On empirical validation of compactness measures for electoral redistricting and its significance for application of models in the social sciences (Extended)*

Christine Chou
National Dong Hwa University
Department of International Business
Hualien 97401,Taiwan
chou@mail.ndhu.edu.tw

Steven O. Kimbrough
University of Pennsylvania
Operations and Information Management
Philadelphia, PA 19104, USA
kimbrough@wharton.upenn.edu

Frederic H. Murphy
Temple University
Marketing and Supply Chain Management
Philadelphia, PA 19122, USA
fmurphy@temple.edu

John Sullivan-Fedock
University of Pennsylvania
Operations and Information Management
Philadelphia, PA 19104, USA
johnsul@design.upenn.edu

C. Jason Woodard
Singapore Management University
School of Information Systems
Singapore 178902
jwoodard@smu.edu.sg

June 5, 2014

Abstract

Use of optimization models in science and policy applications is often problematic because the best available models are very inaccurate representations of the originating problems. Such is the case with electoral districting models, for which there exist no generally accepted measures of compactness, in spite of many proposals and much analytical study. This paper reports on an experimental investigation of subjective judgments of compactness for electoral districts. The experiment draws on a unique database of 116 distinct, legally valid districting plans for the Philadelphia City Council, discovered with evolutionary computation. Subjects in the experiment displayed, in the aggregate, remarkable agreement with several standard measures of compactness, thus providing warrant for use of these measures that has heretofore been unavailable. The exercise also lends support to the underlying methodology on display here, which proposes to use models based on subjective judgments in combination with algorithms that find multiple solutions in order to support application of optimization models in contexts in which they are only very approximate representations.

Keywords: systems design, evolutionary design, emotive design, genetic algorithms, evolutionary computing, districting, zone design, compactness, redistricting, reapportionment, interactive evolutionary computing

1 Introduction

Advances in articulating and solving constrained optimization models often lead to progress in science and progress in solving real-world design problems. Better algorithms for finding optimal solutions, however, can

*This is a slightly extended version of (Chou et al., 2013), containing archival and other material excluded from the *SSCR* version for lack of space.

do little to address real-world problems if the models themselves are very inaccurate representations of the originating problems. In the face of model inaccuracies we have basically three choices, each of which may be appropriate depending upon circumstances:

1. Do the best we can in building the model, obtain an optimal solution for it, act on an optimal solution, and hope for the best.

2. Insist on an accurate model, then wait to decide until one is available and has a solution.

3. Use the admittedly flawed model to support discussion and deliberation, implementing a modification of an optimal solution to the model in response to considerations not fully represented in the model.

Recent work, undertaken in the context of the design of electoral districts, has led to a promising technique in conjunction with the third option, aimed at making the option more attractive in cases where the proposal can be applied (Chou et al., 2012; Gopalan et al., 2013, 2012a). The proposal is designed for cases in which it is problematic to define and parameterize an objective function mathematically, yet human subjective judgments can distinguish better and worse solutions. Such cases are, perhaps surprisingly, common, as the extensive literature in so-called *interactive evolutionary computation (IEC)* attests. (The IEC idea originated in *The Blind Watchmaker*, (Dawkins, 1985). See (Chou et al., 2012; Takagi, 2001) for overview and references on subsequent developments.) Not surprisingly, the downside of relying on human subjective judgments for selecting the good solutions from a large pool of solutions is human fatigue and consequent failures of discrimination. Takagi and Iba (2005) call this "the most pressing problem in IEC systems."

As a partial remedy to this problem, Chou et al. (2012) propose what they call *validated surrogate fitness (VSF)* functions. The nub of the idea is simple: (a) use human subjective judgment to validate a mathematical function that approximates the choices humans make when confronted with the pool of solutions to be considered; (b) use the mathematical function (as the VSF) during automated search (for optimal solutions); (c) produce by automated search a plurality of good solutions (according to the model); and (d) offer these good solutions as members of a consideration set for subsequent deliberation. (Points (c) and (d) instantiate the more general concept of *solution pluralism* (Kimbrough et al., 2011).)

While we agree that the VSF idea ((a)–(d), sketched above) has much to recommend it, the supporting evidence presented in (Chou et al., 2012) was rather modest, as we shall see. The purpose of this research note is to report on an experiment that augments the unusual, even unique, data set produced by (Chou et al., 2012) and investigates the core idea more thoroughly. §2 succinctly presents necessary background. §3 then presents the experiment and discusses the results. We conclude in §4 with some brief comments.

2 Essential Background

Our context is the design of electoral districts. In the United States, after each census it is usually necessary to redraw the boundaries of electoral districts in order to maintain approximately equal populations among them. The process is called (re)districting or reapportionment. It is fraught—contentious and consequential—because there are often many legally valid ways to draw the boundaries of the districts, but different designs will favor different interests. Attracting considerable public attention of late, e.g.,(Draper, 2012; Wang, 2013a,b), the universal worry is that when it comes to reapportionment those in power will normally attend to their own interests ahead of the public's. The question, "And will elected politicians use redistricting to pick their voters, instead of the other way around?" (Altman and Klass, 2005), voices an often-repeated worry that advances in information technology will serve to work against the public interest. The literature is enormous; see, e.g., (Altman et al., 2005; Cain, 1984; Cox and Katz, 2002; Turner and Lamacchia, 1999) for useful reviews.

Electoral districts that are oddly shaped for the purpose of benefitting some party, cause, or even individual (a common occurrence) are said to be *gerrymandered.*[1] There is widespread agreement that gerrymandering is not in the public's interest. Among other things, this has led to efforts to get the public involved by encouraging the development and publicizing of districting proposals by the general public in hopes that putting better plans in play will lead to better outcomes, e.g., (Satullo, 2011).

The antidote to interest-based designs, as in gerrymandering, is generally seen to be an objective, disinterested basis for designing electoral districts. But how? Gerrymandered districts are often apparent to the eye because they sprawl and meander in seemingly arbitrary ways. They are not tight and tidy. To use the common term, they are not *compact*. The received hope has been to find an objective measure of compactness, which may then be used to assess districting plans (Cox and Katz, 2002; Niemi et al., 1990; Young, 1988) (see also (Fryer, Jr. and Holden, 2011, pages 497–8), (Altman and McDonald, 2012), and (Niemi et al., 1990, pages 1156–7), which summarize court sympathies with using some measure of compactness).[2] Points arising:

1. If an objectively assessable measure of compactness could be agreed upon as the objective in reapportionment, then it would be possible in principle to conduct search and optimization by computer to find good designs. One would minimize the compactness measure (assume smaller is better), subject to legal constraints on contiguity and population size ((Cox and Katz, 2002; Kalcsics et al., 2005; Young, 1988); see (Gopalan et al., 2013; Murphy et al., 2012) and references therein for background on the problem in the history of operations research).

2. There is in fact no general agreement regarding measures of compactness for reapportionment. Young (1988) argues in detail that all standard measures of compactness have serious flaws, and hence there can be no mathematical or a priori basis for choosing among them. In a follow-on study, Niemi et al. (1990) support and extend Young's findings. They conclude that combining multiple measures of compactness should be explored as a possible solution. But they do not develop this idea and, to our knowledge, it remains only a suggestion. In short, Young's views have been widely accepted. As Fryer, Jr. and Holden (2011, page 494) note, "There is no consensus on how to adequately measure compactness."[3]

3. Even if agreement on the definition of compactness for reapportionment purposes could be had, limiting the objective of the optimization to compactness is problematic, since other objectives, such as protection of neighborhood integrity (Fryer, Jr. and Holden, 2011; Satullo, 2011), are often important.

This last point is addressed in a principled way by point (d) in the VSF concept (above). The first point offers a hope if points 2 and 3 can be resolved favorably. So, we are left with addressing point 2: the problem of finding a measure of compactness that can be used in automated search and optimization and that is sufficiently warranted to serve as a surrogate for human judgment.

Given Young's analyses (Young, 1988), the justification for a compactness measure will have to be an empirical one, yet, there is very little such work available. This takes us to the paper by Chou et al. (2012). Further points arising:

4. Besides proposing the VSF concept, Chou et al. (2012) succeeded in the technically challenging task of producing 116 good solutions for a difficult districting problem (point (c), above).[4] Every design was required to allocate each of 66 wards to one of ten councilmanic districts for the City of Philadelphia, using data from the 2010 census. In consequence, the raw search space, assuming assignment of at least one ward to each of the ten districts, is sized as a Stirling number of the second kind, $S(66, 10) \approx 2.7295 \times 10^{59}$. They used evolutionary computation to produce this pool of legally feasible solutions by minimizing on a measure of compactness, subject to feasibility on contiguity and population size.

5. The measure of compactness used by Chou et al. (2012) was the *max-max distance*. Their optimization sought to minimize the maximum of the largest intra-district distance within a plan. This is very much a non-standard measure of compactness, use of which was necessitated by the information available at the time.

6. Further, Chou et al. (2012) undertook three experiments that gave qualified support from human judgments for use of this compactness measure as a VSF in the context of electoral districting. Using paired comparisons of district plans that were chosen to be well separated in terms of the max-max distance measure, they were able to show, across 3 experiments, a discernible level of agreement between comparative compactness judgments of the subjects and the verdicts of the max-max distance metric.

We conjecture that a main reason why careful empirical studies have not heretofore been done on measures of compactness for electoral districts is that there have not been libraries of alternative plans available for comparison. In fact, reporting of the production of the 116 plans, and how this was achieved, was, with the VSF idea, the main thrust of (Chou et al., 2012). Given their quite limited and provisional empirical results—items (5) and (6) just above—it becomes important to test the VSF idea on standard measures of compactness. That is just what our experiment was designed to do.

3 Experiment 4

Chou et al. (2012) report three experiments on human judgment of compactness, using maps (district plans) drawn from the 116 produced by their evolutionary algorithm. Their experiment 2 is not directly comparable to our experiment and so we will not discuss it further. In their first and third experiments, the same four pairs of district plans were presented to subjects who were asked to yield judgments of comparative compactness. In terms of table 1, subjects were asked to choose between maps (1 & 2), (3 & 4), (5 & 6), and (7 & 8) with regard to compactness. Most importantly, subjects were asked to choose on the basis of what they judged most *other* people would think with regard to the comparative compactness. Further, subjects were rewarded based on the number of choices they made that agreed with the majority. This is called a *beauty contest* design; it provides incentive for the subjects to report honestly and thoughtfully. Experiment 1 was performed in a laboratory setting; experiment 3 was conducted on Amazon Mechanical Turk (AMT). The results of the two experiments were in good agreement.

Our experiment 4 extended experiment 3 on AMT. We used the same basic setup with these modifications.

1. Subjects were asked to choose between 8 pairs of maps (numbered 1–16), instead of 4, as in experiment 3 (and experiment 1).

2. Maps 1–8 (and pairings) were as in experiment 3 (and experiment 1). Maps 9–16 were paired in the same pattern, now: (9 & 10), (11 & 12), (13 & 14), and (15 & 16).[5]

3. Maps 9–16 were randomly drawn anew. Note especially that, unlike maps 1–8, they were *not* selected to be far apart in objective values. (All 16 maps were drawn without replacement.)

4. Subjects were paid \$0.20 to complete the task, with a bonus of \$2.00 for 7 or 8 agreements with the group, and \$1.00 for 5 or 6.

5. Maps 1 and 2 were presented paired in random order, maps 3 and 4 were presented paired in random order, etc.

6. We limited participation to 200 subjects, each of whom recorded judgments on the 8 pairs of maps. Of these subjects, 20 made 8 judgments that agreed with the majority vote in each case, 38 judged correctly in 7 of the 8 cases, 30 got 6 of the 8 right, and 22 got 5 of 8 right.

Table 1 presents the key data from the experiment, along with other information, including the results of experiments 1 and 3. Points arising:

1. The column labeled "Exp. 4" presents the votes received by individual maps in the pair comparisons. The column labeled "Objective" records the max-max distance score of the associated map. Recall that lower scores are (presumably or hypothetically) associated with better compactness, where the max-max distance consists of finding the largest distance between two points in a district and finding the maximum of those distances over all districts. The votes are all in the "right" direction and consistent with the results of experiments 1 and 3. The p-values are significant in each case: 0.0280 (map 1–map 2 pairing), 7.0085e-04 (maps 3–4), 4.1640e-07 (maps 5–6), and 2.4865e-05 (maps 7–8).

2. The votes are *not* in the "right" direction for the max-max distance measure of compactness for the pairings maps 9–10, maps 11–12, and maps 15–16. The voting differences, however, *are* significant with p-values of 0.0097 (maps 9–10), 0.0018 (maps 11–12), 7.5353e-09 (maps 13–14), and 0.0141 (maps 15–16). Note that in these cases the objective values of the compared maps are much closer together than in the earlier comparisons. This suggests that max-max distance is not salient or "right" for subjects in judging compactness when the compared values are too close together.

Map	Exp. 1	Exp. 3	Exp. 4	Totals	Objective	Avg. Schwartzberg	Distance^2	Perimeter
1	19	83	114	216	341.62	1.678	138.8407	220.03
2	0	54	86	140	399.34	1.845	176.3282	240.28
3	17	73	123	213	346.57	1.761	144.4725	231.35
4	3	64	77	144	399.34	1.840	179.9436	239.65
5	4	58	65	127	399.34	1.847	179.1198	240.50
6	15	79	135	229	346.57	1.666	136.0053	218.68
7	13	86	129	228	346.57	1.677	137.2534	219.89
8	7	51	71	129	391.26	1.709	153.2877	224.81
9			117	117	377.46	1.645	134.7425	216.23
10			83	83	365.49	1.706	136.2504	224.71
11			79	79	369.43	1.831	170.4763	238.10
12			121	121	378.46	1.669	136.1410	219.48
13			140	140	367.58	1.757	146.2849	230.66
14			60	60	399.34	1.894	191.8305	246.67
15			84	84	355.37	1.787	165.1738	230.90
16			116	116	377.46	1.705	136.6441	223.90

Table 1: Summary of results, experiments 1, 3, and 4. (Perimeters given in miles.)

3. Using manual tracing, assisted by computer, we obtained the perimeters and areas of each of the 10 districts in each of the 16 maps. The perimeter values are reported in table 1 in the column labeled "Perimeter." Both the perimeter and the area values were used in computing the Schwartzberg measure, shown in the column labeled "Avg. Schwartzberg." Thus, we augmented the data available in (Chou et al., 2012).

4. The Schwartzberg measure of compactness is the ratio of the perimeter of a district divided by the perimeter of the circle having the same area as the district. Lower values indicate more compact districts on this measure. In the table, in the column labeled "Avg. Schwartzberg," we report the sum of the Schwartzberg measures for each district in the corresponding map, divided by the number of districts (10 in the case of Philadelphia).

5. There is perfect agreement between the Schwartzberg measure and the majority voting results in experiments 1, 3, and 4.

6. There is perfect agreement between the Perimeter measure and the majority voting in experiments 1, 3, and 4 (and with Schwartzberg).

7. The column labeled "Distance^2" in table 1 records the (square of the) total Euclidean distances in each design/map between the centroids of the districts and the centers of their constituent wards. This is a commonly-used compactness metric (often called the *center of gravity* measure), with smaller numbers indicating better values. (This information was in fact available at the time of writing of (Chou et al., 2012), but was not used.)

8. There is perfect agreement between the Distance^2 measure and the majority voting in experiments 1, 3, and 4 (and with Schwartzberg and Perimeter). Of course, Distance itself would preserve the agreement as well. The (Pearson) correlation coefficient between the average Schwartzberg measure and Distance (square root of Distance^2) is 0.9570, suggesting that while these are indeed different measures, they will be largely equivalent for practical purposes.

9. On the basis of these results we find support for the hypotheses that the Schwartzberg measure, the Perimeter measure, and the Distance measure can be used as VSFs in districting because they accord well with human subject judgment. Of course, we have no data to rule out possible measures of compactness, other than max-max distance.

10. We find it plausible that human subjects are sensing and judging on the basis of the Perimeter measure, rather than Schwartzberg, which we suspect is harder to sense and intuit. This, however, is only a

conjecture and will need to be tested experimentally. Similarly, whether subjects rely on Distance or Perimeter (or neither or a combination) remains to be tested.

11. Interestingly, the favorite map of Chou et al. (2012), shown in figure 2 of (Chou et al., 2012), has a Distance^2 score of 132.2226, which is the lowest of the 116 maps we discovered. This is a fact we encountered after Chou et al. (2012) selected the map as the best of the bunch.

12. Figure 1 shows gray scale versions of maps 9 and 10, which subjects were asked to compare on compactness. (Of the 116 maps discovered, 107 were produced in color. These maps may be found at `http://opim.wharton.upenn.edu/~sok/phillydistricts/doc/TeamFred_CompleteMaps.pdf`.) Note that on the max-max distance criterion (in the "Objective" column) map 10 has a lower score than map 9 and hence is counted as more compact by this criterion. The voting, however, is 117 to 83 in favor of map 9. As noted earlier, the average Schwartzberg, the Distance^2, and the Perimeter measures all agree with the voting in counting map 9 as more compact than map 10. We find this a remarkable performance by our Mechanical Turk subjects.

13. Figure 2 shows gray scale versions of maps 15 and 16, which subjects were asked to compare on compactness. Note that on the max-max distance criterion (in the "Objective" column) map 15 has a lower score than map 16 and hence is counted as more compact by this criterion. The voting, however, is 116 to 84 in favor of map 16. As noted earlier, the average Schwartzberg, the Distance^2, and the Perimeter measures all agree with the voting in counting map 16 as more compact than map 15. Again, we find this a remarkable performance by our Mechanical Turk subjects.

4 Discussion

The problem of compactness—of defining it, of using it—is central to the problem of avoiding malapportionment in electoral districting. Yet, as we have noted, it is beset with difficulties. In his review of compactness measures for districting, Young concludes that "compactness is such a hazy and ill-defined concept that it seems impossible to apply it, in any rigorous sense, to matters of law" (Young, 1988, page 113). This is widely accepted, with the following passage reflecting the received view.

> Compactness refers to the shape of a district, but is formally ill-defined. Scholars have proposed over fifty compactness measures, which have not resulted in clarity, since these measures conflict and can be manipulated. (Altman and McDonald, 2012, page 1190)

The approach we have sketched may well afford keeping the baby without the bath water. Young's analysis presumes that the standard measures of compactness are to be assessed in an absolute fashion. If so, then we would agree that the contingencies of geography and policy will defeat any attempt to mandate a threshold score for compactness on any known measure. But there is an alternative. The perspective we would urge is to use them in a relative or comparative sense. We should judge a districting plan as (relatively) compact based on measure M if it scores well *compared to relevant alternative plans,* the relevant plans being that plurality of known solutions that are legally valid and have generally good scores for compactness (comparatively). If one or a pool of compactness measures can be empirically validated, then surely the move we have indicated of focusing discussion on the better-scoring solutions could be used at the least to shift the burden of argument onto those who would advocate significant departures. Very many extant districting plans are in fact in this category. Compare the maps above.

Questions that naturally arise in consequence include: (1) How do disinterested human subjects judge compactness of electoral districts and to what extent do these judgments agree with established measures of compactness? (2) How might pooled judgments regarding compactness be used to improve the reapportionment process? Regarding (1), as the following passage attests, the advisability of undertaking empirical assessment of districting plans and methods is becoming recognized.

> The development of a new generation of software, such as our DistrictBuilder system [`http://www.districtbuilder.org/`, last visited 2013-03-03], has enabled members of the public to create hundreds of real districting plans—plans based on official census data, and satisfying all of

Figure 1: Map 9 (left), map 10 (right)

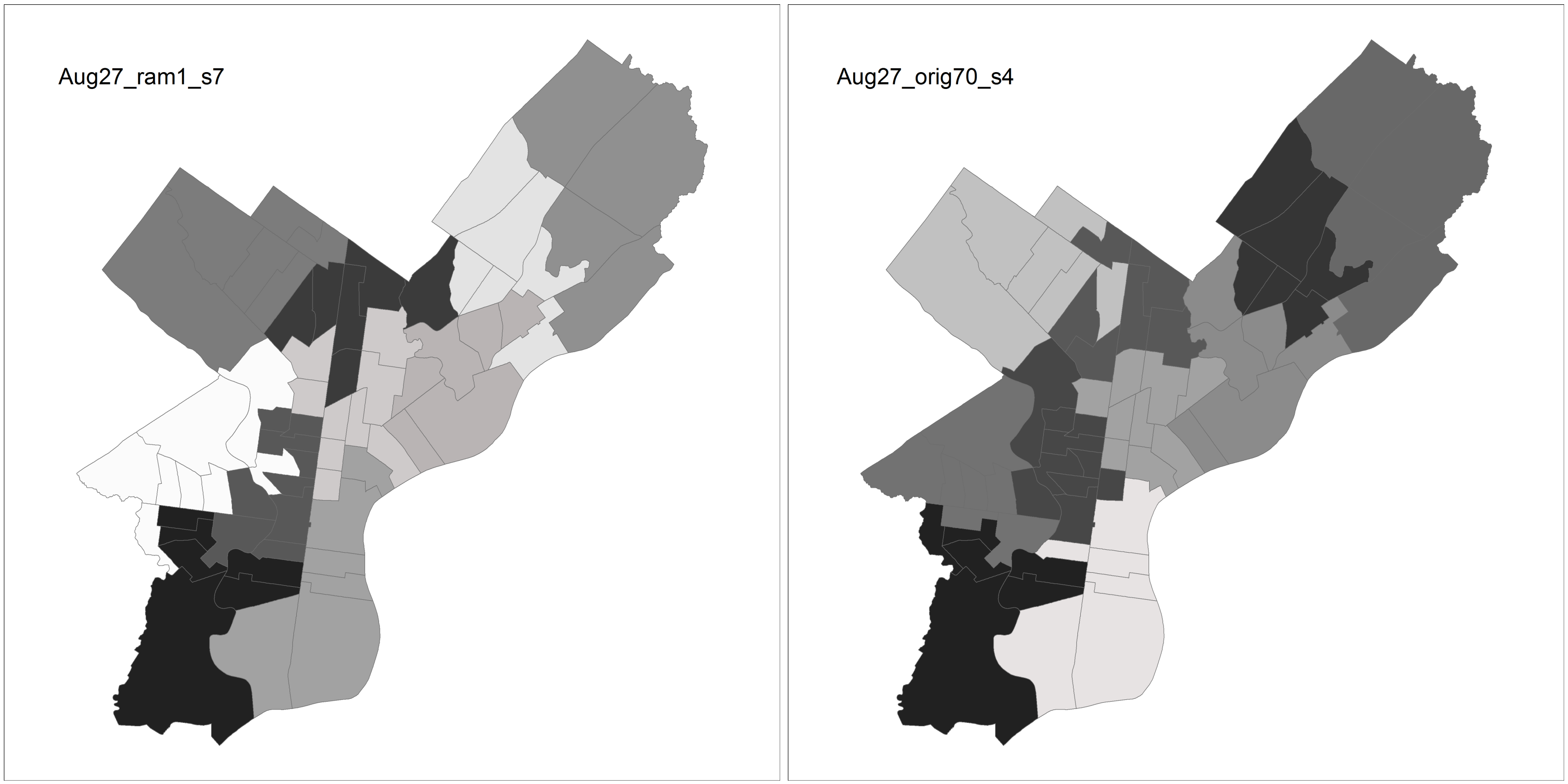

Figure 2: Map 15 (left), map 16 (right)

> the criteria required by law. The existence of this large corpus of redistricting plans for the first time is beginning to enable an empirical analysis of redistricting plans that both expands our understanding of the range of redistricting outcomes that are feasible in practice, and illuminates the trade-offs that members of the public tend to view as most desirable. (Altman and McDonald, 2012, page 1197)

The experiment we report here is among the first, if it is not the first, to compare human judgments of compactness for real districting plans with standard measures of compactness for those plans. Of course, many more experiments are called for. What we have here, however, are encouraging results and a replicable method.

Regarding question (2), even if an acceptable measure of compactness (or pool of such measures) could be agreed to, there remain severe problems of application.

> The ideal data to estimate [compactness] would contain the geographical coordinates of every household in the United States, its political district, some measure of distance between any two households within a state, and a precise definition of communities of interest. This information is not available. (Fryer, Jr. and Holden, 2011, page 506)

The passage is correct as far as it goes, but it greatly understates the difficulties. At the least, additional criteria, such as respect for existing political boundaries and geographic features, will often be important. And arriving at "a precise definition of communities of interest" is an enormous undertaking. The upshot is that for the foreseeable future we shall be under the condition described at the outset of having only models that are very inaccurate representations of the underlying problem.

The VSF idea, supported by the empirical findings reported here, affords the third option identified above: Using an admittedly flawed model to support discussion and deliberation, implementing a modification of an optimal solution to the model in response to considerations not fully represented in the model. In a nutshell, the idea is to use the ability to generate a large number of good solutions to *constrain* the decision process to a suitably characterized pool, each member of which is reasonably compact and meets the legal requirements on contiguity and population size. One might set a loose requirement on compactness, perhaps as measured by several different rules. Evolutionary computation, or indeed any other workable method, could be used to generate a large consideration set of objectively good quality solutions. Let anyone contribute to this pool, but limit the consideration set to the pool. This leaves room for a modicum of politics and for incorporation of other features not incorporated into the computational models, and it secures choice among objectively good solutions. A regime on this order (further details are needed, of course) could in fact be implemented by agreement of non-partisan districting boards, where they exist. Other zone design applications—such as design of sales districts—might follow a similar procedure as well.

In this regard, we note that redistricting problems are normally multi-objective—

> Drawing political maps necessarily involves making difficult choices among a number of competing political values. Among these are population equality, representation of minority groups, geographic compactness and contiguity, respect for the boundaries of other political units at different spatial scales, continuity with previous district boundary, and (of course) partisan advantage. . . (Eagles et al., 1999)

—as are zone design problems generally (Sacks, 2000). Yet, often there is (as in the Philadelphia districting case) a practical impossibility of representing all important aspects of the problem in a single optimization model, if only for lack of data. By producing a plurality of good solutions to a reasonable optimization model, we can hope to provide stakeholders with useful grist for the mill of deliberation.

Notes

[1]After Governor Elbridge Gerry of Massachusetts who in 1812 engineered a districting plan in which one of the districts resembled a salamander. The term "Gerry salamander" was created and soon shortened to "gerrymander" and the name has stuck.

[2]Representation of racial minorities, representation of communities of interest or neighborhoods, and holding inviolate existing political subdivisions are prominent among the issues in reapportionment for which compactness may serve as an

objective criterion for decision making (Barabas and Jerit, 2004). Also, (Altman, 1997; Altman and McDonald, 2010, 2011) constitute a very helpful series of general background articles.

[3]See the working paper by Dopp and Godfrey (2012) for additional visualizations demonstrating problems with standard measures of compactness for electoral districting.

[4]Thanks to Ram Gopalan for using classical OR techniques for finding four contiguous seeds. See Chou et al. (2012).

[5]In all, 107 maps were produced in color. They may be found in a single PDF file at `http://opim.wharton.upenn.edu/~sok/phillydistricts/doc/TeamFred_CompleteMaps.pdf`. Each map is identified uniquely. The correspondence between the map numbers of the experiments and the map identifiers in the PDF file are as follows: map 1, page 7, label Aug27_orig70_s7; map 2, page 89, label Aug27_ram2_s4; map 3, page 32, label Aug27_orig70_s32; map 4, page 94, label Aug27_ram2_s9; map 5, page 88, Aug27_ram2_s3; map 6, page 48, label Aug27_orig70_s48; map 7, page 56, label Aug27_orig70_s56; map 8, page 65, label Aug27_orig70_s65; map 9, page 19, label Aug27_orig70_s19; map 10, page 10, label Aug27_orig70_s10; map 11, page 93, label Aug27_ram2_s8; map 12, page 9, label Aug27_orig70_s9; map 13, page 39, label Aug27_orig70_s39; map 14, page 96, label Aug27_ram2_s11; map 15, page 77, label Aug27_ram1_s7; and map 16, page 4, label Aug27_orig70_s4.

5 Acknowledgments

The work reported here is new, but part of a larger project; see (Gopalan et al., 2012b). Thanks to Azavea for sponsoring the districting contest that occasioned the larger project and for providing us with key information. Thanks to Ram Gopalan and Nicholas Quintus for optimization work and GIS work. The integer programming aspect of our project, which was submitted to the contest, is described in (Gopalan et al., 2013); see also (Gopalan et al., 2012a). Thanks to Nicholas Quintus for preparing the maps in the figures of this paper. Thanks to Knowledge@Wharton for a concise, accurate story, `http://knowledge.wharton.upenn.edu/article/2850.cfm`. Our project repository is at `http://opim.wharton.upenn.edu/~sok/phillydistricts/`.

References

Altman, M. (1997). The computational complexity of automated redistricting: Is automation the answer? *Rutgers Computer and Technology Law Journal*, 23(1):81–142.

Altman, M. and Klass, G. M. (Fall 2005). Current research in voting, elections, and technology. *Social Science Computer Review*, 23(3):269–273.

Altman, M., MacDonald, K., and McDonald, M. (Fall 2005). From crayons to computers: The evolution of computer use in redistricting. *Social Science Computer Review*, 23(3):334–346.

Altman, M. and McDonald, M. P. (2010). Promise and perils of computers in redistricting. *Duke Journal of Constitutional Law & Public Policy*, 5(1):69–112.

Altman, M. and McDonald, M. P. (2011). BARD: Better automated redistricting. *Journal of Statistical Software*, 42(4):1–28.

Altman, M. and McDonald, M. P. (2012). Redistricting principles for the twenty-first century. *Case Western Reserve Law Review*, 62(4):1179–1204.

Barabas, J. and Jerit, J. (2004). Redistricting principles and racial representation. *State Politics and Policy Quarterly*, 4(4):415–435.

Cain, B. E. (1984). *The Reapportionment Puzzle*. University of California Press, Berkeley,CA.

Chou, C., Kimbrough, S. O., Murphy, F. H., Sullivan-Fedock, J., and Woodard, C. J. (2013). On empirical validation of compactness measures for electoral redistricting and its significance for application of models in the social sciences. *Social Science Computer Review*, page (forthcoming). doi: 10.1177/0894439313484262 `http://ssc.sagepub.com/content/early/2013/04/09/0894439313484262`.

Chou, C., Kimbrough, S. O., Sullivan, J., Woodard, C. J., and Murphy, F. H. (2012). Using interactive evolutionary computation (IEC) with validated surrogate fitness functions for redistricting. In Soule, T., editor, *Proceedings of GECCO 2012 (Genetic and Evolutionary Computation Conference)*, pages 1071–1078, New York, NY, USA. ACM. `http://dx.doi.org/10.1145/2330163.2330312`.

Cox, G. W. and Katz, J. N. (2002). *Elbridge Gerry's Salamander: The Electoral Consequences of the Reapportionment Revolution*. Political Economy of Institutions and Decisions. Cambridge University Press, Cambridge, UK.

Dawkins, R. (1985). *The Blind Watchmaker: Why the Evidence of Evolution Reveals a Universe without Design*. W. W. Norton & Company, New York, NY.

Dopp, K. and Godfrey, N. (2012). Legislative redistricting—compactness and population density fairness. SSRN online. `http://ssrn.com/abstract=1945879`.

Draper, R. (2012). The league of dangerous mapmakers. *The Atlantic*. `http://www.theatlantic.com/magazine/archive/2012/10/the-league-of/309084/`.

Eagles, M., Katz, R. S., and Mark, D. (Spring 1999). GIS and redistricting: Emergent technologies, social geography, and political sensibilities. *Social Science Computer Review*, 17(1):5–9.

Fryer, Jr., R. G. and Holden, R. (2011). Measuring the compactness of political districting plans. *Journal of Law and Economics*, 54(3):493–535.

Gopalan, R., Kimbrough, S. O., Murphy, F. H., and Quintus, N. (2012a). The Philadelphia districting contest: Designing territories for city council based upon the 2010 census. Video on the Web. `https://live.blueskybroadcast.com/bsb/client/CL_DEFAULT.asp?Client=569807&PCAT=5277&CAT=5278`.

Gopalan, R., Kimbrough, S. O., Murphy, F. H., and Quintus, N. (2012b). Team Fred home page. World Wide Web. `http://opim.wharton.upenn.edu/~sok/phillydistricts/`.

Gopalan, R., Kimbrough, S. O., Murphy, F. H., and Quintus, N. (2013). The Philadelphia districting contest: Designing territories for City Council based upon the 2010 census. *Interfaces*, 43(5):477–489. `http://dx.doi.org/10.1287/inte.2013.0697`.

Kalcsics, J., Nickel, S., and Schröder, M. (2005). Towards a unified territorial design approach—Applications, algorithms and GIS integration. *Sociedad de Estadística e Investigación Operativa Top*, 13(1):1–74.

Kimbrough, S. O., Kuo, A., LAU, H. C., Murphy, F. H., and Wood, D. H. (2011). Solution pluralism and metaheuristics. In *MIC 2011: The IX Metaheuristics International Conference*, Udine, Italy.

Murphy, F. H., Hess, S. W., and Wong-Martinez, C. G. (2012). Politics. In Gass, S. and Fu, M., editors, *Encyclopedia of Operations Research*. Springer.

Niemi, R. G., Grofman, B., Carlucci, C., and Hofeller, T. (1990). Measuring compactness and the role of a compactness standard in a test for partisan and racial gerrymandering. *Journal of Politics*, 52(4):1155–81.

Sacks, S. R. (Spring 2000). Optimal spatial deployment of police patrol cars. *Social Science Computer Review*, 18(1):40–55.

Satullo, C. (2011). Fixphillydistricts: Voters offer winning visions for council. WHYY Web page. `http://www.newsworks.org/index.php/homepage-feature/item/25988-06csmaps`.

Takagi, H. (2001). Interactive evolutionary computation: Fusion of the capabilities of EC optimization and human evaluation. *Proceedings of the IEEE*, 89(9):1275–1296.

Takagi, H. and Iba, H. (2005). Interactive evolutionary computation. *New Generation Computing*, 23(2):113 – 114.

Turner, M. L. and Lamacchia, R. A. (Spring 1999). The U.S. census, redistricting, and technology: A 30-year perspective. *Social Science Computer Review*, 17(1):16–26.

Wang, S. (2013a). The great gerrymander of 2012. *The New York Times*. `http://www.nytimes.com/2013/02/03/opinion/sunday/the-great-gerrymander-of-2012.html`.

Wang, S. (2013b). Slaying the gerrymander. `http://election.princeton.edu/2013/02/03/slaying-the-gerrymander/`.

Young, H. P. (1988). Measuring the compactness of legislative districts. *Legislative Studies Quarterly*, 13(1):105–115.

`$Id: evaluating-ill-defined-objectives-RN.tex 3684 2013-09-09 20:37:28Z sok $`