APPORTIONMENT SCHEMES AND THE QUOTA METHOD

M.L. Balinski*
H.P. Young*

March 1977

*Graduate School and University Center, City University of New York, and the International Institute for Applied Systems Analysis, Laxenburg, Austria.

This work was supported by the National Science Foundation under Contract MCS 75-07414 A01 with the Graduate School of the City University of New York.

ABSTRACT

This paper answers criticisms [4] recently leveled at the Quota Method for Congressional apportionment, and reconsiders the relative merits of various axioms and methods.

APPORTIONMENT SCHEMES AND THE QUOTA METHOD

M.L. Balinski and H.P. Young

1. *BACKGROUND: HOUSE-MONOTONICITY AND QUOTA*

The *apportionment problem* is the problem of determining how to divide the number of representatives in a legislature proportionally among given constituencies. In the United States the problem is rooted in the Constitution, which requires a distribution of Representatives among the various States "*according to their respective numbers.*" The issue is to find an *operational method* for interpreting this mandate, and to identify the essential properties that any fair and reasonable method ought to have. In a recent paper [4] various properties and methods have been suggested as desirable; the purpose of this paper is to examine these proposals in the light of the problem they purport to address.

Formally, the apportionment problem may be stated as follows. Let $\underset{\sim}{p} = (p_1, p_2, \ldots, p_s)$ be the populations of s *states*, where each $p_i > 0$ is integer, and let $h \geq 0$ be the number of seats in the *house* to be distributed. The problem is to find, for any $\underset{\sim}{p}$ and all house sizes $h \geq 0$, an *apportionment* for h: an s-tuple of non-negative integers $\underset{\sim}{a} = (a_1, \ldots, a_s)$ whose sum is h. A *solution* of the apportionment problem is a function $\underset{\sim}{f}$ which to every $\underset{\sim}{p}$ and h associates a unique apportionment for h, $a_i = f_i(\underset{\sim}{p}, h) \geq 0$ where $\sum_i a_i = h$. A specific apportionment method may give several different solutions, for "*ties*" may occur when using it -- for example when two states have identical populations and must share an odd number of seats. It is useful, for this reason, to define an *apportionment method* $\underset{\sim}{M}$ as a non-empty set of solutions. Two different apportionment solutions $\underset{\sim}{f}$ and $\underset{\sim}{g}$ of a method $\underset{\sim}{M}$ may be identical up to some house h and then branch, depending on how a particular tie is resolved. The restriction of $\underset{\sim}{f}$ to the domain $(\underset{\sim}{p}, h')$, $0 \leq h' \leq h$, will be called a *solution up to* h, $\underset{\sim}{f}^h$, and $\underset{\sim}{f}$ will be called an *extension* of $\underset{\sim}{f}^h$.

The principles that should apply to apportionment have been intensely debated ever since the Constitutional Convention in 1787.

From these debates two basic themes emerge. The first, fundamental to the approaches of Hamilton, Webster, and later contributors is that the ideal or exact number of seats that any state i should receive is $p_i h / \sum_j p_j = q_i$, called the *exact quota* of state i, and in any case no state should receive *less* than its *lower quota*, $\lfloor q_i \rfloor$, or *more* than its *upper quota*, $\lceil q_i \rceil$. Any method whose solutions have this property is said to *satisfy quota*.

One such method, first proposed by Alexander Hamilton, and used from 1850 through 1900 is the following: first give to each state i its lower quota and then distribute the remaining seats, one each, to the states with the largest fractional remainders. A fundamental difficulty with this method -- which begat the second basic theme for debate -- came to light in 1881 when Alabama would have *lost* seats by this method as the house increased from 299 to 300. This behavior is not only shocking to common sense and any reasonable notion of fair division, but has proved to be totally unacceptable politically -- as members of Congress immediately perceived. As Representative John C. Bell put it, "*This atrocity which* [mathematicians] *have elected to call a 'paradox' ... this freak presents a mathematical impossibility.*" (Stated in debate, 8 January 1901.) For this reason the Hamilton Method was abandoned in 1911, and the basic principle was recognized that an apportionment method must be *house-monotone*; that is, if the total number of seats to be apportioned *increases*, then *ceteris paribus* no state should receive fewer seats than it did before.

Yet in [4] it is said that, "*there is no real reason for requiring apportionment to be house-monotone. The objective should be to <u>minimize</u> <u>inequity</u>.*

2. *MINIMIZING "INEQUITY" AND CONSISTENCY*

Intuitively, "*minimizing inequity*" is what the apportionment problem is all about. The real problem is to determine what "*inequity*" means. To say it is desirable to "*minimize the length of the inequity vector in Euclidean s-space*" begs the question. Indeed, as pointed out in [4], " ... *All measures ... of inequity are to some extent arbitrary.*"

Motivated by the need for house-monotone methods E.V. Huntington began in 1921 [8] the investigation of several measures of inequity based on pairwise comparisons of states' relative representation. Given populations $\underset{\sim}{p}$ and an apportionment $\underset{\sim}{a}$ for h, p_i/a_i and a_i/p_i represent, respectively, the *average district size* and the "*share of a representative*" of state i. If $p_i/a_i > p_j/a_j$ then state j is *better off* than state i. Let $T(i,j) = T(j,i) \geqq 0$ be a measure of inequity between states i and j. An apportionment $\underset{\sim}{a}$ is *in equilibrium* if no transfer of one seat from a better off state j to a less well off state i reduces the value of $T(i,j)$. Certain T's admit no equilibrium apportionments, but Huntington showed ([7][8]) that others do and that five different apportionment methods devolve from these. For example, $T(i,j) = |p_i/a_i - p_j/a_j|$ yields Harmonic Mean ($\underset{\sim}{HM}$) apportionments whereas $T(i,j) = |a_j/p_j - a_i/p_i|$ gives Webster ($\underset{\sim}{W}$) apportionments. Huntington argued that the most natural choice was the "*relative difference*" $T(i,j) = |p_i/a_i - p_j/a_j|/\min(p_i/a_i, p_j/a_j)$ and showed this choice leads to the Method of Equal Proportions ($\underset{\sim}{EP}$). He was persuasive: the U.S. Congress adopted $\underset{\sim}{EP}$ as the law beginning in 1941. Nevertheless this choice of measure of inequity remained arbitrary.

Huntington unified his five methods -- many of which were anticipated by others in one guise or other -- through a computational approach. We generalize it. Let $r(p,a)$ be any real-valued function of two variables called a *rank-index*. Then an apportionment method $\underset{\sim}{M}$ is obtained by taking all apportionment solutions $\underset{\sim}{f}$ defined recursively as follows: (i) $f_i(\underset{\sim}{p},0) = 0$, $1 \leqq i \leqq s$; (ii) if $a_i = f_i(\underset{\sim}{p},h)$ and k is some one state for which $r(p_k,a_k) \geqq r(p_i,a_i)$ for $1 \leqq i \leqq s$, then $f_k(\underset{\sim}{p},h+1) = a_k + 1$ and $f_i(\underset{\sim}{p},h+1) = a_i$ for $i \neq k$. These we call *Huntington Methods*. For example, $\underset{\sim}{HM}$ has the rank-index $r(p,a) = p/\{2a(a+1)/(2a+1)\}$, $\underset{\sim}{W}$ has $r(p,a) = p/(a+\frac{1}{2})$, and $\underset{\sim}{EP}$ has $r(p,a) = p/\{a(a+1)\}^{\frac{1}{2}}$.

Clearly all Huntington methods are house-monotone. But they also satisfy a condition which epitomizes the very idea of "*method*": namely, the decision as to which state of any pair most deserves the extra seat as the house size is increased by 1 depends only upon the populations and seats already allocated to those states singly, and not on the vector $\underset{\sim}{p}$ or the vector $\underset{\sim}{a}$ of seats so far allocated. Consider a method $\underset{\sim}{M}$ and suppose that it has a solution $\underset{\sim}{f}$ allocating to a state with p^* votes a^* seats and to a state with $\bar{p}$ votes $\bar{a}$ seats in a house h, while $\underset{\sim}{f}$ allocates to the star state $a^* + 1$ seats and to the bar-state $\bar{a}$ seats in a house $h + 1$. Then the star state is said to have *weak priority* at that point and this is written $(p^*,a^*) \geq_{\underset{\sim}{M}} (\bar{p},\bar{a})$. A natural criterion for any method is that the relative claims for an extra seat between two states should depend *only* on their respective populations and apportionments. Specifically, if for some other problem with populations $\underset{\sim}{p}'$ there are states having p^* and $\bar{p}$ which are allocated, by a solution of $\underset{\sim}{M}$, a^* and $\bar{a}$ seats respectively, and $(\bar{p},\bar{a}) \geq_{\underset{\sim}{M}} (p^*,a^*)$ then then the states are said to be *tied*, and this is written $(p^*,a^*) \sim_{\underset{\sim}{M}} (\bar{p},\bar{a})$. A method is said to be *consistent* if it treats tied states equally, that is, if $(p^*,a^*) \sim_{\underset{\sim}{M}} (\bar{p},\bar{a})$ implies $\underset{\sim}{f}^h$ has both an extension giving the star state a^*+1 seats at $h+1$, and an extension giving the bar state $\bar{a} + 1$ seats at $h + 1$. Any two states will naturally compare their resultant numbers of seats: a change in priorities could not but be viewed as conflicting with common sense.

Theorem 1 [2]. *An apportionment method* $\underset{\sim}{M}$ *is house-monotone and consistent if and only if it is a Huntington method.*

3. QUOTAS AND PSEUDO-QUOTAS

It is a major defect of Huntington methods that *none* of them satisfies quota ([3]),p.712).

However, one may arrive at certain of Huntington's methods by the device of defining "*pseudo-quotas*" (see [3],p.709). In [4], the "*radically different resolution of the Alabama paradox ... apportionment by σ-quota*" ([4], last paragraph, p.684) is one such instance. The approach is to define $p_i/\sigma = q_i(\sigma)$ as the σ-quota of state i, with σ a maximum (*not* minimum as said in [4]) allowable average population per district of any state. Letting $a_i(\sigma) = \lceil q_i(\sigma) \rceil$, a σ is sought for which $\sum_1^s a_i(\sigma) = h$ (the size of the house). We omit the improbable case of a tie.

Then, $a_i(\sigma) \geqq p_i/\sigma$ or $\sigma \geqq p_i/a_i$. The house-monotone method which results is known as Smallest Divisors ($\underset{\sim}{SD}$), was known to Huntington in his 1928 paper [7], and was described in precisely this way on p. 709 of [3].

The idea of redefining quota in a manner similar to that of the σ-quota is one which is solidly planted in American history. Jefferson advanced it in 1792 (see [3], p.703).

In [4] it is suggested that "$\underset{\sim}{EP}$ *should not be considered unacceptable because it fails to satisfy 'quota' -- a short-coming that is easily cured, moreover.*" Presumably this means that an altered quota idea can be used to explain $\underset{\sim}{EP}$. This is true. An $\underset{\sim}{EP}$ apportionment for house size h is found by choosing a σ such that if $a_i(\sigma) = \lfloor\{p_i^2/\sigma^2 + \frac{1}{4}\}^{\frac{1}{2}} + \frac{1}{2}\rfloor$ then $\sum_1^s a_i(\sigma) = h$ (see [3], p.709). This hardly seems to commend $\underset{\sim}{EP}$ -- or any method based on some pseudo-quota notion -- as a natural method to adopt.

Can satisfying quota be reconciled with monotonicity and consistency? Indeed it can. If consistency is weakened to apply only when upper quota is not violated, then there exists a unique method, the *Quota Method* ($\underset{\sim}{Q}$), which satisfies the three properties [1], [3]. It is said that "$\underset{\sim}{Q}$ *uses a much more arbitrary and extreme measure of inequity than* $\underset{\sim}{EP}$" ([4], p.685). But the fact is that $\underset{\sim}{Q}$ is not based on *any* measure of inequity. The description of $\underset{\sim}{Q}$ "*that the* <u>*augmented*</u> *representatives shall be as nearly as possible proportional to the populations*" [3], p.685) is false.* In fact the former defines the Huntington method $\underset{\sim}{J}$ first proposed by Jefferson (see [3], p.703), also much used in Europe but known as the method of d'Hondt, and cited by Birkhoff as $\underset{\sim}{GD}$ (Greatest Divisors) which he claims is superior to $\underset{\sim}{Q}$.

$\underset{\sim}{Q}$ is defined recursively as follows: (i) $f_i(\underset{\sim}{p},0) = 0$, $1 \leqq i \leqq s$; (ii) if $a_i = f_i(\underset{\sim}{p},h)$, $E(h+1)$ is the set of states which can receive an extra seat without violating upper quota at $h+1$, and $k \in E(h+1)$ is some one state satisfying $p_k/(a_k+1) \geqq p_i/(a_i+1)$ for all $i \in E(h+1)$, then $f_k(\underset{\sim}{p},h+1) = a_k + 1$ and $f_i(\underset{\sim}{p},h+1) = a_i$ for $i \neq k$.

---

*Birkhoff's example is incorrect. If $p_1 = 23{,}500{,}000$ and $p_2 = 1{,}500{,}000$ then $\underset{\sim}{Q}$ first gives State 2 a second seat when State 1 has 31 seats (not 35).

4. "BINARY FAIRNESS"

In [4] a new apportionment principle called "*binary fairness*" is advanced. This apparently reasonable condition states that if q_i and q_j are the exact quotas of states i and j, and if a_i and a_j are their apportioned numbers of seats, then it should not be possible to transfer a representative from a state i to a state j and reduce $|a_i - q_i| + |a_j - q_j|$. It is, of course, true that Hamilton's method satisfies this condition. But also, we perceive the truth of

Theorem 2. An apportionment solution satisfies the binary fairness property if and only if it is a Hamilton method solution.

Corollary. There exists no house-monotone method satisfying binary fairness.

This is immediate, since any solution satisfying binary fairness is a Hamilton method solution and no Hamilton method solution is monotone. It can only be concluded that binary fairness is inappropriate to the problem of apportionment.

5. WELL-ROUNDING AND THE WEBSTER METHOD

In [4] Birkhoff introduces a condition he calls "binary consistency," and proceeds to attack the quota method $\underset{\sim}{Q}$ as the only method--of the five proposed by Huntington, the Hamilton method and $\underset{\sim}{Q}$--which "fails to have [it]." Here we express this condition in a slightly more natural form and show that it, in fact, uniquely characterizes the Webster method in the class of Huntington methods.

Let $\underset{\sim}{a}$ be an apportionment and $\underset{\sim}{q}$ the exact quotas. If $a_i > q_i + 1/2$ we say that state i's apportionment a_i is *over-rounded*, while if $a_j < q_j - 1/2$ that state j's apportionment is *under-rounded*. If there exists no pair of states i and j, with a_i over-rounded and a_j under-rounded, then $\underset{\sim}{a}$ is said to be *relatively well-rounded*. This is equivalent to satisfying "binary consistency."

Theorem 3. The Webster method $\underset{\sim}{W}$ is the unique method that is house-monotone, consistent, and relatively well-rounded.

Proof. We use the facts (see, e.g., [2]) that: (i) a Webster apportionment $\underset{\sim}{a}$ is characterized by

(1) $$\max_i \frac{p_i}{a_i + 1/2} \leqq \max_i \frac{p_i}{a_i - 1/2} \quad \text{for} \quad a_i \geqq 1; \quad \text{and}$$

(ii) Webster apportionments may be found recursively by

(2)

(a) $\mathbf{f}(\mathbf{p},0) = 0$

(b) if $\mathbf{a} = \mathbf{f}(\mathbf{p},h)$ and k is some one state for which $p_k/(a_k+1/2) = \max_i p_i/(a_i+1/2)$ then $f_k(\mathbf{p},h+1) = a_k + 1$, $f_i(\mathbf{p},h+1) = a_i$ for $i = k$.

First, that the Webster method is consistent and house monotone is clear by (2). Suppose it is not relatively well-rounded. Then there exists an apportionment $\mathbf{a}$ for h, with states i and j satisfying $a_i > q_i + 1/2$ and $a_j < q_j - 1/2$. Therefore, $a_i - 1/2 > q_i = p_i h/\Sigma_k p_k$ and $a_j + 1/2 < q_i = p_j h/\Sigma_k p_k$, implying

$$\frac{p_i}{a_i - 1/2} < \frac{\Sigma_k p_k}{h} < \frac{p_j}{a_j + 1/2}$$

violating (1). Thus, $\mathbf{W}$ satisfies the three conditions.

Conversely, suppose that $\mathbf{M}$ is consistent, house-monotone and relatively well-rounded, but is not a set of Webster apportionments. Then there must exist populations p, q having $\mathbf{M}$-apportionments a, b which are Webster apportionments, but

(3) $$(p,a) \gtrsim_{\mathbf{M}} (q,b) \text{ whereas } p/(a+1/2) < q/(b+1/2) \quad ,$$

equivalently, $q(a+1/2) > p(b+1/2)$. By consistency this implies that the two-state problem (p,q) has an $\mathbf{M}$ apportionment (a+1,b). But, then, the exact quota of the p-state at $h = a + b + 1$ is

$$\frac{p(a+b+1)}{p+q} = \frac{p(a+1/2+b+1/2)}{p+q} < \frac{p(a+1/2) + q(a+1/2)}{p+q} = (a+1) - 1/2$$

showing that this state is over-rounded. The corresponding exact quota of the q-state is

$$\frac{q(a+b+1)}{p + q} = \frac{q(a+1/2+b+1/2)}{p + q} > \frac{p(b+1/2) + q(b+1/2)}{p + q} = b + 1/2$$

showing that this state is under-rounded. Therefore, the apportionment (a+1,b) is not relatively well-rounded, a contradiction. This completes the proof.

It should be remembered, however, that in spite of having this property, the Webster method does not satisfy quota.

6. "BIAS"

The axiomatic approach to apportionment proceeds by making a choice concerning the *principles* which any fair apportionment should satisfy, and then identifying that method (or methods) that satisfy the principles. The advantage of beginning with agreed-upon fairness principles is that subsequent squabbles over particular numbers resulting from these principles are avoided.

Nevertheless given any method it is an almost irresistable temptation to analyze particular numerical solutions by adding and substracting different combinations of the numbers to show that the method is in some peculiar sense unfair to certain groups of states. Thus one may question whether a particular solution gives more than a just share to the "*larger*" states versus the "*smaller*" states (or the "*middle*" states) or to the North versus the South, or to the states with large fractions versus those with small fractions, and so forth. These investigations may generally be called ones of "*bias*" and they purport to establish empirically that certain "*new*" principles are violated; principles which by the very nature of the case are different from those already agreed upon as defining the method. For the notion of bias to even make sense, a normative principle must be postulated; one may then ask what methods (if any) satisfy this principle instead of other principles.

It is stated [4] that SD is "*unfair to populous states for a simple reason: every nonpopulous state 'entitled' to 1.1 representatives must be given two representatives ...*" This can only mean that if the exact quota of a "*nonpopulous state*" is at least 1.1 then SD assures this state two representatives. This is false, as the following example shows.

						Totals
State populations	4533	4686	5049	6183	9549	30,000
Exact quotas	1.51	1.56	1.68	2.06	3.18	10
SD solution ($\sigma = 4533$)	1	2	2	2	3	10

But, in any case, no normative principle is advanced to support the claim that the numerical example shows bias.

In [4] an argument is given "*to show that Q is biased*" against nonpopulous states. The argument consists of comparing the H and Q solutions for four 50-state examples, and selecting from the 50 states in each case a subset of "*nonpopulous*" states which Q rounds down and H rounds up, and a subset of "*populous*" states for which the contrary occurs ([4], Tables 2-5).* It is then observed that Q allots less than H to the nonpopulous states chosen, and more than H to the populous states chosen. It would be as pertinent to remark that virtually any apportionment solution gives some states less than their exact quotas and others more, and that the two sets will in general be different for different methods.

It is true of course that Q has a tendency of rounding up the exact quotas of large states more often than those of small states. This is unavoidable -- being a necessary consequence of the fairness principles uniquely satisfied by the Quota Method.

If H is taken as a norm for comparison, Q is then "*biased*" in that it does not necessarily round up the exact quotas of those states having the largest fractional remainders. This is the procedure which constitutes H, so Q can hardly but be "*biased*" according to this measure. But H violates the essential house-monotonicity axiom so cannot be taken as a reasonable norm for comparison. Birkhoff goes on to propose "H *as a good compromise between* Q *and ...* EP, *which goes so far in the opposite direction*

---

*The sets of "*nonpopulous*" or "*populous*" states are different and conveniently chosen in each case.

that it violates 'quota'" ([4], p.685). Thus, according to the argument of [4], EP is also biased, but preferred to Q.

Setting aside the pejorative notion of "*bias*," there *is* a precise sense in which one can talk about one method "*favoring*" large (or small) states *in comparison with* another method. This notion is defined in a precise manner and Theorem 1 ([3], p.708) compares the five Huntington methods with respect to "*favoring*" large over small states. But no Huntington method satisfies quota so those comparisons, while interesting, shed little or no light on the supposed "*bias*" of Q.

7. *THE ROLE OF AXIOMS AND QUOTA METHOD*

The lessons of history clearly point to the necessity of arriving at a fundamental understanding of the *properties* of methods. Put in other terms, political apportionment must be based on principles of fair division rather than on *ad hoc* choices of measures of inequity. Thus axiomatics finds a political role!

In [4] Birkhoff attacks Q for a variety of reasons. First, Q is faulted because it fails to satisfy the "binary fairness" property, although it is ignored that this property uniquely determines the Hamilton method (which is not house-monotone). Second, Q is noted to violate "binary consistency", although it is not observed that this property uniquely determines the Webster method (W) in the class of Huntington methods (moreover W is not recommended by Birkhoff). Third, a description of an admittedly arbitrary measure of inequality supposedly "*used*" by Q is attacked, but this measure is not used by Q (it characterizes Jefferson's method J, or GD, one of three methods recommended by Birkhoff). Fourth, house-monotonicity is discarded as having "*no real reason,*" while minimizing any inequity measure is deemed preferable.

This confused state of affairs can only be cleared up through a careful construction of fundamental axioms which satisfy precedents explicitly or implicitly determined by the U.S. Constitution, its framers and interpreters, and by the members of Congress. Further, in the words of Zecharia Chaffee, Jr., "*the preservation of a respect for the law will in the long run be best obtained by the adoption of a plan which is least likely to produce a sense of unfairness in those who are forced to obey legislation*" ([6], pp.1043-1044).

REFERENCES

[1] Balinski, M.L. and Young, H.P. (1974), A New Method for Congressional Apportionment, *Proc.Nat.Acad.Sci. USA* 71, 4602-4606.

[2] Balinski, M.L. and Young, H.P. (1976), On Huntington Methods of Apportionment, RR-77-2, International Institute for Applied Systems Analysis, Laxenburg, Austria.

[3] Balinski, M.L. and Young, H.P. (1975), The Quota Method of Apportionment, *Amer.Math.Month.* 82, 701-730.

[4] Birkhoff, G. (1976), House-monotone Apportionment Schemes, *Proc.Nat.Acad.Sci. USA* 72, 684-686.

[5] Bliss, G.A., Brown, E.W., Eisenhart, L.P., and Pearl, R. (1929), Report to the President of the National Academy of Sciences, 9 February, in *Congressional Record*, 70th Congress, 2nd Session, 70, 4966-4967.

[6] Chaffee, Z., Jr. (1929), *Congressional Reapportionment*, Harvard Law Review, XLII, 1015-1047.

[7] Huntington, E.V. (1928), The Apportionment of Representatives in Congress, *Trans.Amer.Math.Soc.* 30, 85-110.

[8] Huntington, E.V. (1921), The Mathematical Theory of the Apportionment of Representatives, *Proc.Nat.Acad.Sci. USA* 7, 123-127.

[9] Morse, M., von Neumann, J. and Eisenhart, L.P. (1948), Report to the President of the Academy of Sciences, 28 May.