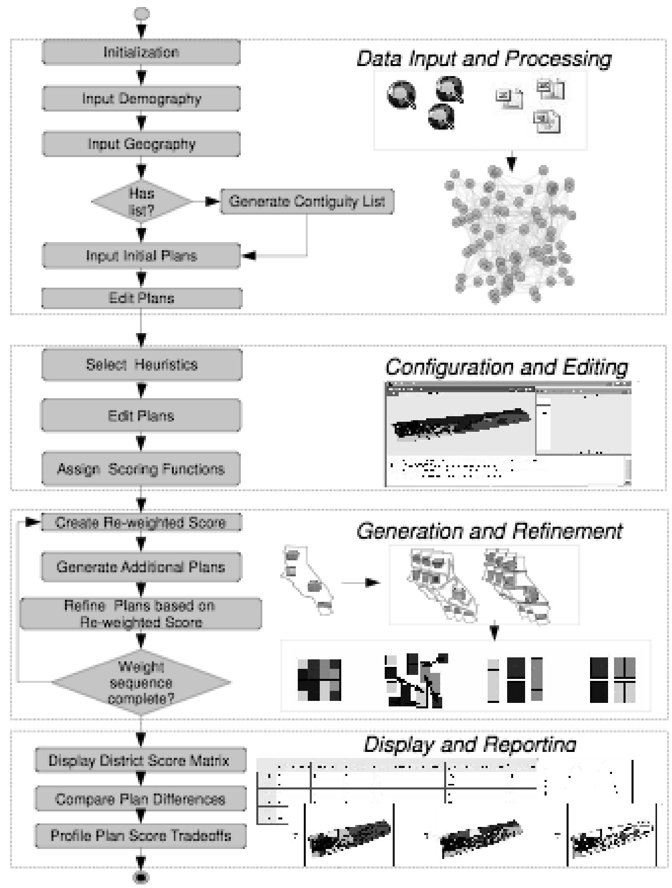
84. The optimization problem induced by school districting is similar but not identical to that of political districting.

85. *See generally* Marie desJardins et al., *Heuristic Search and Information Visualization Methods for School Redistricting*, 28 AI MAGAZINE 59 (2007).

heuristics can reduce the potential for hidden bias, it does not eliminate it.

Figure 2

BARD System Diagram and Screenshots[86]



86. Micah Altman & Michael P. McDonald, *BARD: Better Automated Redistricting*, 34 J. STAT. SOFTWARE (forthcoming 2010).

The second reason automated redistricting is difficult is that in practice there is a large gap between identifying representational values and creating criteria reflecting those values that can be optimized in a computer. Even conceptually simple values like "contiguity" require a host of technical choices when fed into a computer. Is a district contiguous if it is connected in some place only by telephone lines? By areas containing roads but not people? By areas containing no land (with or without bridges)? Are two districts contiguous if they cross each other at one or more points, or if one entirely contains the other? When measuring compactness, which of the dozens of standards should be used? Should entirely unpopulated areas be counted? Should areas covered by water be excluded or apportioned to neighboring land units? How should map orientation and scale be selected? The answer to each of these questions may yield substantially different outcomes, all of which could be claimed to yield the "best" contiguous, compact, equal-population district.

Although semi-automated approaches foster transparency, these approaches also can easily obscure it if the computer algorithms are not designed and documented carefully. These obstacles warrant caution, but they may be accounted for if there is sufficient flexibility and transparency in the redistricting process. Allowing a range of semi-automated systems using open-source documented algorithms, open criteria, and publicly available data may help redistricting authorities find plans that better meet their stated goals and better articulate why they chose a particular plan. Current commercially available automated systems, however, are closed, black box systems that do not lend themselves to verification that criteria were implemented correctly and do not enable advocates for new criteria to incorporate their recommendations into the software.

VI. QUANTITATIVE EVALUATION—RELIABLE PREDICTION, NOT CAUSAL LINKAGE

Redistricting criteria often are advocated as methods of detecting or limiting gerrymandering. Violations of criteria may reveal improper manipulation of districts, and such gerrymandering may be limited by requiring redistricting authorities to draw districts subject to certain prophylactic constraints. Although the use of redistricting criteria preceded the development of computers, computers have enabled a more expansive set of criteria to be calculated, including a set of statistical methods that provide relatively robust predictions of the hypothetical electoral outcomes expected to arise from proposed redistricting plans.

These criteria are typically quantitative, facially objective, strongly propounded by their advocates (at least for a time), and generally fall into three categories. The first category is a neutral proxy that correlates with (at least extreme) gerrymanders, and is otherwise "ostensibly" neutral. Compactness standards most often are propounded on this basis. The second category of criteria predicts probable electoral outcomes of redistricting plans. Among the many criteria in this category that have been proposed, Andrew Gelman's and Gary King's bias and responsiveness indices have received the most recent scholarly recognition.[87] The third set of criteria, which are more rarely used, aim to detect gerrymandering directly to reveal where a redistricting authority intended to achieve an improper goal in creating a district plan. Advances in computer technology have rendered a wide variety of criteria in each of these categories relatively easy to calculate. Computers, however, have not helped overcome inherent limitations associated with each type.

The fundamental limitation of the first type of criterion is neutrality. We say these criteria are "ostensibly" neutral because legislators and others in the public policy sphere have a long history of advocating *prima facie* neutral reasons and methods which are in fact chosen to achieve politically-motivated goals.[88] Consider, for example, the 1973 racial gerrymander of Mississippi county supervisors' districts, chronicled by Frank Parker and illustrated below.[89] In response to *Allen v. State Board of Elections*,[90] which invalidated at-large elections for Mississippi's county supervisors, Hinds County officials drew districts with the goal of equalizing population, land area, county road mileage, and the number of bridges (road maintenance was one of the responsibilities of county

87. *See generally* Andrew Gelman & Gary King, *A Unified Method of Evaluating Electoral Systems and Redistricting Plans*, 38 AM. J. POL. SCI. 513 (1994).

88. Robert Nozick uses the term "second-level bias" to refer to this selection of standards or procedures, which will be applied evenly, but which are chosen to advantage a particular group. ROBERT NOZICK, THE NATURE OF RATIONALITY 103–05 (1994).

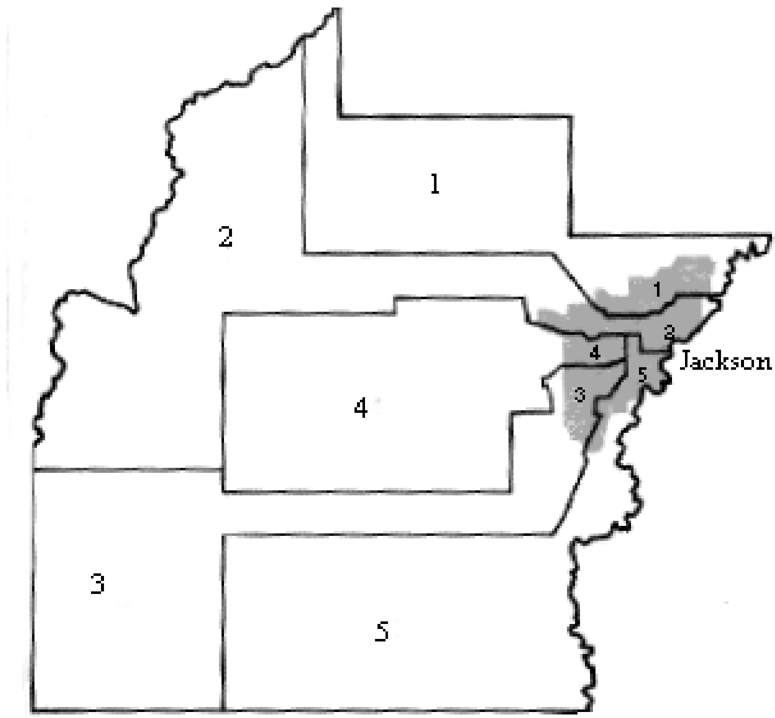
89. FRANK R. PARKER, BLACK VOTES COUNT: POLITICAL EMPOWERMENT IN MISSISSIPPI AFTER 1965 155–66 (1990).

90. Allen v. State Bd. of Elections, 393 U.S. 544 (1969).

supervisors).[91] Although these goals are facially neutral, the effect of their application was not:[92] simultaneously equalizing road mileage and population created districts that split the major city in Hinds County, where African-American voters were concentrated, and resulted in a racial gerrymander.[93]

Figure 3

Hinds County Supervisor's Districts Adopted by the Count Board of Supervisors in 1973. The gray areas represent concentrations of black population.[94]



91. PARKER, *supra* note 89, at 153–56.
92. *Id.*
93. *Id.*
94. *Id.*

Compact districts may advantage groups or political parties supported by geographically-dispersed populations, a scenario referred to in both the dissent and concurrence in *Vieth v. Jubelirer*.[95] Respecting existing political boundaries may also produce bias by a similar mechanism.[96] An important caveat is that these generalizable outcomes may not be applicable to all redistricting scenarios in all jurisdictions. Drawing compact districts may result in different outcomes in a state with several small to mid-sized cities compared to a state with a large metropolis.

The second set of indicia is best represented by what Richard Niemi and John Deegan called "neutrality,"[97] and by what Bernard Grofman and Gary King refer to as "partisan symmetry."[98] These indicia are based on the seats–votes relationship. This empirical relationship, first described by F.Y. Edgeworth[99] and interpreted as a measure of electoral manipulation by Edward Tufte,[100] has been formalized and estimated in different ways.[101] The most popular of these methods currently is derived from predictions of how changes in the statewide average district vote for a political party's candidate will translate into an expected statewide fraction of seats for the political party. As Grofman and King argue, advances in statistical methods (made practical by faster computers) now allow relatively easy and reliable estimates of these quantities.[102]

The prediction of plans' electoral characteristics may be easily conflated with the detection of gerrymandering based on these estimates. Whereas calculating these estimates has become fairly easy, using them to detect gerrymandering remains deeply challenged. The fundamental limitation of these criteria is that, as Gelman and King explain, they are predictive, not causal models, and are formally unrelated to gerrymandering. For example, non-zero bias and low

95. *See generally* Micah Altman, *Modeling the Effect of Mandatory District Compactness on Partisan Gerrymanders*, 17 POL. GEOGRAPHY 989 (1998) (citing Vieth v. Jubelirer, 542 U.S. 267 (2004) (Kennedy, J., concurring)).

96. MCDONALD, *supra* note 59.

97. Niemi & Deegan, *supra* note 37, at 1304–08.

98. Grofman & King, *supra* note 27.

99. Edgeworth, *supra* note 26.

100. Tufte, *supra* note 25.

101. *See generally* William LeBlanc, Party Positions and the Seats/Votes Relationship with Ideological Voters (Aug. 6, 2007) (unpublished Ph.D. dissertation, Massachusetts Institute of Technology), *available at* http://dspace.mit.edu/handle/1721.1/42388.

102. Grofman & King, *supra* note 27.

responsiveness (in the Gelman–King sense) may result from causes other than partisan gerrymandering, such as: geographic constraints, attempts at reducing the majoritarian-winner bonus (arguably *increasing* fairness) in situations where the normal vote is not expected to be fifty percent, a change in the proportion or location of moderate voters, or the pursuit of other legitimate goals.[103] To be useful in policy, predictive redistricting criteria must carry a substantive interpretation. For example, Grofman and King state that they view measures of partisan symmetry as a "substantive standard of fairness in districting" and not as evidence of improper intent to create a gerrymander.[104] However, courts have rejected the pursuit of bias and responsiveness exclusive to other redistricting criteria.[105] Additionally, as discussed in the previous section, neither the courts nor social scientists have reached consensus on what predictive criteria should apply.

The third category (intended to detect gerrymandering by a redistricting authority) avoids the limitations posed by the first two by focusing directly on the likelihood that a plan was caused by an improper motive—whether or not the plan is actually a classic gerrymander. Formally, this can be framed as a Bayes' factor test, which measures the ratio of the probability that the plan would have particular characteristics, such as the number of Democratic seats, based on the redistricting authority's purely partisan motives over the probability that a plan would have the same characteristics conditioned on a permissible motive.

Unfortunately, although such tests are easy to formulate, they are impossible to estimate because the distribution of plans under each outcome is unknown, and it is not possible to sample or simulate from it. Because redistricting is an NP-hard problem, it is not tractable to enumerate the population of possible districting plans of any reasonable size. As a consequence, no known algorithm can provide a

---

103. *See generally* Andrew Gelman & Gary King, *A Unified Method of Evaluating Electoral Systems and Redistricting Plans*, 38 AM. J. POL. SCI. 513 (1994); J.M. Kousser, *Estimating the Partisan Consequences of Redistricting Plans—Simply*, 22 LEG. STUD. Q. 521 (1996); *see also* LeBlanc, *supra* note 101 (illustrating an innovative formal analysis).

104. Grofman & King, *supra* note 27, at 9–13.

105. League of United Latin American Citizens (LULAC) v. Perry, 548 U.S. 399, 420 (2006) ("Without altogether discounting its utility in redistricting planning and litigation, we conclude asymmetry alone is not a reliable measure of unconstitutional partisanship.").

truly random sample of districting plans (set partitions).[106] Moreover, it is unlikely that such an algorithm exists, as its presence would imply that NP-hard decision problems could be solved using randomization, which is widely believed to be false.

Instead, current criteria based on "random" samples of districts use heuristics similar to the automated redistricting algorithms discussed above. Heuristics provide no formal guarantees of outcomes' characteristics and are always presumably biased toward some sort of outcome. Although the direction of the bias may be impossible to determine, this is more than a theoretical problem. Several recently published studies have drawn unsupportable inferences about the characteristics of redistricting plans based on random districting.

A study of South Carolina concludes that the state was racially gerrymandered in the 1990s because their "random" redistricting heuristics failed to produce plans that yielded the same number of minority seats as the actual plan.[107] This conclusion is not a valid statistical inference, as there is no reason to believe that the "randomized" method produced a set of plans that constituted an unbiased random sample of possible compact, contiguous, equal-population districting plans.[108] Further, the analysis rejects only the hypothesis that the actual plan is not a plan drawn based solely on traditional districting criteria. It does not reject other plausible hypotheses that are different from gerrymandering—for example, that the actual plan is a random sample of plans based on other legitimate non-racial factors, such as moderate partisanship, communities of

---

106. Our lengthy search of the computer science literature revealed a single method for producing partitions of sets with a known (in this case, uniform) random sampling distribution. These methods were invented over twenty years ago, *see generally* ALBERT NIJENHUIS & HERBERT S. WILF, COMBINATORIAL ALGORITHMS (1975), but are not well known. In theory, it is trivial to adapt these methods to sample redistricting plans using the rejection sampling method—sampling partitions at random, rejecting any that do not constitute feasible redistricting plans (because of violation of contiguity, population constraints, etc.) and using the remainder to estimate the distribution of redistricting plans. Unfortunately, in practice this approach is computationally intractable, since as the number of blocs in the plan gets larger the ratio of feasible plans to random partitions grows exponentially smaller, and the execution time for calculating the distribution continues to grow.

107. *See generally* Cirincione et al., *supra* note 6.

108. Although bias is an inevitable result of the use of heuristics, in this case the direction of bias in small samples can be readily demonstrated. The appendix shows how enumeration of all feasible redistricting plans on a 3x2 grid demonstrates how this heuristic rule biases towards the creation of compact districts, even though compactness is not a stated criterion of the heuristic.

interest, and protection of incumbents. Thus the analysis makes an unwarranted causal inference that gerrymandering caused the discrepancy.

A study of Japanese districting concludes that the Japanese Diet was gerrymandered because the districting plan differed in characteristics from those created by their heuristic.[109] This analysis makes errors in statistical and causal inference identical to the South Carolina study.

A study of the United States House of Representatives draws inferences about the effect of redistricting on congressional polarization from a set of districts created by randomly allocating split-county units to districts without replacement.[110] Split-counties are formed by dividing counties into whole numbers of 10,000 person blocks and discarding remainders. Although this is a purely random sample, and hence unbiased, it does not sample from the population of legal districting plans. Instead it includes non-contiguous plans and plans that are of unequal population and excludes legal plans created from census tracts and blocks. The authors are aware of this problem and use ad-hoc adjustments to this method to attempt to correct it, but the amount and direction of bias and how it affects their conclusions are indeterminate.

One of the most prominent computer science textbooks warns against a similar assumption when designing random number generators.[111] Another textbook cautions quite specifically, "Generating random permutations [and other combinatoric objects] is an important little problem that people stumble upon and often botch up. . . you must be very careful with random [combinatorial] generation. We recommend that you try reasonably extensive experiments with any random generator before believing it."[112] In other words, it is trivial to create an algorithm that does "'random things,"' but it is risky to assume that such arbitrary behavior produces statistically random results.

---

109. Sakaguchi & Wada, *supra* note 78.

110. *See generally* Nolan M. McCarthy, Keith Poole & Howard Rosenthal, *Does Gerrymandering Cause Polarization?*, 53 AM. J. POL. SCI. 666 (2009).

111. DONALD E. KNUTH, 2 THE ART OF COMPUTER PROGRAMMING: SEMINUMERICAL ALGORITHMS 4–6 (3d ed. 1997).

112. STEVE SKIENA, THE ALGORITHM DESIGN MANUAL 248 (1998).

VI. OPEN REDISTRICTING—THE NEXT WAVE

The next major change in the use of computers in redistricting will be centered on systems that enable the public to participate in the map drawing process. If the potential of automated redistricting and its variants cannot be achieved due to theoretical and computational constraints, perhaps computing advances can be leveraged to implement more modest goals.

The idea that software, and mapping software in particular, can enhance public participation dates back to at least the 1960s, as previously discussed. In the last decade, however, a number of geographers have begun to crystallize the notion of "participatory" GIS systems, and to study "public participation GIS" more systematically.[113] In the last several years, Google's release of the Google Map service and related advances in online mapping have led to more interest in what is now more colloquially known as "collaborative" or "social" mapping.

Limited attempts at participatory redistricting pre-date the widespread use of GIS systems. In the 1990s, some states attempted to mitigate barriers to public participation in redistricting by providing a public computer terminal, typically located at a state library or state office, loaded with the necessary software and data to draw districts.[114] However, it was not until shortly after 2000 that GIS systems became sufficiently inexpensive to contemplate their large-scale public use in redistricting.[115]

Of the eighteen states that reported providing public terminals in 2000, Arizona illustrates the potential benefits that public participation in the redistricting process provide, as well as the barriers that may exist. Although Arizona provided a public terminal, only two well-organized interest groups presented complete congressional plans to the Arizona Independent Redistricting Commission: the Coalition for Fair Redistricting—an alliance of minority groups—and Democratic state legislative members, lead by

113. *See generally* Renee Sieber, *Public Participation Geographic Information Systems: A Literature Review and Framework*, 96 ANNALS OF THE ASS'N OF AM. GEOGRAPHERS 491 (2006).
114. Altman et al., *supra* note 16.
115. *See id.*

Democratic Congressman Ed Pastor.[116] Still, these plans, along with testimony from other members of the public, had an impact on Arizona's adopted redistricting plan. Perhaps the most apparent effect was observed in the First Congressional district, where Arizona's northwest communities are connected with the Hopi Reservation in the northeast corner of the state by a narrow neck extending through the Grand Canyon. This oddly-shaped district was drawn at the urging of representatives of the Hopi tribe, who did not wish to have the same representation as their traditional enemies (and the more numerous) Apache, who live in a reservation that surrounds the Hopi.

In contrast to the very small number of maps created by the public in the last round of redistricting, a simplified "redistricting game" probably has led to more people drawing (hypothetical) redistricting maps in the last three years than in the entire history of the nation. Released in 2007 by creator Chris Swain of the USC Game Innovation Lab, this redistricting system is interested in redistricting if provided with a user-friendly means to participate.[117]

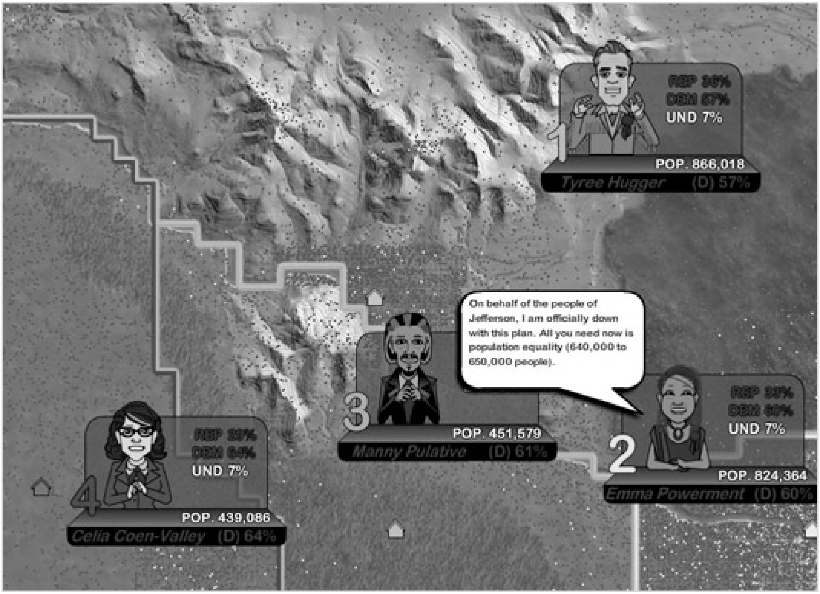
---

116. Reporter's Transcript of Proceedings for Public Session on July 17, 2001, Arizona Independent Redistricting Commission, http://www.azredistricting.org/Meetings/PDF/AIRCTranscriptsPublicSession7-17-01.pdf (last visited Apr. 14, 2010); *see* Arizona Independent Redistricting Commission, Meetings & Transcripts, http://www.azredistricting.org/?page=meetings (last visited Apr. 14, 2010) (providing the comprehensive transcripts of all AIRC meetings).

117. *See* Michael Falcone, *A Gamers Guide to Redistricting*, THE CAUCUS, June 14, 2007, http://thecaucus.blogs.nytimes.com/2007/06/14/a-gamers-guide-to-redistricting/.

Figure 4

Screen from the "Redistricting Game", by Chris Swain



A recent Ohio mapping exercise illustrates the potential of adapting this approach to real redistricting plans. In the summer of 2009, the Ohio Secretary of State's office found that there may be greater public interest in mapping redistricting plans than the Arizona experience indicates. The Secretary of State's office held a redistricting competition, inviting the public to draw and submit congressional plans evaluated on several criteria based on 2000 census data.[118] Three of the fourteen submitted maps—two from Ohio citizens and one from an Illinois state legislator and physics professor named Mike Fortner—were declared "'winning'" plans in that they satisfied threshold criteria values.[119] Although three plans were rejected because they did not have an African American majority district, all fourteen scored as politically fairer (under the Ohio

118. *See* Ohio Secretary of State, Ohio Redistricting Competition, http://www.sos.state.oh.us/SOS/Text.aspx?page=12303 (last visited Apr. 14, 2010) (explaining the Ohio Secretary of State's redistricting competition).

119. *Id.*

Secretary of State's measure) than the bipartisan congressional plan adopted by the state legislature.

Compared to the redistricting game, the Ohio competition attracted few participants. Yet the set of plans it created suggest that by opening up the redistricting process more broadly to encourage citizen participation, it is possible to discover a menu of legal redistricting plans that are "fair" by nominally objective standards advocated by various proponents of reform.[120] By disseminating these alternative plans, redistricting authorities may change their behavior by using suggested district configurations to aid in fashioning their own plans, or by reacting to media scrutiny shaming them into producing a better plan than they might have otherwise.

Scholarly study of public participation GIS systems suggests that it is useful to characterize participation as having different levels of involvement and influence, ranging from observation to joint public control.[121] The last round of redistricting saw some increase in observation in the lowest level of involvement, as measured by "hits" on state redistricting websites disseminating data and other information. While we do not expect to see collaborative redistricting systems that give the public joint control over the process in the near future, we do believe that the level and quality of participation can be "upgraded." Using collaborative redistricting systems, the public can take an active role in the public dialogues and deliberations over redistricting.

We envision that with an open online redistricting system, public interest groups could draw and actively lobby for their maps, and courts would no longer need to choose between only the maps offered to them by the political parties during litigation. Academic scholars and expert witnesses could use the program to explore hypothetical scenarios to test the motivations and outcomes of redistricting. Perhaps even politicians would be interested in the software as a means to remove "politics" from the process. However, on this last point, we should not be blind to the prospect of politicians using the program to help maximize their political goals.

---

120. Altman & McDonald, *supra* note 21.
121. Sieber, *supra* note 113, at 500–01

VII. THE NEED FOR TRANSPARENCY IN "OPEN" SYSTEMS

Wherever political systems are readily manipulated, it is better to have ten thousand eyes watching than twenty. The drawing of electoral districts is among the most frequently manipulated and least transparent processes in democratic governance. All too often redistricting authorities maintain their monopoly on the process by imposing high costs to public participation. By providing the public with information similar to that held by official decisionmakers, increasing transparency and public participation can be a powerful counterbalance to this monopoly on redistricting districts. This information can lead to different outcomes and better representation.

The redistricting process does not need to be closed to the public. Transparency can be facilitated through greater access to resources by interested outsiders, and redistricting authorities can encourage public participation through formal requests for public comment and the consideration of publicly submitted redistricting plans. The optimal redistricting plan may not be discoverable, but open redistricting may produce a "better" outcome through the exploration of a menu of redistricting plans generated by the public (or automated methods) and through robust public debate over these plans' goals. In this manner, the adopted plan may not necessarily be the one created behind closed doors that embodies only the preferences of political professionals.

Increasing openness has many direct benefits. Openness can aid in educating the public about the electoral process, empowering them to participate in a process that engages them in shaping the representation of their neighborhoods and communities, permiting them to show legislators plans they support, and promoting broad commentary and discussion.

Fostering public participation enables the public to identify their neighborhoods and communities, promotes the creation of alternative plans, and facilitates an exploration of a wide range of representational possibilities. Publicly-drawn plans can provide a measuring stick against which an official plan may be compared, and promote the creation of a "market" for plans that support political fairness and the representational goals of the community.

VIII. RECOMMENDATIONS FOR TRANSPARENCY

Complex systems are inherently difficult to make transparent. As explained above, automated redistricting systems often give the appearance of clarity while embedding bias in technical details. Similarly, web-based redistricting software, if not thoughtfully implemented and deployed, could create impressions of accessibility and transparency that are merely illusory. Software and systems that are used to analyze or generate plans, or reports based on that software, can be impossible to reproduce or correctly interpret without access to the code used to generate them. Thus, transparency requires that:

- Software used to automatically create or improve redistricting plans that are either open-source or provide documentation sufficient for a third party to replicate the results using independently developed software.
- Reports analyzing redistricting plans include documentation of data, methods, and procedures sufficient to allow a third party to verify the report.

It is becoming widely recognized that software transparency is required for scientific replication purposes,[122] trusted systems,[123] and the transparency of government actions relying heavily on computer systems, such as electronic voting.[124] Because the technical definitions of redistricting criteria are so complex and varied, it is critical to document the exact algorithms used when determining contiguity, measuring compactness, "optimizing" districts, and measuring competitiveness. Because the goals of redistricting evolve with our understanding of representation, it is critical to allow the public to modify the tools used for redistricting in order to better adapt them to their own views of representation.

122. *See generally* MICAH ALTMAN, JEFF GILL & MICHAEL P. MCDONALD, NUMERICAL ISSUES IN STATISTICAL COMPUTING FOR THE SOCIAL SCIENTIST (2003).
123. *See generally* Rebecca T. Mercuri, *Trusting in Transparency*, 48 COMM. OF THE ACM 15 (2005).
124. LAWRENCE LESSIG, CODE: VERSION 2.0 141–43 (2006).

Without the ability to inspect redistricting software, it will be impossible to verify that algorithms for creating or evaluating plans are implemented correctly. Without a license to modify and reuse the software, the system can only accommodate a limited set of representational goals. Without access to previous versions of the code, it will be impossible to replicate earlier versions of the code. Distribution of code under an open source license through a well-known repository, such as Sourceforge, solves these problems: an open source license establishes the rights to inspect the code and to build new tools with it and the repository captures the history of changes to the system.

Still, without data, software for creating and analyzing plans is useless. Yet state redistricting authorities can, with modest effort, make available online a wealth of information that would substantially lower the barriers to public participation in a transparent manner. It is technically feasible for states to maintain websites that provide access to redistricting data, proposed redistricting plans, and software that enables the public to create their own community maps and entire redistricting plans.

Some data is already readily available. The population data used for redistricting purposes is known as the PL94-171 file—named after the federal public law mandating its release. The Census Bureau provides online access to the geographic and population information necessary to use this data for redistricting purposes. In some instances, however, redistricting authorities may further adjust or enhance these geographic and population data sets. For example, state law could require that geographic entities not found in the Census Bureau' be respected during the redistricting process, such as townships or communities of interest. Alternatively, a state may adjust population counts by excluding military, students, or prisoners from their calculations. Most states except those with specific prohibitions enhance their population data by merging it with election returns so probable election outcomes may be forecasted.

Creating and evaluating redistricting plans requires access not only to demographic data but to community, and in many cases, electoral data, as well. The public needs equal access to this data to participate in plan creation and to verify claims made about official plans. To determine what is necessary to guarantee transparency, we need only look to the science,[125] in which replication and verification are fundamental requirements. Thus, all data needed to create legal redistricting plans should be distributed for public use, under a license allowing reuse of the data for non-commercial purposes. Further, to ensure that data have not been intentionally or unintentionally modified in such a manner as to affect the redistricting outcomes, all data should provide clear documentation of the original source, the chain of ownership (provenance), and all modifications made to it.

The public needs easy access to the redistricting plans generated by a redistricting authority or submitted by the public in order to evaluate these plans. Redistricting plans released as images of districts or by describing how districts' borders align with existing features—known as "'metes and bounds"'—are difficult to import into mapping software. Full transparency thus requires that redistricting plans be made available in non-proprietary formats that are easily read into commonly used GIS systems.

Finally, on the Internet, as the joke goes, "no one knows you're a dog." Websites and online service must be transparent to be trustworthy. Services offered to the public to evaluate or create redistricting plans are opaque and subject to misinterpretation unless adequately documented. Transparency requires that such sites and services provide a clear privacy policy; offer users the ability to publish plans and make available all published plans in non-proprietary, machine analyzable formats; and provide documentation of any organizations providing significant contributions to the operation of the service.

IX. PROMOTING PARTICIPATION

Even the most transparent redistricting computer system will not curb gerrymandering if it is not used. Surveys consistently find that the public has little knowledge about the most basic components of the redistricting process.[126] For a redistricting system to encourage

---

125. *See, e.g.,* Nature Staff, *Special: Data Sharing*, 461 NATURE 145 (2009); COMMISSION ON BEHAVIORAL AND SOCIAL SCIENCES AND EDUCATION, SHARING RESEARCH DATA (Stephen E. Fienberg, Margaret E. Martin & Miron L. Straf eds., 1985).

126. *See, e.g.*, Press Release, Pew Research Center for the People & The Press, Most Have Heard Little or Nothing About the Redistricting Debate: Lack of Competition in Elections Fails to Stir Public (Oct. 27, 2006) (finding that eighty-nine percent of voters had heard little or

broad participation it must overcome this information barrier. It must be accessible, easy to use, and be accompanied by sufficient training materials and evaluation tools to help novices draw legal redistricting plans. It may be that these tools will be insufficient for the masses to draw redistricting plans, but even the participation by a few people may illuminate different approaches to drawing districts within in a state.

Social networking has blossomed over the past decade, and we suspect that this technology can be applied readily to redistricting. Map drawers can work collaboratively to improve redistricting plans that make sense for their communities. For example, a rural resident might start drawing their redistricting plan by importing an urban district that the NAACP believes is in compliance with the Voting Rights Act. Redistricting authorities and other map-drawers among the public may benefit from public commentary on various proposed redistricting plans in order to better understand what may or may not work in a given community.

CONCLUSION

Computers have decreased the costs of redistricting, but they are a means and not an end to themselves. By virtue of their ability to quickly sort through large amounts of data, computers permit the exploration of a greater number of alternative configurations of districts within the short period of time between the census and the next election. Despite this technical innovation, however, the motivations of professional political actors who currently control the process in most states remain the same and have at best only marginally been more easily realized by the advent of computerized redistricting.

The common goal of those who advocate for automated methods or use of indicia as to constrain the motivations of professionals have monopolized their power by keeping the process opaque. We are skeptical that technical and philosophical practicalities limit the application of these methods. We believe, however, that these methods can be harnessed to produce a more modest and achievable outcome. Computers can transform redistricting by encouraging the

---

nothing about redistricting, and forty-seven percent of voters did not who was in charge of the process).

production of a broad variety of legal redistricting plans, and by encouraging a robust public discussion over the inherent tradeoffs these plans embody.

The primary effect of these efforts is probably *not* going to be immediate or ignite aggressive reform of the redistricting process. Rather, the existence of publicly-drawn plans empowers citizens to participate in a process that is generally closed, provides a measuring stick against which an official plan may be compared, and establishes the existence of a "market" for plans that support political fairness and community representational goals. Redistricting outputs may change where a redistricting authority is responsive to citizen submitted maps or, where partisan gridlock necessitates court involvement. Judges may look more favorably on plans drawn by citizens than those drawn by redistricting authorities or politicians. Reform may follow if a plan adopted by a redistricting authority is not as strong as that demanded by the public, which now has both knowledge of and a vested interest in the redistricting process.

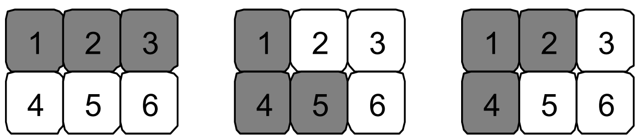
APPENDIX: PROOF OF BIAS IN THE CDR METHOD

A simple example suffices to prove the basic "computationally-intensive" method used by Cirincione, Darling, and O'Rourke (CDR) for drawing contiguous districts does not produce a representative sample of the population of districting plans. Consider a simple "state" which is composed of six identically sized and populated block groups on a 2x3 grid. Each block group is contiguous with its horizontal and vertical neighbors.[127] We divide the state into two districts, each containing exactly three blocks and we fully enumerate the possible districting plans and calculate the probability the CDR algorithm will find each solution. The algorithm is biased if all solutions do not have the same probability of discovery.

In this hypothetical state, there are three possible districting plans (assuming that the numbering of districts is unimportant, but this assumption does not fundamentally affect our conclusions), as shown below:

Figure 1

Feasible contiguous redistricting plans on a 2x3 grid.



The contiguous district generation algorithm used by CDR is very simple, and thus amenable to formal analysis. As they describe it:

> The first algorithm, the *contiguity algorithm*, begins by randomly selecting a block group to serve as the "base" of the first district. It then constructs a "perimeter list" containing the unassigned block groups contiguous to the base block group. The program then randomly selects a block group from the perimeter list to add to

127. By constructing the geography with "holes" so that no population units meet at a single point, we avoid the issue of whether to treat such units as contiguous. This simplification is taken only to clarify the exposition. In fact, decisions about how to measure contiguity (and other criteria) are completely independent from the sampling behavior of the algorithm. Our example only requires that contiguous relationships not be impossible, *a priori*.

> the emerging district and adjusts the perimeter lists. The process continues until the population of the emerging district achieves the desired population level . . . . The next district begins with the random selection of a census block group from among those that touch one of the complete districts. [And the process continues until a legal plan is generated, or until no more legal districts can be created, in which case the process is restarted.][128]

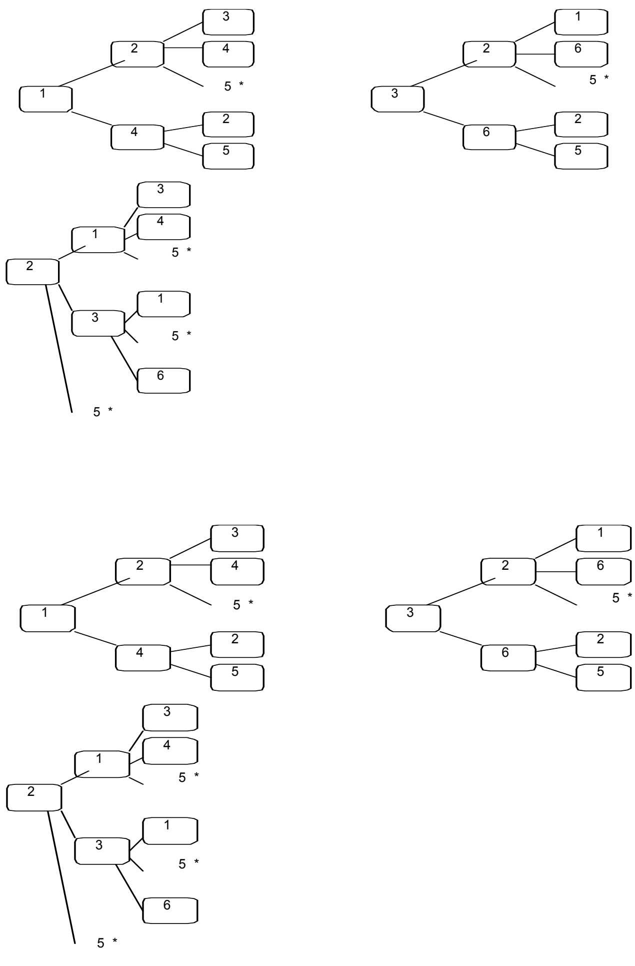
In a true random sample of contiguous districting plans, the probability of the method generating each plan would be one-third. What is the probability of generating each districting plan using CDR's algorithm? Using the computationally intensive "sampling" method, the probability is lower for {1,2,3} than for the other two plans.[129] The tree below shows all of the possible sequences of choices starting from the base block groups 1,2,3 (the paths from the bases 4,5,6 are symmetric):

128. Cirincione et al., *supra* note 6, at 196.

129. At first glance, one might think this bias is intended to select "compact" districts. In fact, this method, as developed by Cirincione et al., is supposed, to select from the universe of contiguous plans. They use a different algorithm to select compact plans.

Figure 2

Event-trees showing the generation of the district plans in Figure 2. Each sub-tree is equally likely (P=1/6), and the probability of following any branch at each node is equal to 1/(number of branches). Starred nodes indicate illegal plans, which cause the algorithm to restart with subtree selection.



The total probability of generating plan {1,2,3} during a single run of CDR's algorithm is the probability of the sum of the probabilities of the paths

{{1,2,3}, {3,2,1}, {2,1,3}, {2,3,1}},

multiplied by two (for symmetry with starting points (4,5,6)). This equals

(1/36+1/36+1/54+1/54)*2=5/27.

The probability of generating plan {1,4,5} (which is the same as the probability of {1,2,4} by symmetry) is the sum of the probabilities of the paths

{{1,2,4},{1,4,2},{2,1,4},{3,5,6}}*2,

which equals

(1/36+1/24+1/54+1/24)*2=7/27.

(The probability of having to start over is 8/27, but this does not affect the asymmetry of accepting each plan in later rounds.) Thus, the basic algorithm used by CDR is statistically biased, in small samples.

CDR offers no evidence at all that their method is asymptotically unbiased. And note that unlike sampling techniques based on statistical theory, the heuristic techniques used by CDR and others carry with them no guarantees of unbiasedness or any other asymptotic properties.