

LEVEL II

(2)

AD A057061

(14) BKD - TR-3-251

(13) ALLOCATION FOR
AUTHORIZATION MANAGEMENT,

(13) John P. Mayberry Ph.D.

(11) 1 April 1978

(13) 72 p.

RECEIVED
APR 3 1978
F

DDC FILE COPY

This report was prepared for the

(15) Office of Naval Research
under Contract N00014-77-C-0386

B-K Dynamics, Inc.
15825 Shady Grove Road
Rockville, Maryland 20850

Approved for Public Release; Distribution Unlimited

387 575
78 07 26 057

BOK
DYNAMICS, INC.

113

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER TR-3-251	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Allocation for Authorization Management		5. TYPE OF REPORT & PERIOD COVERED
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) John P. Mayberry, Ph.D.		8. CONTRACT OR GRANT NUMBER(s) N00014-77-C-0386
9. PERFORMING ORGANIZATION NAME AND ADDRESS B-K Dynamics, Inc., 15825 Shady Grove Road Rockville, Maryland 20850		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research, Code 450 800 North Quincy Street Arlington, Virginia 22217		12. REPORT DATE 1 April 1978
		13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) PERSONNEL RESOURCE MANPOWER ASSETS ALLOCATION		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) This report explains the methodology for operationally calculating and allocating manpower and personnel authorizations in the U.S. Navy. ↗		

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Allocation for Authorization Management

TABLE OF CONTENTS

	Page
INTRODUCTION	1
1. The Quantitative Apportionment Problem	3
2. Lower and Upper Limits on the Quotas	4
3. Quantitative Apportionment with Relative Priorities	4
4. Prohibition of Over-Manning in Under-Manned Communities	5
5. Special "CNO Priority-1" UIC's	5
6. Computation of Pro-Rated Allowances (PRA)	6
7. Grade Allocation for CNO Priority-1 UIC's	9
8. Non-Integral Grade-by-Grade Allocation	11
9. Integral Grade-by-Grade Allocation	14
10. Allocation of Men by Grade to UIC's	14
11. Qualitative Priority	17
12. Which Quota Method?	19
13. Conclusion	20
APPENDIX I - Methods of Apportionment	I-1
I.1 The Apportionment Problem	I-1
I.2 The Work of Balinski and Young	I-2
I.3 Notation and Definitions	I-2
I.4 Duality and the Quota Method	I-4
I.5 The Alabama Paradox	I-5
I.6 Huntington's Work: Divisor Methods	I-6
I.7 Consistency	I-8
I.8 The Dual Quota Method	I-9

ACCESSION for	White Section	Buff Section	SP CIAL
NTIS	✓		
DOC			
UNCLASSIFIED			
CLASSIFICATION			
CLASSIFICATION/AVAILABILITY CODES			
-J/ or SP CIAL			
A			

APPENDIX II - Description of Quota Methods	II-1
II.1 The Classical Quota Methods	II-1
II.2 The Primal Quota Method	II-1
II.3 The Dual-Quota Method	II-2
II.4 Direct Quota Methods	II-3
II.5 Other Quota Methods	II-6
II.6 Comments on the Ascending Quota Methods	II-10
APPENDIX III - Theorems on Apportionment	III-1
III.1 General	III-1
III.2 Restrictions on Divisor Functions	III-3
III.3 Linear Divisor Functions	III-6
III.4 Primal Quota and Dual Quota Methods	III-10
III.5 Influence of the Upper Bounds and Lower Bounds	III-15
III.6 Periodicity of Apportionment Solutions	III-18
III.7 Backtracking	III-24
III.8 Ties	III-27
III.9 Non-Integral Populations	III-28
III.10 Recommendations	III-29
REFERENCES	III-30

TABLES

	Page
Table I	
Calculation of Pro-Rated Allowances for Hypothetical UIC	8
Table II	
Example Illustrating Allocation Procedure	10
Table III	
Non-Integral Allocation of Available Men Against Pro-Rated Allowances	12
Table IV	
Non-Integral Allocation of Cumulative Available Men Against Cumulative Pro-Rated Allowances	12
Table V	
Integral Allocation of Cumulative Available Men Against Cumulative Pro-Rated Allowances	15
Table VI	
Integral Allocation of Available Men Against Pro-Rated Allowances	15
Table III.1	
Example Showing That Both HM and EP Can Violate Upper Quota	III-7
Table III.2	
The Apportionments Generated by Eleven Distinct Methods	III-20

ALLOCATION FOR AUTHORIZATION MANAGEMENT

Introduction

This paper explains the method proposed by B-K Dynamics, Inc. for operationally calculating and allocating manpower and personnel authorizations in the U.S. Navy. We are concerned with an allocation which is intended to be a useful forecast of actual personnel assignments two years in the future. This proposal incorporates many of the known important considerations, including:

- (i) predicted availability of personnel of various pay grades within each community;
- (ii) approved allowances by rating and by grade for each Unit Identification Code (UIC);
- (iii) equity among UIC's of equal priority;
- (iv) relative priorities associated with each UIC;
- (v) directives that (for certain special UIC's, called "CNO Priority-1 UIC's") authorizations must agree exactly with allowances; and
- (vi) substitution (within a rating) of one grade for another, when necessary (minimizing both the number of such substitutions and the grade-differences involved).

In putting forward these proposals, B-K is acutely aware of several other factors, which are important in themselves and could indeed be crucial to the success and usefulness of the entire Authorization Management System, but which do not form a part of the allocation procedure as we describe it herein. Among those factors are the following:

- (i) the division of all personnel among "distributable communities" -- which are often either designators or NOBC's -- each of which is assumed to be completely homogeneous and completely disjoint from all the others;
- (ii) the numerical approved allowances by grade within each distributable community, which are simply taken as inputs by our procedure; and
- (iii) the numerical prediction of available personnel by grade within each distributable community, which are again simply taken as inputs by our procedure.

Each of the "approved allowances" and "available personnel" are predictions as of some specific date -- typically about 24 months after the time when the calculations are performed.

For expository reasons, this paper will describe a succession of allocation problems, beginning with the simplest quantitative apportionment problem and successively introducing the complicating factors. We shall give a brief verbal description, and an example, of each portion of our procedure.

1. The Quantitative Apportionment Problem

This problem is deceptively easy to state: we wish to "divide a given whole number into integral portions, which shall stand as nearly as possible in the same respective proportions as a given set of numbers." The phrase "as nearly as possible" admits a wide variety of different interpretations, and several of the most natural interpretations have been found to produce results which possess serious flaws. Because of the requirement in the U.S. Constitution that each state shall have a number of representatives in the Congress proportional to its population, the above problem has received much attention -- political, legislative, and (more recently) mathematical -- over the last two hundred years. Ref. [QMA] provides a fascinating summary of this history, and refs. [NMC A] and [QMN U] present some more recent results. Appendix I of this paper describes some of the flaws mentioned above, and describes several of the most reasonable interpretations of the phrase "as nearly as possible"; we merely remark here that the demand for "equity among UIC's of equal priority" can also be interpreted in several ways, corresponding to various possible "methods" of legislative apportionment; that the methods known as "Quota" methods avoid the most serious difficulties; and that the several known Quota methods have each their own idiosyncrasies in the context of the manpower-allocation problem. We shall refer to "the quota" of a particular UIC as a definite number, calculated from the total to be apportioned and the proportions to be employed; we defer until Appendix III below the question of exactly which quota method will be used at each stage of our procedure.

2. Lower and Upper Limits on the Quotas

Many apportionment methods permit a minimum and a maximum to be imposed on the portion allocated to each UIC, over-riding the requirement for proportionality or "fair share". Phase I of our allocation procedure is to compute each UIC's modified fair share of the total personnel available in the distributable community, imposing minima to ensure that each UIC receives at least one man if possible.

3. Quantitative Apportionment with Relative Priorities

In case relative priorities can be assigned to the various UIC's which are to share the resources of some community, we further modify the notion of "fair share", taking now the ideal ratios as proportional to the respective allowances multiplied by the respective priorities. (We shall retain the condition that, whenever possible, each UIC should receive at least one man from the community.) Thus, if some UIC has an allowance of 10 slots and a priority of 1.3, while a second UIC has an allowance of 6 slots and a priority of 1.1, the allocation would be reckoned as if they had allowances of 13.0 and 6.6 respectively.

4. Prohibition of Over-Manning in Under-Manned Communities

Compliance with existing U.S. Navy personnel assignment policies requires that two additional considerations be taken into account at this stage of our procedure: first, if the community as a whole is not over-manned, then no UIC is to be over-manned; and, second, if the community as a whole is over-manned, the relative priorities are to be disregarded in the quantitative allocation. If analogous methods are considered for use in manpower and personnel planning by other organizations, it seems likely that the first of these rules would generally be retained, but the second is perhaps more specifically tailored to the Navy's situation; in any case, these two constraints help to ensure that high priorities for important UIC's will not cause such undesirable effects as ships being crowded with more men than can be effectively utilized, or even bunked.

5. Special "CNO Priority-1" UIC's

The last factor which we shall consider, as an influence on the initial quantitative allocation, is the special treatment accorded to UIC's denoted "CNO Priority-1". Those UIC's must have exactly 100% manning, by numbers and by grade, regardless of the impact on other UIC's. In Phase I, if sufficient men are available in the community as a whole, the correct total number of men is withdrawn from the quantitative allocation process; if sufficient men are not available, we thenceforth disregard the special status of those UIC's, and use their allowances and relative priorities to compute a quantitative allocation in the ordinary way.

Sections 1 to 5 above have completed the description of Phase I -- the Quantitative Allocation Phase -- of our allocation process. To summarize: within each distributable community, the CNO Priority-1 UIC's are given their exact allowance if possible; if the community is not over-manned, the total remaining available personnel are allocated among the other UIC's, with the quota of each determined by the product of its total allowance by its relative priority, subject to over-riding minima of one man per UIC if possible and maxima equal to the allowances; if the community is over-manned, the total remaining available personnel are allocated among the other UIC's with quotas determined by allowances only.

6. Computation of Pro-Rated Allowances (PRA)

The next step is to pro-rate the quantitative allocations of each UIC among the pay grades. This is done by multiplying the approved allowances for a UIC by its "fill-ratio" -- the ratio of its quantitative allocation to its total allowance. Thus a UIC whose allocation was exactly equal to its total allowance would have a fill-ratio of 1.00 (or 100%), and its pro-rated allowances (p.r.a.) would be equal to its approved allowances; a UIC which was allocated, from some community, only 10 men against a total allowance of 15 slots, would have fill-ratio of $10/15 = 2/3$, and its pro-rated allowances would be two-thirds of its approved allowances; a UIC which was allocated 12 men against allowances of 10 slots would have a fill-rate of 1.20 (or 120%), and its pro-rated allowances would be 1.20 times its approved allowances. (The fact that these p.r.a. numbers will generally not be integers is not a problem, since we do not intend to actually allocate those numbers of men -- the aggregate grade-distribution of the p.r.a. is unlikely to exactly match the grade-distribution of available men anyway -- but merely to use them as a

measure of the relative entitlement of the various UIC's for men of the various grades.)

Instead of considering the allowances for the various grades separately, we accumulate them before we pro-rate them. Table I presents the detailed computations for a UIC with $2/3$ fill-ratio and an allowance of 15 slots, distributed as shown over grades. In particular, this UIC has an allowance of 4 slots at the G-3 level, and 7 slots at grade G-3 or higher; the $2/3$ fill-ratio, applied to those 7 slots, gives a cumulative pro-rated allowance (c.p.r.a) of 4.67 slots at grade G-3 or higher, and a pro-rated allowance of 2.67 slots at grade G-3.

When we perform the computations as indicated by the table, we find that the sum of the pro-rated allowance figures in the last column is exactly equal to the total allocation for this UIC; thus we may disregard small round-off errors. Other advantages of working with the cumulative allowances will be seen in later sections.

The calculations described in Section 6 constitute Phase II of our proposed procedure.

Pay Grade	Approved Allowance	Cumulative Allowance	Cumulative Pro-Rated Allowances (C.P.R.A.)	(P.R.A.) Pro-Rated Allowances
G-7	0	0	0.00	0.00
G-6	1	1	0.67	0.67
G-5	1	2	1.33	0.66
G-4	1	3	2.00	0.67
G-3	4	7	4.67	2.67
G-2	4	11	7.33	2.66
G-1	4	15	10.00	2.67

TABLE I
Calculation of Pro-Rated Allowances
For Hypothetical UIC

7. Grade Allocation for CNO Priority-1 UIC's

Having settled in Phase I the question of how many men each UIC shall receive, and in Phase II decided the UICs' relative entitlement to men of various grades, we proceed to determine the grade-distribution of the men to be allocated to each UIC. We do this in successive stages, first allocating men to fill the pro-rated allowances of highest grade, then allocating men to fill the residue of the p.r.a. for the two highest grades, continuing thus until we finally fill the residue of the p.r.a. for all grades. (At that point, since the allowances were pro-rated so that the total p.r.a. equals the total of the men available, we will find that the available men are exactly used up.)

However, before we begin the detailed work of fairly sharing the men of various grades among the competing UIC's, we consider any CNO Priority-1 UIC's which received fill-ratio 100% in Phase I; they should receive exactly the numbers of men of each grade specified in their approved allowances. If this is possible, we assign the correct numbers of men of each grade, and exclude those men and those UIC's from the remainder of the entire allocation-process; but if an exact grade match is not possible for all those UIC's, we disregard their special status. (In this latter case, we base their grade-mix purely on their qualitative priorities, as described below.

178/6

		Pay Grade							
		G-7	G-6	G-5	G-4	G-3	G-2	G-1	Total
UIC #1									
*Priority 1.10									
* Allowance		1	0	1	2	1	3	2	10
Cum. Allow.		1	1	2	4	5	8	10	
C.P.R.A		0.80	0.80	1.60	3.20	4.00	6.40	8.00	
P.R.A.		0.80	0.00	0.80	1.60	0.80	2.40	1.60	8.00
Cum. Alloc		1	1	2	3	4	5	8	
Alloc.		1	0	1	1	1	1	3	8
UIC #2									
*Priority 1.50									
* Allowance		0	1	2	1	4	7	5	20
Cum. Allow.		0	1	3	4	8	15	20	
C.P.R.A		0.00	1.00	3.00	4.00	8.00	15.00	20.00	
P.R.A.		0.00	1.00	2.00	1.00	4.00	7.00	5.00	20.00
Cum. Alloc.		0	0	3	3	7	13	20	
Alloc.		0	0	3	0	4	6	7	20
UIC #3									
*Priority 1.00									
* Allowance		0	1	1	2	3	4	3	14
Cum. Allow.		0	1	2	4	7	11	14	
C.P.R.A		0.00	0.79	1.57	3.14	5.50	8.64	11.00	
P.R.A.		0.00	0.79	0.78	1.57	2.36	3.14	2.36	11.00
Cum. Alloc.		0	0	1	3	5	7	11	
Alloc.		0	0	1	2	2	2	4	11
Total (All UICs)									
Allowance		1	2	4	5	8	14	10	44
Cum. Allow.		1	3	7	12	20	34	44	
C.P.R.A		0.80	2.59	6.17	10.34	17.50	30.04	39.00	
P.R.A.		0.80	1.79	3.58	4.17	7.16	12.54	8.96	39.00
Cum. Alloc.		1	1	6	9	16	25	39	
Alloc.		1	0	5	3	7	9	14	39
*Available		1	0	5	3	7	9	14	39
Cum. Available		1	1	6	9	16	25	39	

TABLE II

Example IllustratingAllocation Procedure

Notes: Asterisks denote data input to the procedure; the "allowances" are numbers of manpower slots to be filled; "C.P.R.A" denotes cumulative prorated allowances (see text); "alloc." denotes men finally allocated by the process; "totals" are obtained by summing over UICs; a UIC is a Naval installation, identified by a "unit identification code." See text for details.

8. Non-Integral Grade-by-Grade Allocation

At each stage, we apply the highest-grade men against the highest-grade p.r.a. not already filled; in effect, we place the p.r.a. in descending grade sequence, place all the available men in descending grade sequence, and match up those sequences -- highest man to highest p.r.a. and lowest man to lowest p.r.a. Unfortunately, this conceptually elegant process does not complete the actual allocation of men to slots; as mentioned above, the p.r.a. are in general not integers. We may regard this non-integral allocation as a provocative proposal, like King Solomon's proposed division of the child equally between the two women, which precedes and facilitates a realistic and equitable solution. Quota procedures prove to be unnecessary in converting this proposal into the required integral allocation.

To clarify the procedure so far, we present an example involving three UIC's whose allowances and priorities are given in Table II. The basic input data of the problem -- including the available resources -- are indicated with asterisks; the other numbers given there are intermediate results.

Numerical results applicable to those three UIC's will be given as triples of numbers; thus the relative priorities are (1.10, 1.50, 1.00) and the allowances are (10, 20, 14), so that the 39 available men must be allocated in proportion to the products (11.00, 30.00, 14.00). Exact proportionality would give (7.80, 21.27, 9.93) but UIC #2 must not receive more than its allowance in this undermanned community, so we give it 20.00 and share the remainder, producing the exact quotas (8.36, 20.00, 10.64); a quota process then produces quantitative allocations of (8, 20, 11), completing Phase I of the example.

		Grades of Available Men							
		G-7	G-6	G-5	G-4	G-3	G-2	G-1	Total
Pro-Rated Allowance By Grade	G-7	0.80							0.80
	G-6	0.20		1.59					1.79
	G-5			3.41	0.17				3.58
	G-4				2.83	1.34			4.17
	G-3					5.66	1.50		7.16
	G-2						7.50	5.04	12.54
	G-1							8.96	8.96
	Total	1.00	0.00	5.00	3.00	7.00	9.00	14.00	39.00

TABLE III

Non-Integral Allocation of Available Men
Against Pro-Rated Allowances

		Grades of Available Men						
		G-7	≥ G-6	≥ G-5	≥ G-4	≥ G-3	≥ G-2	≥ G-1
Cumulative Pro-Rated Allowance By Grade	G-7	0.80	0.80	0.80	0.80	0.80	0.80	0.80
	≡ G-6	1	1	2.59	2.59	2.59	2.59	2.59
	≡ G-5	1	1	6	6.17	6.17	6.17	6.17
	≡ G-4	1	1	6	9	10.34	10.34	10.34
	≡ G-3	1	1	6	9	16	17.50	17.50
	≡ G-2	1	1	6	9	16	25	30.04
	≡ G-1	1	1	6	9	16	25	39.00

TABLE IV

Non-Integral Allocation of Cumulative Available Men
Against Cumulative Pro-Rated Allowances

The cumulative allowances are then pro-rated, giving the rows labeled "C.P.R.A." in Table II; for UIC #1, the Cum. Allowances are multiplied by 8/10, for UIC #2 the Cum. Allowances are multiplied by 20/20, and for UIC #3 they are multiplied by 11/14. Then the "P.R.A." rows -- the pro-rated allowances -- are computed by differencing the preceding rows, and Phase II of the example is completed.

Now we attempt to match the sequence of total pro-rated allowances (in decreasing grade sequence) to that of the available men. We see from Table II that there are P.R.A.'s of 0.80 for G-7's, 1.79 for G-6's ..., 8.96 for G-1's, totalling 39.00 slots, and 1 available G-7, 1 G-6, ..., 14 G-1's (also totalling 39, of course). We allocate 0.80 of the top man to the P.R.A. for G-7's, the other 0.20 of him to the P.R.A. for G-6's, and fill the remaining 2.59 G-6 slots with available G-5's. The other 3.41 of the G-5's are appointed to fill G-5 slots, and the remaining 0.17 G-5 slot must be filled with G-4 men. Continuing in this way, we fill in Table III.

Table IV expresses the same information as Table III, but both grades of available men and grades of pro-rated allowances are cumulated, as indicated in the marginal labels for the rows and columns. Although Table III is easier to understand, Table IV is much easier to compute, as can easily be seen.

9. Integral Grade-by-Grade Allocation

Next we produce integral grade-by-grade allocations (which can actually be implemented), by simply rounding each entry in the Cumulative Non-Integral Grade-by-Grade Allocation array to the nearest integer. (For our example, we round Table IV to get Table V.) The result is then differenced to provide an Integral Grade-by-Grade Allocation, which shows how many men at each grade will be used to fill slots at each grade. (In the example, Table V is differenced to give Table VI, which reveals, for instance, that none of the available G-4's will be used to fill a slot at either the G-3 or G-5 level, but that 2 of the 8 slots at grade G-3 will be filled by G-2's.)

These calculations conclude Phase III. (Note that some of the entries in Table II have not yet been explained.)

10. Allocation of Men by Grade to UIC's

As mentioned above, we employ a sequence of several Quota apportionments to allocate men to the UIC's by grade. We do this by following the path of non-zero entries in the Integral Grade-by-Grade Allocation array from top left to bottom right; each such entry involves a separate stage of the computation, and an additional Quota apportionment. At each stage, the allocations of men to UIC's in the previous stages are taken as minima (considered as already committed), while the total allocations (as computed in Phase I) are taken as maxima, with the cumulative pro-rated allowances (down to the grade-level of the slots being

178/6

Grades of Available Men

	G-7	G-6	G-5	G-4	G-3	G-2	G-1
Cumulative Pro-Rated Allowance by Grade							
G-7	1	1	1	1	1	1	1
G-6	1	1	3	3	3	3	3
G-5	1	1	6	6	6	6	6
G-4	1	1	6	9	10	10	10
G-3	1	1	6	9	16	18	18
G-2	1	1	6	9	16	25	30
G-1	1	1	6	9	16	25	39

TABLE V

Integral Allocation of Cumulative Available Men
against Cumulative Pro-Rated Allowances

Grades of Available Men

	G-7	G-6	G-5	G-4	G-3	G-2	G-1	Total
Pro-Rated Allowance by Grade								
G-7	1							1
G-6			2					2
G-5			3					3
G-4				3	1			4
G-3					6	2		8
G-2						7	5	12
G-1							9	9
Total	1	0	5	3	7	9	14	39

TABLE VI

Integral Allocation of Available Men
against Pro-Rated Allowances

	Grades							Total
	G-7	G-6	G-5	G-4	G-3	G-2	G-1	
UIC #1								
Cum. Alloc.	1	1	2	3	5	5	8	
Alloc.	1	0	1	1	2	0	3	8
UIC #2								
Cum. Alloc.	0	0	3	4	8	14	20	
Alloc.	0	0	3	1	4	6	6	20
UIC #3								
Cum. Alloc.	0	0	1	2	5	6	11	
Alloc.	0	0	1	1	3	1	5	11

TABLE VII

Effect of Qualitative Priority Factor

filled) taken as the basis for proportionality. The sequential nature of this computation guarantees that, if a UIC happens to gain a fraction of a man in one stage, it will tend to lose a fraction in the succeeding stage, and vice versa. In this way we exploit the information in the Non-Integral grade-allocation matrix represented for our example by Table IV.

Continuing with the example begun above, we see from Table VI that one man of grade G-7 fills one slot of the same grade. From Table II we see that the three UIC's have respectively (0.80, 0.00, 0.00) slots at grade G-7; thus the one G-7 man goes to UIC #1.

The next entry in Table VI tells us that 2 G-5 men are to be assigned against pro-rated allowances of grade G-6. Table II gives C.P.R.A.'s, at grade G-6 and above, as (0.80, 1.00, 0.79) for the three UIC's, and we perform a second Quota computation which allocates these first three men as (1, 1, 1) -- one to each of the UIC's. In other words, those two G-5 men have filled one G-6 slot in UIC #2 and one G-6 slot in UIC #3.

The third entry in Table VI shows 3 G-5 men assigned to fill slots at the G-5 level. The new Quota computation partitions 6 men (the three previously allocated, and the three new men) against the total P.R.A.'s for G-5 and above -- of which there were (1.60, 3.00, 1.57) in the three UIC's. There is a clear though slight preference for assigning the sixth man to UIC #1 rather than to UIC #3, so the cumulative assignment is (2, 3, 1), and we see that one of those last three G-5's was assigned to UIC #1 and the other two to UIC #2.

As these computations proceed, we successively fill in the rows of Table II labelled "Cumulative Allocation"; of course, when we reach the last cell of Table VI we are assigning the last of the G-1 men to the last remaining slots, so the final entries in the "Cum.Aloc." row agree with the original result of Phase I -- the quantitative allocation to the UIC's.

Finally, the "Allocation" rows in Table II are found by differencing the preceding rows. This concludes the major portion of Phase IV.

11. Qualitative Priority

It has seemed reasonable to modify the process described above in one other respect, by introducing the notion of "Qualitative Priority". The purpose of this last twist in the complicated path is to ensure that, if a UIC happens to be deprived of a fraction in the quantitative allocation, it will be given the benefit of any flexibility of grade-allocation in Phase IV.

Instead of using the C.P.R.A. (cumulative pro-rated allowances) as the basis for proportionality in the successive stages of Phase IV, we shall use the products of those C.P.R.A.'s with a Qualitative-Priority factor, which is the square of the ratio of the original basis for proportionality (the product of Allowance by Priority) to the quantitative allocation actually made in Phase I.

In our example, the Phase-I proportions were (11, 30, 14) while the Phase-I allocations were (8, 20, 11); thus the qualitative-priority factors are the squares of (11/8, 30/20, 14/11) -- approximately (1.89, 2.25, 1.62), showing that Phase I treated UIC #3 relatively well, while UIC #2 was penalized by its not being permitted to exceed its allowance of 20. Incorporating this factor into our computation, the Allocation rows of Table VII result. Note that the allocations of Table II are changed slightly -- both UIC #1 and UIC #2 are slightly upgraded, while UIC #3 is somewhat downgraded.

We should note that, generally, the effect of the qualitative priority factor is this: a UIC which gained a fraction in Phase I will tend to have each of its pro-rated allowances filled with the lowest grade of man who fills any slots at that grade, while a UIC which lost a fraction in Phase I will tend to have each of its pro-rated allowances filled with the highest grade of man who fills any slots at that grade level.

Sections 10 and 11 constitute Phase IV, the final phase of our proposed allocation-process.

12. Which Quota Method?

As mentioned above, references [QMA] and [QMNU] have introduced new methods of apportionment, in addition to those previously described and investigated in refs. [MTAR], [ARC], and earlier works (ref. QMA has a bibliography of earlier methods.)

For reasons given above, we recommend that the MFQ ("Major Fractions Quota") method be used in the quantitative-allocation phase; it lies midway between the two most extreme of the new Quota methods, and thus can justify the claim that it does not tend to favor either the smaller UIC's or the larger. (For example, the Quota method of [QMA], called Primal Quota herein, if used to allocate a slightly undermanned community would conclude that many -- perhaps all-- of the UIC's were tied for their last man; thus a community with one man short would receive no guidance at all as to which UIC should be deprived of a man. Similarly, the Dual Quota method if used for a community which was slightly overmanned, would find so many ties that it would be of little value.) MFQ has at least the advantage that it will be as distant as possible from both those risks.

When Phase IV is performed, we recommend that the Dual Quota method be employed for guiding the grade-allocation among UIC's. Under the Dual Quota method, every UIC will be tied for the first man to be allocated, and there will be a strong tendency for every UIC to receive one man before any receives its second (though that tendency will not overcome a large discrepancy in the ideal ratio of their manning.) As a result,

when the higher-grade availables are allocated (prior to the lower-grade availables) there will be a tendency for each UIC to receive at least one man at the highest grade which occurred in its allowance; this will have the desirable consequence that the "team leader", or highest-grade slot in an UIC, will tend to be filled with a man of the appropriate grade. The attractiveness of this result, in the context of a team of specialists, is evident, and would conform with a policy goal in the allocation of U.S. Navy enlisted men.

For these reasons we propose that Dual Quota should be used for Phase IV of the above procedure, even though MFQ is used in Phase I.

The detailed algorithms for both Dual Quota and MFQ are given above.

13. Conclusion

This paper has proposed a procedure for allocating available Naval manpower resources to requirements, in case where the tasks are all similar in nature, but where a hierarchy of skill-- or training--levels exists. The procedure depends on the availability of one or more appropriate apportionment methods, such as have been recently developed for legislative apportionment. The qualitative acceptability of the procedure is currently being evaluated in trials which are performing the computations necessary to allocate the Navy's entire enlisted force.

APPENDIX I

METHODS OF APPORTIONMENT

I.1 The Apportionment Problem

The apportionment problem is to divide a given integer h , called the house, into s integral portions $a_1, a_2, \dots, a_s$, respectively proportional to s given numbers $p_1, \dots, p_s$, subject to over-riding minima $r_1, \dots, r_s$ and maxima $b_1, \dots, b_s$. If the portions could be non-integral, an easy calculation (formalized below) would determine numbers, called the exact quotas, which would give the correct apportionment. But the portions must be integers, and therefore suitable integers must be used to approximate the exact quotas; we may think of an apportionment method as an effective interpretation of the words "suitable" and "to approximate".

This problem, which must be solved decennially to determine the number of seats for each state in the U.S. House of Representatives, has an extensive history; ref [QMA] provides a fascinating introductory survey. We shall therefore use terminology appropriate to that specific application, and call the given numbers $p_1, \dots, p_s$ the populations of the respective states 1, 2, ..., s .

I.2 The Work of Balinski and Young

In refs [N M A] and [Q M A], Balinski and Young showed that a new apportionment method, the Quota Method, satisfied three axioms intended to summarize the essential desiderata. The first axiom -- "monotonicity" -- excludes the Alabama paradox (see I.5 below). The second axiom -- "the quota condition" -- limits the discrepancy between the exact quota and any acceptable apportionment. Their third axiom -- "consistency" -- excludes capricious or discriminatory methods. Balinski and Young then proved that the Quota method is the unique method to satisfy all three axioms.

I.3 Notation and Definitions

Bold-face letters denote s -tuples of real numbers indexed by i , where i is restricted to be one of the integers $1, 2, \dots, s$, and all summations are over i . An apportionment problem is a set $(\mathbf{p}, \mathbf{r}, \mathbf{b}, h)$ as above, with p_i, r_i, b_i , and h integral, $p_i > 0$, $0 \leq r_i \leq b_i$, and $\sum r_i = h_* \leq h \leq h^* = \sum b_i$. A problem-set $(\mathbf{p}, \mathbf{r}, \mathbf{b})$ is the set of all apportionment problems which share the given values of $\mathbf{p}, \mathbf{r}$, and $\mathbf{b}$. An apportionment for the problem $(\mathbf{p}, \mathbf{r}, \mathbf{b}, h)$ is an s -tuple $\mathbf{a} = (a_1, \dots, a_s)$ of integers called portions, with $r_i \leq a_i \leq b_i$ for each i , and $\sum a_i = h$.

An apportionment solution is a function f , which to any such problem assigns an apportionment $\underline{a} = f(\underline{p}, \underline{r}, \underline{b}, h)$. Note that the "pure" problem, without maxima or minima, may be identified with the special case $\underline{r} = (0, \dots, 0)$ and $\underline{b} = (h, h, \dots, h)$. An apportionment method is a non-empty set of solutions. A method M is called monotone if, for every problem and any f in M , $f(\underline{p}, \underline{r}, \underline{b}, h) \leq f(\underline{p}, \underline{r}, \underline{b}, h+1)$ unless $h = h^*$ so that the right side is undefined.

We define the exact quotas q_i by

$$q_i = \max(r_i, \min(b_i, \lambda \cdot p_i)), \text{ where } \lambda = \lambda(h) \text{ is}$$

$$\text{chosen so that } \sum q_i = h.$$

Since $g(\lambda) = \sum \max(r_i, \min(b_i, \lambda \cdot p_i))$ is a continuous non-decreasing function of λ , with $g(0) = h_*$ and $g(\lambda) = h^*$ for λ large enough, the q_i defined above are unique (even though $\lambda(h)$ may not be.) Lower quotas and upper quotas are defined respectively by $\ell_i = \lfloor q_i \rfloor$ (the greatest integer not exceeding q_i) and $u_i = \lceil q_i \rceil$ (the least integer not less than q_i .) An apportionment method is said to satisfy lower quota if always $f(\underline{p}, \underline{r}, \underline{b}, h) \geq \underline{\ell}(\underline{p}, \underline{r}, \underline{b}, h)$; to satisfy upper quota if always $f(\underline{p}, \underline{r}, \underline{b}, h) \leq \underline{u}(\underline{p}, \underline{r}, \underline{b}, h)$; and to satisfy quota if both conditions hold. There is excellent justification, in the context of the manpower allocation problem, for requiring that an apportionment method should always satisfy quota; e.g., if the "fair share" of a UIC is 3.89 men from a certain rating, any method which failed to satisfy quota by allocating fewer than 3 or more than 4 men would surely be unacceptable.

Note that the quota condition defined by Balinski and Young in $[QMA]$ and $[NMA]$ differs in several respects from the condition defined above: (i) their formal treatment disregarded the possible importance of upper bounds $\underline{b}$ on the

portions; (ii) they did not define an exact quota for the constrained problem -- i.e., the problem with lower bounds $\underline{r}$ on the portions; (iii) their definition of upper and lower quotas for the constrained problem permitted a difference of more than 1 between them (as shown by their example, Table 6, p. 718, of [QMA].) As pointed out in [QMNU], the distinction actually causes a difference of one seat in the apportionment for their hypothetical 1984B populations of the U.S.

1.4 Duality and the Quota Method

Once maxima as well as minima are considered in apportionment problems, a duality can be defined in which: maxima correspond with minima; an upward induction (using house sizes increasing from h_* to h^*) corresponds with a downward induction; "greater than" corresponds with "less than"; and lower quota corresponds with upper quota.

Under this duality, the Quota Method of Balinski and Young will correspond with another algorithm, which I have called the Dual Quota (DQ) Method, and which will have exactly as much basis for acceptability as the Quota Method -- which will henceforth be called the Primal Quota (PQ) method. Similarly, from the proof that the Primal Quota method is the unique method satisfying the three axioms of Balinski and Young we can produce a proof that the Dual Quota method is the unique method satisfying three equally-reasonable axioms -- viz, monotonicity, quota, and a "dual-consistency" condition which is dual to their "consistency" condition.

Examples have shown that Primal Quota and Dual Quota solutions will often differ; Appendix III below shows why it is unlikely for them to agree. Furthermore, we have devised algorithms which permit us to define a spectrum of methods, each of which satisfies both quota and monotonicity, and each of which avoids capricious or arbitrary behavior by employing a divisor function -- see section I.6 below. Those other methods will be described in Appendix II.

I.5 The Alabama Paradox

The importance of monotonicity can be seen by considering the "Vinton" or "Hamilton" method -- viz: "Give each state its lower quota, and give one more seat to each of the $h - \sum \ell_i$ states with greatest remainders $q_i - \ell_i$." The pure problem with $\underline{p} = (1, 3, 3)$ makes $q(3) = (3/7, 9/7, 9/7)$ and $q(4) = (4/7, 12/7, 12/7)$ and thus produces the apportionments (1, 1, 1) when $h = 3$, and (0, 2, 2) when $h = 4$; the first state loses a seat when the house increases from 3 to 4. Much colorful discussion has made clear that the Congress finds this "Alabama Paradox" (named in honor of its first victim) unacceptable, and consequently the Vinton method, although frequently used between 1850 and 1910, is no longer seriously proposed for the apportionment of Congress. Neither do we propose to use it in our manpower allocations.

I.6 Huntington's Work: Divisor Methods

Any monotone solution $\underline{f}$ may be characterized by identifying, for any $\underline{p}$, $\underline{r}$, and $\underline{b}$, the sequence in which the states successively gain seats as the house increases from h_* to $h_* + 1, \dots, h^*$. The divisor methods comprise a family of monotone methods, each of which is defined by a divisor function $d(a)$; the state which gains the $(h + 1)$ 'th seat is one which achieves the minimum of $d(a_i)/p_i$, where $\underline{a} = (a_1, a_2, \dots, a_s) = \underline{f}(\underline{p}, \underline{r}, \underline{b}, h)$ is the apportionment for a house of size h . (Huntington considered the maximum of $p_i/d(a_i)$; we have reformulated the condition to avoid division by zero.)

A reasonable divisor function d must be a monotone-increasing function, and must satisfy $a \leq d(a) \leq a + 1$ for all non-negative integral a . (See Theorem 3 of Appendix III.)

Huntington's paper, ref. [ARC], describes five methods, corresponding to these five divisor functions:

- (i) $d(a) = a$, called by Huntington the method of Smallest Divisors (SD);
- (ii) $d(a) = a + 1$, apparently first devised by Thomas Jefferson but known as the method of d'Hondt, and called by Huntington the method of Greatest Divisors (GD);
- (iii) $d(a) = a + \frac{1}{2}$, apparently due to Daniel Webster but often called the method of Major Fractions (MF);
- (iv) $d(a) = 2a(a + 1)/(2a + 1)$, called the method of the Harmonic Mean (HM); and
- (v) $d(a) = \sqrt{a(a + 1)}$, called by Huntington the method of Equal Proportions (EP).

Note that the first two represent the extreme possibilities, and the last three are respectively the arithmetic, harmonic, and geometric means between those extremes.

Although other functions, such as $d(a) = 2a + 1 - \sqrt{a(a+1)}$, $d(a) = 0.3a + 0.7 \sqrt{a(a+0.5)}$, et al., could be used for the divisor function $d(a)$, only the above five methods seem to have been seriously considered between the publication of Huntington's analysis [MTAR] in 1921 and the appearance of [NMCA] in 1974. The latter paper stated that no Huntington method satisfies quota, a result which was apparently not emphasized in the many congressional discussions between 1921 and the passage in 1941 of P.L. 291, "An Act to Provide for Apportioning... Congress ... by the equal proportions method".

In fact, GD is the only one of the above five methods -- and probably the only divisor method -- which satisfies lower quota; and, dually, SD is the only one of the five -- and probably the only divisor method -- which satisfies upper quota. (See Theorem 2, ref. [QMA] . The proof, which was omitted, is straightforward -- even when upper and lower bounds are included, and even when the above narrower definition of "upper quota" is used.)

I.7 Consistency

We express in our notation some salient portions of the definition of "consistency" from p. 4604 of ref. [NMCA].

For any solution f , we define the eligible set at $h + 1$ to be the set $E(h + 1)$ = $\left\{ i \mid f_i(h) < q_i(h + 1) \right\}$ of states which could receive the $(h + 1)$ 'th seat without violating upper quota; $f_i(h)$ is the previous portion of the i 'th state.

A monotone method M is called consistent if the choice of state to receive the $(h + 1)$ 'th seat is governed entirely by priority among the states eligible at $(h + 1)$, where relative priority between any two states is determined only by their populations and (immediately) previous portions. Without completing the details, it is clear that this consistency condition explicitly protects the solution from violating upper quota, but not from violating lower quota; it is thus rather natural that the Balinski-Young Quota method -- the only method consistent in the above technical sense which is monotone and satisfies quota -- is related to GD, which intrinsically satisfies lower quota.

In fact, the (Primal) Quota method assigns the next -- i.e., the $(h + 1)$ 'th-- seat to a state which achieves the minimum of $(a_i + 1)/p_i$, where the minimum is taken over the set $E(h + 1)$ of those states eligible to receive the $(h + 1)$ 'th seat without violating upper quota. Thus GD and Primal Quota place the states in the same priority order, but GD does not impose the eligibility requirement.

The proof that the PQ method satisfies upper quota is easy; it is only necessary to observe that $E(h+1)$ can never be empty. Although it is "natural" that PQ "inherits" from GD the property of satisfying lower quota, the proof, that PQ actually does satisfy lower quota, is far from trivial.

I.8 The Dual Quota Method

The concepts of dual-eligibility and dual-consistent can be derived from the above definitions; if we consider the sequence of states which lose seats as the house decreases from $h^* - 1, \dots, h_*$, we define the dual-eligible set at h as the set $E'(h) = \left\{ i \mid f_i(h+1) \geq q_i(h) \right\}$ of states which could lose the $(h+1)$ 'th seat without violating lower quota. We then define a solution f as dual-consistent if the choice of losing state is governed by priority within the dual-eligible set, where relative priority of two states is determined by their populations and previous (i.e., at the next-higher house-size) portions. It is natural that in this case an analog of the method of Smallest Divisors (which intrinsically satisfies upper quota) has the desired properties. The modified proof mentioned above may be translated mechanically into a proof of the following

Theorem: There exists a unique dual-consistent house-monotone method satisfying quota.

That method is called the Dual-Quota method (DQ), and is defined in section II.3 of Appendix II.

APPENDIX II

DESCRIPTION OF QUOTA METHODS

II.1 The Classical Quota Methods

We give first the defining algorithms for the Balinski-Young Primal Quota Method (introduced in [N M C A] , which exposed the possibility of a method being both monotone and quota) and the Dual Quota Method. The former builds upward from h_* , and the latter builds downward from h^* .

II.2 The Primal Quota Method

The Primal-Quota (PQ) method is the set of all solutions φ defined as follows:

- (i) $\varphi_i (p, r, b, h_*) = r_i$ for all i ;
- (ii) Given $a_i = \varphi_i (p, r, b, h)$ for some h with $h_* \leq h < h^*$, define $E(h+1)$ as the non-empty set $\left\{ i \mid a_i < q_i(h+1) \right\}$, and let k be some member of $E(h+1)$ such that $(a_k + 1)/p_k = \min \left((a_i + 1)/p_i \right)$, where the minimum is taken over $E(h+1)$; then we set $\varphi_k (p, r, b, h+1) = a_k + 1$, $\varphi_i (p, r, b, h+1) = a_i$ for all $i \neq k$.

Various solutions of the PQ method result, according to how k is selected in the case of a tie for the minimum.

II.3 The Dual-Quota Method

The Dual-Quota method (DQ) is the set of all solutions $\mathcal{Y}$ defined as follows:

- (i) $\mathcal{Y}_i(\underline{p}, \underline{r}, \underline{b}, h^*) = b_i$ for all i ;
- (ii) Given $a_i = \mathcal{Y}_i(\underline{p}, \underline{r}, \underline{b}, h)$, we define the set $E'(h-1) = \left\{ i \mid a_i > q_i(h-1) \right\}$; let k be a state in $E'(h-1)$ such that $(a_k - 1)/p_k = \max \left\{ (a_i - 1)/p_i \right\}$, where the maximum is over $E'(h-1)$; then we set $\mathcal{Y}_k(\underline{p}, \underline{r}, \underline{b}, h-1) = a_k - 1$, $\mathcal{Y}_i(\underline{p}, \underline{r}, \underline{b}, h-1) = a_i$ for $i \neq k$.

Various solutions of the DQ method result, depending on how k is selected when two or more states tie for the maximum.

Note that, in the DQ algorithm, we are deciding whether the i 'th state shall retain its a_i seats, or only $a_i - 1$; thus, evaluating the divisor-function at the smaller of the two portions being considered, we get $d(a_i - 1) = a_i - 1$. In the PQ algorithm, while deciding whether or not to add another seat to the a_i which it had previously, we were also evaluating the divisor-function at the smaller of the two portions being considered, obtaining $d(a_i) = a_i + 1$.

II.4 Direct Quota Methods

The Primal Quota method, by its definition of the eligible set $E(h+1)$, explicitly prevents any solution $\underline{Y}$ from violating upper quota. It is stated without proof in [NMCA] and [QMA] that GD satisfies lower quota; the proof is not difficult. One may say that any solution $\underline{Y}$ of PQ satisfies lower quota "because of" the similarity between PQ and GD; in any case, [QMA] proves that PQ does satisfy quota. (That proof is far from trivial; our inclusion of upper limits, and the slight change in our definition of lower quota, requires some added complication but no essential change in the proof.)

Similarly, the Dual Quota definition explicitly precludes any solution $\underline{Y}$ of DQ from violating lower quota, and (perhaps "because of" the similarity between DQ and SD) such a $\underline{Y}$ cannot violate upper quota either.

Let us now generalize the definition of II.2. The direct ascending quota method with divisor d is the set of solutions $\underline{Y}$ defined recursively as follows:

- (i) $\mathcal{Y}_i(\underline{p}, \underline{r}, \underline{b}, h_*) = r_i$;
- (ii) Given $a_i = \mathcal{Y}_i(\underline{p}, \underline{r}, \underline{b}, h)$ for $h_* \leq h < h^*$, define the set $S(h+1)$ of supereligible states at $h+1$ to be the set of states which could receive the $(h+1)$ 'th seat without violating quota, and let k be a state in $S(h+1)$ such that $d(a_k)/p_k = \min d(a_i)/p_i$, where the minimum is taken over $S(h+1)$. Then define $\mathcal{Y}_k(\underline{p}, \underline{r}, \underline{b}, h+1) = a_k + 1$, $\mathcal{Y}_i(\underline{p}, \underline{r}, \underline{b}, h+1) = a_i$ for $i \neq k$. If at any stage the set $S(h+1)$ is empty, we say that the direct ascending quota (d.a.q) method fails for that divisor-function d and that problem $(\underline{p}, \underline{r}, \underline{b}, h)$.

It is clear that the d.a.q. method with divisor-function $d(a) = a + 1$ is simply the PQ method, and we know that $S(h+1)$ is never empty in that case. Unfortunately, we have acquired empirical evidence supporting the

Conjecture: Every d.a.q. method, except the PQ method, fails for some problem.

Discussion: We can show that, for any divisor of the form $d(a) = a + c$ with $0 \leq c < 1$, there is a problem which results in an empty $S(h+1)$ at some stage; it suffices to take the pure problem with two components of p having the values x , and $2x$ components having the values 1, where x is an integer exceeding $(1 - c)^{-1} + 2$. The only standard methods which do not have divisor-functions of the form $a+c$ are HM and EP, and both their divisor-functions succumb to the same examples with $c = 1/2$. There remains the possibility that some baroque divisor-function, such as those mentioned at the end of section II.6 above, might happen to allow a d.a.q. method.

One can define the dual concept of a direct descending quota (d.d.q.) method with divisor d ; the only caveat is that we must use $d(a, -1)$, as mentioned in section II.3 above. We note that the d.d.q. method with divisor-function $d(a) = a$ is simply the DQ method, for which the analogous set $S'(h-1)$ is never empty. Dual to the previous conjecture is the

Conjecture: Every d.d.q. method, except DQ, fails for some problem.

It is clear that the d.a.q. method with divisor-function $d(a) = a + 1$ is simply the PQ method, and we know that $S(h+1)$ is never empty in that case. Unfortunately, we have acquired empirical evidence supporting the

Conjecture: Every d.a.q. method, except the PQ method, fails for some problem.

Discussion: We can show that, for any divisor of the form $d(a) = a + c$ with $0 \leq c < 1$, there is a problem which results in an empty $S(h+1)$ at some stage; it suffices to take the pure problem with two components of p having the values x , and $2x$ components having the values 1, where x is an integer exceeding $(1 - c)^{-1} + 2$. The only standard methods which do not have divisor-functions succumb to the same examples with $c = 1/2$. There remains the possibility that some baroque divisor-function, such as those mentioned at the end of section II.6 above, might happen to allow a d.a.q. method.

One can define the dual concept of a direct descending quota (d.d.q.) method with divisor d ; the only caveat is that we must use $d(a_i - 1)$, as mentioned in section II.3 above. We note that the d.d.q. method with divisor-function $d(a) = a$ is simply the DQ method, for which the analogous set $S'(h-1)$ is never empty. Dual to the previous conjecture is the

Conjecture: Every d.d.q. method, except DQ, fails for some problem.

In short, if we attempt to generalize the notion of consistency (in the sense of refs [NMCA] and [QMA]), and to define a direct quota method to be either a d.a.q. or a d.d.q. method, we do not find any new apportionment methods which satisfy quota and are monotone.

II.5 Other Quota Methods

If, in spite of the sombre conjectures of the preceding section, we attempt to define an ascending quota method for some divisor function d other than $d(a) = a+1$, we find, for many problems, that a sequence of non-empty supereligible sets $S(h_1+1)$, $S(h_2+1)$, ..., $S(h^*)$, is generated, so that a systematic monotone solution is generated which satisfies quota while showing great similarity to one of the Huntington methods. In order to have a method which is always applicable we must allow for the eventuality that the set $S(h+1)$ is empty. When such an eventuality arises, the previous apportionment cannot be extended to an apportionment of the next-higher house without violating monotonicity, and therefore must be excluded from the solution which is being generated by the algorithm; in other words, we must simply back-track, choose the next-best in place of the minimum previously chosen, and attempt to proceed. ("Next-best" means by "best" the extremum being sought -- minimum of $d(a_i)/p_i$ for the ascending methods.)

Since there do exist solutions of the type we seek -- monotone methods which satisfy quota -- it is clear that we must eventually find one. Even very simple examples show that, in general, different divisors will result in different methods; Appendix III explains some of those differences. Although we have as yet few theorems limiting the amount of "backtracking" which might be needed, in practice there is rarely any backtracking at all. In any case, the solutions found in this way are neither arbitrary nor capricious, and clearly deserve to be called systematic, even though they do not satisfy the Balinski-Young definition of "consistency".

This discussion was intended to motivate an algorithm which we shall proceed to define, after some further preliminary definitions.

Any monotone apportionment solution $\underline{f}$, applied to the problem-set $(\underline{p}, \underline{r}, \underline{b})$ -- by which we mean the set of problems $(\underline{p}, \underline{r}, \underline{b}, h)$ as h ranges from $h_* = \sum r_i$ to $h^* = \sum b_i$ -- determines a sequence, to be called the gaining-state index-sequence $\text{GSIS}(\underline{f}, \underline{p}, \underline{r}, \underline{b})$, which is a sequence of length $h^* - h_*$ whose j th term indicates the state which gains a seat when the house increases from h_{*+j-1} to h_{*+j} . Since $\underline{f}(\underline{p}, \underline{r}, \underline{b}, h_*) = \underline{r}$ and $\underline{f}(\underline{p}, \underline{r}, \underline{b}, h^*) = \underline{b}$, we see that $\text{GSIS}(\underline{f}, \underline{p}, \underline{r}, \underline{b})$ must contain exactly $b_i - r_i$ occurrences of the index i .

Not only does such an $\underline{f}$ determine a GSIS, but it is clear that we can specify any monotone apportionment function $\underline{f}$ entirely by giving the mapping from the set PS of all problem-sets to the set of positive finite-dimensional integral vectors.

That mapping may be arbitrary, provided it assigns to any $(\underline{p}, \underline{r}, \underline{b})$ a GSIS which includes exactly $b_i - r_i$ instances of the index i . Thus the problem of defining an apportionment function may be aided by defining a GSIS for each problem-set in some reasonable way. Furthermore, whether or not $\underline{f}$ satisfies quota when applied to a particular problem-set $(\underline{p}, \underline{r}, \underline{b})$ can be determined from the properties of the GSIS $(\underline{f}, \underline{p}, \underline{r}, \underline{b})$; it is merely necessary to check the number of occurrences of the index i in the first j terms of the GSIS, for each i and j , against the exact quota at house size $h_* + j$ for the i 'th state. We thus define, for each problem-set $(\underline{p}, \underline{r}, \underline{b})$, the set QIS $(\underline{p}, \underline{r}, \underline{b})$ of quota index-sequences for $(\underline{p}, \underline{r}, \underline{b})$, which includes exactly those GSIS which define apportionment functions satisfying quota.

Now we define a relation called d-precedence (which depends on the divisor function d) between elements σ, σ' of QIS $(\underline{p}, \underline{r}, \underline{b})$, as follows:

σ d-precedes σ' if

- (i) $\sigma_j = \sigma'_j$ for all $j=1, 2, \dots, \ell$,
- (ii) $\sigma_{\ell+1} \neq \sigma'_{\ell+1}$,
- (iii) $d(a_k)/p_k < d(a_{k'})/p_{k'}$, where $k = \sigma_{\ell+1}$, $k' = \sigma'_{\ell+1}$

and $\underline{a}$ is the apportionment of $(\underline{p}, \underline{r}, \underline{b}, h_* +)$ induced by the initial sequence common to σ and σ' .

(The relation is clearly transitive, irreflexive, and therefore acyclic.)

Now we define an ascending quota solution (a.q.s.) with divisor d as a function $\underline{f}$ which assigns to every problem $(\underline{p}, \underline{r}, \underline{b}, h)$ an apportionment determined by applying to the problem-set $(\underline{p}, \underline{r}, \underline{b})$ some quota index-sequence σ of QIS $(\underline{p}, \underline{r}, \underline{b})$ which is not d -preceded by any other σ' of QIS $(\underline{p}, \underline{r}, \underline{b})$. We define an ascending quota method with divisor d as a non-empty set of a.q.s with the same divisor, and denote such a method by $AQM(d)$. Such methods must exist for any divisor-function d , since no QIS is empty and d -precedence is acyclic. In particular, if d is defined by $d(a) = a + 1$, the $AQM(d)$ is PQ.

Given any problem $(\underline{p}, \underline{r}, \underline{b}, h)$ and any divisor-function d , an apportionment $\underline{f}(\underline{p}, \underline{r}, \underline{b}, h)$ belonging to some solution $\underline{f}$ in $AQM(d)$ may be computed by this

Ascending Quota Algorithm:

- (i) Initially define σ to be an empty sequence, its length ℓ to be 0, the house-size h' to be h_* , and the current apportionment

$$\underline{f}(\underline{p}, \underline{r}, \underline{b}, h') \text{ to be } \underline{f}(\underline{p}, \underline{r}, \underline{b}, h_*) = \underline{r}.$$

- (ii) Defining $\underline{a} = \underline{f}(\underline{p}, \underline{r}, \underline{b}, h')$, compute the exact quotas $q(\underline{p}, \underline{r}, \underline{b}, h'+1)$ and find the supereligible set $S(h'+1)$ of states which could receive the $(h'+1)$ 'th seat without violating quota. Define $\text{Seq}(h'+1)$ to be the sequence of indices of $S(h'+1)$, arranged in non-decreasing order of $d(a_i)/p_i$; $\text{Seq}(h'+1)$ will be empty if $S(h'+1)$ is empty. Set $n(h'+1) = 1$.

- (iii) If Seq $(h'+1)$ has fewer than $n(h'+1)$ terms, proceed to (vi); otherwise, let k denote the $n(h'+1)$ 'th term.
- (iv) Augment the sequence σ with the index ℓ_k , increase by 1, increase h' by 1, and define $f_k(\underline{p}, \underline{r}, \underline{b}, h')$: a_{k+1} , $f_i(\underline{p}, \underline{r}, \underline{b}, h') = a_i$ for $i \neq k$.
- (v) If $h' < h^*$, go to step (ii); if $h' = h^*$, go to step (vii).
- (vi) Since it is impossible to extend the current sequence σ while remaining within QIS $(\underline{p}, \underline{r}, \underline{b})$, we must delete the last index of σ , decrease its length ℓ by 1, decrease h' by 1, and then increase $n(h'+1)$ by 1 before returning to step (iii).
- (vii) The desired apportionment is then found by employing the first $h-h_*$ terms of σ to define $\underline{f}(\underline{p}, \underline{r}, \underline{b}, h)$.

II.6 Comments on the Ascending Quota Methods

A dual concept -- the Descending Quota Methods -- may be defined by an analogous algorithm, which begins with house size h^* and works downward, withdrawing a seat from the supereligible state which maximizes the criterion $d(a_i-1)/p_i$, backtracking as necessary when an empty supereligible set is reached, and finally ending with house size h_* . If the selected state is always appended

to the beginning of the state index-sequence, we will terminate, for each problem-set (p, r, b), with a sequence belonging to QIS (p, r, b) -- in fact, with a sequence which satisfies an extremum condition dual to the condition used above to define an ascending quota solution and an Ascending Quota Method. The DQM always attempt to maximize the criterion d/p for the rightmost state-index of the QIS set, while the AQM attempt to minimize the same criterion for the leftmost state-indexes of the same set. Because the total number of appearances of each index in any QIS is fixed, these criteria are not in direct conflict; but empirical evidence strongly suggests that we do not need to choose between the methods AQM(d) and DQM(d), because they seem to be identical. That conjecture has so far been proven only for some special cases -- e.g., for problems with three states and one of the five divisor-functions described by Huntington, and for $d(a) = a + 1/2$.

Having found one apportionment in AQM(d), we can proceed to find all others by reconsidering all cases in which a tie occurred for the choice of gaining state; if there were no such ties, the apportionment is unique.

In the legislative apportionment problem, exact ties are so unlikely that they are merely nuisances in the mathematical theory; but in the manpower allocation problem, ties are of crucial importance.

In either problem, the incidence of ties may be greatly reduced by redefining the relation of d -precedence so that, among states with equal ratios d/p , the state with larger population should gain the seat first (or, lose it last).

Numerical calculations should be arranged so that exact integers are employed in lieu of approximate ratios -- this is again of greater importance in the manpower allocation problem, where exact ties are more likely.

In the unlikely event that an attractive Huntington method is found which is not a divisor method, appropriate modifications could be made in the above definitions.

Although the computation of an AQ M(d) solution is effective, it is long and seems very inefficient if h^* is much greater than h . Some conjectures and theorems are provided in Appendix III, which promise to reduce the labor of computing these new Quota solutions.

Since the computational methods provided do all make explicit use of the upper bounds $\underline{b}$ on the portions of the several states, it is not self-evident that the introduction of those upper bounds (for a problem which did not naturally have upper bounds) will have no influence on the answer. The Constitutional upper limit on the size of a state's delegation ("..shall not exceed one for every thirty thousand [of population]") does provide such a natural upper limit in case of the apportionment of the U.S. House of Representatives, and Appendix III settles a corresponding question for manpower allocations, but the situation will remain slightly incomplete until conjectures of Appendix III are proven.

Thus, corresponding to each of Huntington's "workable methods", and to any other divisor function, we have defined a similar method which satisfies quota and is monotone. (In fact, if the AQ M(d) and DQ M(d) are found to be not always identical, there will sometimes be two analogs for each Huntington method!)

Appendix III will discuss the advantages and disadvantages of each of those, in the context of the legislative and the manpower applications.

APPENDIX III

THEOREMS ON APPORTIONMENT

III.1 General

In this appendix, we provide some definitions, conjectures, and theorems. We define the linear divisor functions and the linear divisor methods, which seem to have more tractable properties than other Huntington methods. We show that our Primal Quota method (which differs from the Quota method of Balinski and Young only in having an altered definition of "quota" and including upper bounds on the portions) satisfies their axioms, and give a detailed proof that the new Dual Quota method (see II.3 above and Mayberry [6].) satisfies the duals of those axioms.

We show that, under certain circumstances (which we hope to widen by more general theorems), the upper bounds $\underline{b}$ and lower bounds $\underline{r}$ will have no effect on an apportionment. We show that linear divisor methods are periodic in the house size h , that the only important non-linear methods (HM and EP) are ultimately periodic, and that the Quota analogs of all those methods (defined in section II.7 above) share the same periodicity properties. We show that those same methods will give rise to "ties" at predictable stages within those periods.

Finally, we offer the opinion that the manpower allocation process should use the MFQ method (a Quota method related to the method of Major Fractions), and that the appointment of Congress should either use MFQ or the Quota analog EPQ of the presently-approved EP method.

We have two definitions for the quota analog of each divisor method; only in the self-dual case of MFQ can we prove that they are equivalent. However, even if examples can be constructed for which the two definitions lead to different results, there is ample empirical evidence for the contention that they are usually the same; and both methods will satisfy quota, avoid the Alabama paradox, and closely resemble the corresponding Huntington method.

We have conjectures about the similarity of the new quota methods to the previously-known Huntington methods in cases where the results of the latter satisfy quota, but few proofs; we also remark that the computations require unusual attention to detail because of the algorithms' susceptibility to round-off error and because of the tediously recursive form of the basic definitions. (The magnitude of the computational task is not a significant factor in selecting an algorithm for the apportionment of Congress, but will be extremely important when a Quota method is selected to allocate manpower in the U.S. Navy.)

III.2 Restrictions on Divisor Functions

A divisor method of apportionment according to Huntington is a monotone method, defined with the aid of a divisor function d , where the state which gains the $(h+1)$ 'th seat is the state achieving the maximum of $p_i/d(a_i)$; here a denotes the apportionment at house size h . Recalling that the purpose of the divisor function is to facilitate making the portions nearly proportional to the respective populations, it is certainly natural to require that, as the house size increases without limit, the ratio a_1/a_2 of the portions of two specified states shall tend to the ratio p_1/p_2 of their populations. This implies that the ratio $d(a)/a$ should tend to a non-zero limit as $a \rightarrow \infty$.

Since multiplying the divisor-function by a constant does not change the resulting apportionments, we may, without loss of generality, assume that $d(a)/a \rightarrow 1$ as $a \rightarrow \infty$; this is equivalent to assuming that $d(a) = a + c(a)$, where $c(a)/a \rightarrow 0$ as $a \rightarrow \infty$.

Now let us also insist that a divisor-function d shall not preclude the attainment of exact proportionality, whenever the house size and populations permit it. More formally, we define a divisor-function d to be acceptable if, given integers x and y with $x/y = p_1/p_2$, state 2 will never receive its $(y+1)$ 'th seat until state 1 has received its x 'th seat. This imposes constraints on d , which does not seem to have been mentioned by Huntington in [4] and [5] nor by Balinski and Young in [3a], viz:

Theorem 1: The divisor function d is acceptable if both

- (i) $0 \leq c(a) \leq 1$ for all a , and
- (ii) if there is an a_1 with $c(a_1) = 0$, then there is no a_2 with $c(a_2) = 1$.

Proof: (Note that Huntington's use of the terms "greatest divisors" and "smallest divisors" suggests that he may have been aware of these conditions, although the author considers it more likely that those names refer only to the fact that, in considering an increase of a portion from a to $a+1$, the divisor to represent "current portion" could reasonably be taken as either a , or $a+1$, or some compromise between them.)

Suppose that, for some integer $y > 0$, we have $c(y) < 0$. (Note that $d(a)$ must surely be ≥ 0 , so $c(0) \geq 0$.) Then choose $p_2 = y$, and define $t = d(y)/y$. Since $t < 1$, and $d(a)/a \rightarrow 1$ as $a \rightarrow \infty$, we see $d(a)/(a+1) \rightarrow 1$, and we can find z large enough that $d(z)/(z+1) > t$. Now, choosing $p_1 = x = z+1$, we find of course that $x/y = p_1/p_2$; but $d(y)/p_2 = t < d(x-1)/x = d(x-1)/p_1$, so that state 2 would indeed obtain its $(y+1)$ 'th seat before state 1 obtained its x 'th, and d is not acceptable. A similar argument excludes $c(a) > 1$.

The other half of the theorem is shown by noting that $x/y = p_1/p_2$ and $a \leq d(a) \leq a+1$ imply $d(y)/p_2 \geq y/p_2 = x/p_1 \geq d(x-1)/p_1$, where the first inequality is strict unless $d(y) = y$, and the last is strict unless $d(x-1) = x-1 + 1$; thus, such a function d is acceptable. This completes the proof.

Corollary 1: Any acceptable divisor function d is a strictly monotone increasing function of the non-negative integer a . (It is therefore superfluous to explicitly require monotonicity in d .)

Corollary 2: All linear divisor methods with $0 \leq c \leq 1$, and all five of the methods studied by Huntington (see Section 1.6 above) are acceptable.

III.3 Linear Divisor Functions

We say d is a linear divisor function if $d(a) = a + c$ where c is a constant and $0 \leq c \leq 1$. A divisor method based on a linear divisor function will be called a linear divisor method, and identified as $LDM(c)$ since the value of the constant c specifies the method completely.

Theorem 2: No linear divisor method, except GD, which is $LDM(1)$, satisfies lower quota.

Proof: We construct a counter-example for $LDM(c)$ with $0 \leq c < 1$. Since $(x-2+c)/x \rightarrow 1$ as $x \rightarrow \infty$, we can choose an integer x with $(x-2+c)/x > c$. Then we define the pure problem with $x+1$ states, $p = (x, 1, 1, \dots, 1)$, and $h = 2x-2$. Since $d(0)/p_2 = c < (x-2+c)/x$, which in turn $= d(x-2)/p_1$, $LDM(c)$ produces the apportionment $(x-2, 1, 1, \dots, 1)$, which violates lower quota since $q_1 = x-1$.

Theorem 2': No linear divisor method, except SD, which is $LDM(0)$, satisfies upper quota.

Proof: (This is of course immediate by duality from the preceding, but we provide brief details.)

Given $LDM(c)$ with $0 < c \leq 1$, take $x > (1+c)/C$, p as above, and $h = 2$. Then $q_1 = 1$, but state 2 gets the second seat and $a_2 = 2$, violating upper quota.

State	Population	Exact Quota at $h = 23$	HM and EP Apportionments for $h = 23$
1	62	15.500	17
2	5	1.250	1
3	5	1.250	1
4	5	1.250	1
5	5	1.250	1
6	5	1.250	1
7	5	1.250	1
Total	92	23.000	23

Table III.1
Example Showing That Both HM and EP
Can Violate Upper Quota (See Theorem 3)

Theorem 3: Both HM and EP violate both upper and lower quota.

Proof: Table III.1 and the first column of Table III.2 show the necessary counter-examples.

Theorem 4: SD satisfies upper quota, and GD satisfies lower quota.

Proof: (Note that this theorem extends, to problems involving upper bounds as well as lower, an unproven part of Theorem 1 of [NMCA] and Theorem 2 of [QMA].)

Formally, SD is the set of solutions $\mathcal{L}$ defined by

(i) $\varphi_i(h_*) = r_i$ for $i = 1, \dots, s$;

(ii) Recursively for $h' = h_*+1, h_*+2, \dots, h$, setting

$\underline{a} = \varphi(h'-1)$, we find k such that $a_k/p_k = \min_U a_i/p_i$

where the set $U = \{i \mid a_i < b_i\}$, and we set

$\varphi_k(h') = a_k + 1$, $\varphi_i(h') = a_i$ for $i \neq k$,

If ever such a $\mathcal{L}$ fails to satisfy upper quota, let h' be the first (i.e., lowest) such house size in the problem $(\underline{p}, \underline{r}, \underline{b}, h)$. If the (h') th seat was assigned to state k , then every state $i \neq k$ satisfies upper quota at h' because it did so at $h'-1$.

We see $\sum a_i = h' - 1 < \sum q_i(h')$, so that for some state m , $a_m < q_m(h')$; m is in U because $q_m(h') \leq b_m$. Also, $q_m(h') > a_m \geq r_m$ shows that $q_m(h') = \min(b_m, p_m \cdot \lambda(h'))$. Thus, $a_m < q_m(h') \leq p_m \cdot \lambda(h')$.

Then, by the choice of k , $a_k/p_k \leq a_m/p_m < \lambda(h')$, and $a_k < \lambda(h') \cdot p_k$, which is in turn $\leq q_k(h')$ unless $q_k(h') = b_k$, and k is in U so $a_k < b_k$. In any case, $a_k < q_k(h')$ so that $\varphi_k(h') = a_k + 1 < q_k(h') + 1$, showing that φ does satisfy lower quota at h' and completing the proof.

(The second part of the theorem is dual to the above, and can be shown by a mechanical translation of each step; one must of course begin with the downward-recursive definition of GD.)

III.4 Primal Quota and Dual Quota Methods

We have repeatedly described the alterations made in the original (Primal) Quota method of Balinski and Young as "minor." In this section, we give that assertion a definite meaning by showing that both the Primal and Dual Quota methods, as we define them, do satisfy the axioms introduced in [1] (or rather, in the case of the consistency axiom and the Dual Quota method, the dual axiom). Because the two proofs are dual to each other, it would be wasteful to prove both; and we prefer to prove the above assertion in some detail for the dual quota method, since it differs more from the proof of [2]. In spite of the distinctions, the proof given below follows the essence of the proofs of Theorems 3 and 4 of [2].

Theorem 5: The Dual Quota method DG is monotone, satisfies quota, and is dual-consistent.

Proof: (Refer to the definition of DG in II.3 above.)

By the construction, it is obvious that DQ is monotone (since the algorithm withdraws one seat from some state for each decrease of the house size), and dual consistent (since the selection of losing state, within the dual-eligible set E' , depends on a criterion -- the maximization of $(a_i - 1)/p_i$ -- governed entirely by populations and previous portions).

It is also obvious that any solution $\underline{Y}$ in DQ satisfies lower quota, since $\underline{Y}(h^*) = \underline{b} = \underline{q}(h^*)$, and it is only the dual eligible states -- viz, those whose portions could be decreased without violating lower quota -- from which the algorithm may withdraw a seat. $E'(h-1)$ is surely non-empty; since $\sum a_i = h$ and $\sum q_i(h-1) = h-1$, there must be some state i with $a_i > q_i(h-1)$. (Note this argument is simpler than that of Theorem 4 of [2] -- we have thrown some of the effort back into the definition of "exact quota.")

Thus we need "only" show that $\underline{Y}$ satisfies upper quota. Suppose the contrary; let $h_0 < h^*$ be the first (i.e., greatest) house size at which $\underline{Y}(\underline{p}, \underline{r}, \underline{b}, h)$ fails to satisfy upper quota, and let j be a state with $\underline{Y}_j(h_0) > u_j(h_0)$. Since j satisfied upper quota at h_0+1 , j is not the state which lost the (h_0+1) 'th seat; and, since $\underline{Y}_j(h_0) \leq \underline{Y}_j(h^*) = b_j$, we have $q_j(h_0) \leq u_j(h_0) < b_j$.

Now define the set M of states under-represented at h_0 ;
 $M = \left\{ m \mid \underline{Y}_m(h_0) < q_m(h_0) \right\}$. Obviously j is not in M ; but M cannot be empty since $\sum \underline{Y}_i(h_0) = h_0 = \sum q_i(h_0)$ and $\underline{Y}_j(h_0) > q_j(h_0)$.

Since $\underline{Y}$ satisfies lower quota, for each m in M we have $\underline{p}_m(h_0) = \underline{Y}_m(h_0) < q_m(h_0)$, and the non-integral $q_m(h_0)$ must equal $\underline{p}_m(h_0)$. Every state of M has surely lost at least one seat at some house size h in the range $h_0 \leq h < h^*$. Denote by t the last state of M to lose a seat for any house size in that range, and by h_t the house size after that loss.

If h_t were $= h_0$, and t had lost the (h_0+1) 'th seat, then
 $\gamma_t(h_0) = \gamma_t(h_0+1) - 1 < q_t(h_0)$ because t is in M , while
 $\gamma_j(h_0+1) - 1 = \gamma_j(h_0) - 1 \geq q_j(h_0)$ by definition of j . Thus
 $(\gamma_t(h_0+1) - 1)/q_t(h_0) < 1 \leq (\gamma_j(h_0+1) - 1)/q_j(h_0)$, so that
 $(\gamma_t(h_0+1) - 1)/(p_t \cdot \lambda(h_0)) < (\gamma_j(h_0+1) - 1)/q_j(h_0) \leq (\gamma_j(h_0+1) - 1)/(p_j \cdot \lambda(h_0))$,
 since t is in M and $q_j(h_0) \neq b_j$. But j was surely dual-eligible
 at h_0 ; thus t was not selected to lose the (h_0+1) 'th seat, and $h_t \neq h_0$.

Now define K as the set of states to lose a seat at house sizes
 h in the range $h_t > h \geq h_0$; formally $K = \{k \mid \gamma_k(h_0) < \gamma_k(h_t)\}$. K is not
 empty, but (by definition of t) $K \cap M$ is empty.

In four steps, we shall show that

$$(\gamma_k(h_t+1) - 1)/p_k > (\gamma_t(h_t+1) - 1)/p_t \text{ for every } k \text{ in } K:$$

First, t is not in K , so $\gamma_k(h_t+1) - 1 = \gamma_k(h_t) - 1$, which in turn
 is $\geq \gamma_k(h_0)$ by definition of K , which again is $\geq q_k(h_0)$, since k is not in
 M . Thus, $(\gamma_k(h_t+1) - 1)/p_k \geq q_k(h_0)/p_k$.

Second, $q_k(h_0) \leq \gamma_k(h_0) < \gamma_k(h_t) \leq b_k$, so $q_k(h_0) =$
 $\max(r_k, p_k \cdot \lambda(h_0)) \geq p_k \cdot \lambda(h_0)$, and $q_k(h_0)/p_k \geq \lambda(h_0)$.

Third, since t is in M , $\gamma_t(h_0) < q_t(h_0) = p_t \cdot \lambda(h_0)$, so that
 $\lambda(h_0) > \gamma_t(h_0)/p_t$.

Fourth, since t is not in K , $\gamma_t(h_0) = \gamma_t(h_t)$, which (by definition
 of t) is equal to $\gamma_t(h_t+1) - 1$.

Assembling the results of those four steps, we see that $(\psi_k(h_{t+1})-1)/p_k > (\psi_t(h_{t+1})-1)/p_t$, so that any state k in the dual-eligible set $E'(h_t)$ would have precluded t from attaining the maximum at house size h_t . But t did attain that maximum, and lost the (h_t+1) 'th seat, hence $K \cap E'(h_t)$ must be empty, and $\psi_k(h_t) = \psi_k(h_{t+1}) \leq q_k(h_t)$.

Finally, we deduce a contradiction from:

$$h_t - h_0 = \sum_K (\psi_k(h_t) - \psi_k(h_0)) \text{ by definition of } K;$$

$$\sum_K \psi_k(h_0) \geq \sum_K q_k(h_0) \text{ since } K \cap M \text{ is empty;}$$

$$\sum_K \psi_k(h_t) \leq \sum_K q_k(h_t) \text{ from preceding paragraph;}$$

$$\sum_L q_i(h_0) < \sum_L q_i(h_t), \text{ where } L \text{ is the set of states not}$$

in K , because t is not in K and $q_t(h_0) = p_t \cdot A(h_0) < q_t(h_t)$.

Assembling those inequalities we obtain:

$$\begin{aligned} h_t - h_0 &= \sum_K (\psi_k(h_t) - \psi_k(h_0)) \\ &\leq \sum_K (q_k(h_t) - q_k(h_0)) \\ &< \sum_L (q_i(h_t) - q_i(h_0)) \\ &= h_t - h_0, \end{aligned}$$

a contradiction which shown that DQ satisfies upper quota and completes the proof.

Theorem 6: The Primal Quota method is the unique monotone and consistent method which satisfied quota.

Proof: This theorem rests entirely on the similar theorem of [2].

The only differences between our Primal Quota method and the Quota method of Balinski and Young, are (i) we have permitted a more general problem, which includes upper bounds on the portions, and (ii) we have used a slightly different -- and more restrictive -- definition of quota. Since ref. [2] has proved uniqueness for their definition, it follows that no method other than our Primal Quota can have the three desired properties for our problem.

III.5 Influence of the Upper Bounds and Lower Bounds

In this section, we collect such theorems as we have which concern the importance of the upper bounds and the lower bounds of the apportionments which result. Where we have not a theorem, we present a counter-example; where we have neither, we present a conjecture. The hope is, that results of this class can free us from the necessity to compute step-by-step solutions for an entire problem-set (as specified in some of the algorithms of Sections II.5 and II.6).

Theorem 7: Changing the upper bounds for a problem will not change the Primal Quota apportionment unless it changes the exact quotas.

Proof: Formally, the hypothesis states that $q(\underline{p}, \underline{r}, \underline{b}, h) = q(\underline{p}, \underline{r}, \underline{b}', h)$, and the conclusion is that there exists a Primal Quota solution $\mathcal{P}$ such that $\mathcal{P}'(\underline{p}, \underline{r}, \underline{b}, h) = \underline{a}$, where $\underline{a} = \mathcal{P}(\underline{p}, \underline{r}, \underline{b}, h)$ is a Primal Quota apportionment for the first problem.

(Note that this theorem does not allow variation of the lower bounds $\underline{r}$.)

It is easy to see that the same value of $f(h)$ may be used to compute the two exact-quota vectors, and thus that any upper bounds, which are constraining for either problem at house size h , must be equal. Then for smaller house sizes h' , it is true a fortiori that any constraining

upper bounds are equal, so the exact-quota vectors are equal at h' also. Reference to the definition of the Primal Quota algorithm (Section II.2) shows that the resulting apportionment depends only on the sequence of exact quotas and constraining upper bounds, proving the theorem.

Dual to the above is Theorem 7': Changing the lower bounds for the problem will not change the Dual Quota apportionment unless it changes the exact quotas.

Notation: The divisor method with divisor-function d will be denoted by $M(d)$.

Definitions: For any problem $(p, \underline{r}, \underline{b}, h)$, any divisor-function d , any state i , and any number α satisfying $r_i < \alpha \leq b_i$, we call α a potential i -seat, and define the d -criterion for such a potential i -seat to be the quantity $d(\alpha - 1)/p_i$.

Theorem 8: Suppose $\mathcal{V}$ is a solution belonging to a divisor method $M(d)$, and we are given two problems, $(p, \underline{r}, \underline{b}, h)$ and $(p, \underline{r}', \underline{b}', h)$, with the apportionment $\mathcal{V}(p, \underline{r}, \underline{b}, h) = \underline{a}$. If the two problems are so related that we have, for each i , either $r_i \leq r'_i \leq a_i$ or $r'_i < r_i < a_i$, and also, for each i , either $a_i \geq b'_i \geq b_i$ or $a_i > b_i > b'_i$, then there is a solution $\mathcal{V}'$ belonging to $M(d)$ with $\mathcal{V}'(p, \underline{r}', \underline{b}', h) = \underline{a}$.

Proof: $M(d)$ solves the first problem by first assigning r_i seats to each state i , and then selecting, from the total $h^* - h_*$ potential

seats, the $h - h_*$ whose d -criterion values are smallest; $a_i - r_i$ potential i -seats are thus selected. If $r_i \leq r'_i \leq a_i$, then $r'_i - r_i$ of the potential i -seats selected for the first problem are assigned without competition when the second is solved; if $r'_i < r_i < a_i$, then $r_i - r'_i$ more potential i -seats must compete when the second problem is solved, but the monotonicity of d shows that they would be selected in any case before the (a_i) th potential i -seat, which was actually selected.

Similar arguments for the upper bounds complete the proof.

III.6 Periodicity of Apportionment Solutions

When these apportionment methods are applied to such problems as manpower allocation, the possibility arises that the number to be apportioned -- the "house size" -- might exceed, perhaps by a substantial factor, the sum of the numbers -- the "populations" -- on which the proportionality is based. Such a circumstance, in the legislative-apportionment or proportional-representation problems, would imply that the number of seats in the legislature exceeded the total electorate, and so has been inevitably disregarded heretofore.

Under the circumstances when the house size greatly exceeds the sum of the populations, we find that most methods which have ever been proposed are either periodic, or ultimately become periodic. (It is even true of the Hamilton method.)

Theorem 9: If P denotes the sum $\sum p_i$, and $\underline{p}$ denotes an s -vector all of whose components are equal to P , then $\underline{\mathcal{Y}}(\underline{p}, \underline{0}, \underline{p}, P) = \underline{p}$ if $\underline{\mathcal{Y}}$ is a solution belonging to any divisor method, any quota method, or the Hamilton method.

Proof: (Recall that the definition of "(apportionment) problem" required that the p_i be positive integers.)

Because the exact quotas are just the populations, a method which satisfies quota must produce the integral apportionment $\underline{p}$.

Because exact proportionality is possible for this problem, any acceptable divisor method will achieve it. (See Section III.2 above.)

Because $q(P) = p$, $h - \sum \ell_i = 0$, and the Hamilton method is trivial for this problem.

Theorem 10: Linear divisor methods, when applied to a sequence of pure problems with house size increasing without limit, produce a sequence of gaining states which is periodic with period $P = \sum p_i$.

Proof: By the preceding theorem, state i must get p_i of the first P seats, for each i . Then the choice of gaining seat at house size $P+1$ will be governed by the criteria $d(p_i)/p_i$, for $i = 1, 2, \dots, s$. But the linear divisor methods have divisors such that $d(a) = a+c$, so that $d(p_i)/p_i = (p_i+c)/p_i = 1 + d(0)/p_i$; thus the order-relations for the criteria at $h = P+1$ are identical with those at $h = 1$. If ties are broken for h between P and $2P-1$ as they were for h between 0 and $P-1$, we find that each criterion at house size $P+h$ is greater by 1 than the corresponding criterion at house size h , and the periodicity is established.

Theorem 11: All the quota methods of Appendix II which depend on a linear divisor-function are periodic in h with period P , when applied to a sequence of pure problems with common population-vector p .

Proof: Examination of the algorithms in Sections II.2, II.3, II.5, and II.6 shows that the selection of gaining or losing state is always based

House Size	SD HM EP	SDQ=DQ HMQ EPQ	MF	MFQ	GD	GDQ=PQ	Hamilton
0		0000		0000		0000	0000
1		1000		0001		0001	0001
2	(1100)	1001		0011		0011	0011
3	(1110)	1011		0012		0012	0012
4	(1111)	1012		0022		0022	0022
5	1112	1022		0023		0023	0023
6	(1122)	1023		1023	(0024)	0033	1023
7	1123			1123		0034	1123
8	(1133)	1124	(1133)	1124	(0035)	1034	1034
9	1134			1134	0045	1035	1134
10	1135			1135	(1046)	1045	1135
11	1145			1145		1046	1145
12	1146			1146		1146	1146
13	1156			1156		1156	1156
14	1157			1157		1157	1157

Table III. 2

The Apportionments Generated by

Eleven Distinct Methods:

SD (Smallest Divisors) and its Quota Analog SDQ (= Dual Quota)
HM (Harmonic Mean) and its Quota Analog HMQ
EP (Equal Proportions) and its Quota Analog EPQ
MF (Major Fraction) and its Quota Analog MFQ
GD (Greatest Divisors) and its Quota Analog GDQ (= Primal Quota)
Hamilton (= Vinton Method of 1850)

(All these methods applied to the Pure Problems with
 $p = (1, 1, 5, 7)$ and $0 \leq h \leq 14$.)

- Notes: (i) All ties were resolved by giving the seat to the
"first among equals";
(ii) Apportionments in parentheses violate quota;
(iii) Alabama Paradox exhibited at $h = 8$.

on the exact quotas, and the criteria $d(a)/p$. When the house size increases by P , each exact quota increases by p_i and each previous portion increases by p_i , and each criterion $d(a)/p$ increases by 1; thus, the eligible sets are identical, and the state to be selected within the eligible set is identical; an empty eligible set (causing a backtrack) will also be repeated at intervals of P in the house size.

For the Dual Quota and other descending quota methods, the allocations must begin with some multiple of P seats; then the same sequence of losing seats will be found within each consecutive block of P seats. The sequence does not depend on what multiple of P is used as the initial house size; Theorem 9 shows that the initial allocation for a house of size $n.P$ is simply $n.p$.

Theorem 12: When applied to a sequence of pure problems with common population p and increasing house sizes, the method EP, HM, EPQ, and HMQ ultimately attain one of the same periods as the methods MF, MF, MFQ, and MFQ, respectively.

Proof: Define $f(a)$ to be the difference $a + 1/2 - \sqrt{a(a+1)}$ between the divisor $a + 1/2$ used in MF and the divisor $\sqrt{a(a+1)}$ used in EP.

The Major Fractions criterion, when applied to all the P potential seats in a house of size P , takes on at most P distinct values; there is a least positive difference e between unequal values of that criterion.

(That least difference is the same in each succeeding block of P seats, because each of the criteria is increased by 1.)

Since $f(a) \rightarrow 0$ as $a \rightarrow \infty$, we can find x such that $f(a) < e$ whenever $a > x$. For h large enough, all portions will exceed x , and thereafter EP will duplicate a periodic sequence produced by MF .

The same argument shows that HM ultimately duplicates a period of MF and (noting that the sequence of supereligible sets is also periodic) that each of EPQ and HMQ must ultimately duplicate a periodic behavior of MFQ .

Theorem 13: The periodic sequence of P gaining states which results when $LDM(c)$ is applied to a sequence of pure problems with population p may be chosen to be the reverse of the sequence of gaining states resulting from application of $LDM(1-c)$ to the same pure problems.

Proof: It is only necessary to note that, if $LDM(c)$ has given m seats to state 1 and n seats to state 2, then the next gaining state is chosen by comparing criteria like $(m+c)/p_1$ and $(n+c)/p_2$; while a descending algorithm for $LDM(1-c)$, which has withdrawn m seats from state 1 and n seats from state 2 after beginning with the apportionment p at house size P , will select the next losing state by comparing $(p_1-m-1+(1-c))/p_1$ with $(p_2-n-1+(1-c))/p_2$; but the former is equal to $1 - (m+c)/p_1$ and the latter to $1 - (n+c)/p_2$, so the ordering (when definite) will be reversed. Since the ascending algorithm awards the next seat to the state whose

criterion is least, while the descending algorithm withdraws the seat from the state whose criterion is greatest, the downward sequence of losses in $LDM(1-c)$ can be chosen to duplicate the upward sequence of gains in $LDM(c)$.

Corollary: The reverse of a valid P-sequence for MF, which is $LDM(1/2)$, may be chosen as a valid P-sequence for the same sequence of pure problems. (We cannot say that a P-sequence for MF must be palindromic; that would be obviously impossible if more than one of the p_i were odd.)

Theorem 14: The sequence of gaining states in the ascending quota method based on $LDM(c)$ can be chosen to be identical with the sequence of losing states in the descending quota analog of $LDM(1-c)$, for pure problems.

Proof: Can be easily verified by checking the definitions of supereligible sets, criteria, and quotas; the relation among those entities parallels the relations of the preceding theorem.

Corollary: For the pure problem, the ascending quota method based on MF, which is $LDM(1/2)$, is identical with the descending quota method based on MF. This quota method can thus be called the MFQ method.

Theorem 15: If $LDM(c)$, or its quota analog, is applied to a sequence of pure problems with one or more of the $p_i = 1$, those "singleton" states will tend to get their single seat when the large states have obtained the fraction c of their share of the first P seats.

Proof: The criterion, for the potential seat of a singleton state, will be $d(0)/1 = c$; thus the singleton states will be nearly tied with a large state (j , say) when the latter has received a_j seats, where $d(a_j)/p_j = c$ nearly; but if p_j is large, that implies a_j is nearly $c \cdot p_j$.

III.7 Backtracking

As stated in Section II.5, backtracking will be required (for some problem-set) in every ascending quota method except PQ, and in every descending quota method except DQ.

Table III.2 shows examples of several such methods applied to the pure problems with $p = (1, 1, 5, 7)$. Note that the apportionment $(1, 1, 1, 2)$ satisfies quota but cannot be continued in the ascending direction; the exact quota for $h = 6$ is $(3/7, 3/7, 15/7, 3)$, so both states 3 and 4 would have to gain the 6'th seat in order to satisfy quota. Therefore, we know that the apportionment $(1, 1, 1, 2)$ cannot result from any monotone solution satisfying quota.

The apportionment $(0, 0, 4, 5)$, which is complementary to the above, also satisfies quota; however, this latter cannot be continued in the descending direction because the exact quota for $h = 8$ is $(4/7, 4/7, 20/7, 4)$, so both states 3 and 4 would have to lose a seat in order to satisfy quota. No monotone quota solution can produce $(0, 0, 4, 5)$ either.

These examples show why it is necessary to find apportionments for an entire problem-set (including all house sizes from h_* to h^*), and not only for one or other of the intervals from h_* to h , or from h to h^* ; although $(0, 0, 4, 5)$ is inaccessible to any ascending quota algorithm, it may be found by certain descending quota algorithms -- in fact, it is found by descending GDQ -- though it must later be

eliminated by a backtrack. (Our conjecture, Section II.6, would imply that the descending GDQ method is identical with ascending GDQ, which is of course simply the Primal Quota method PQ.)

Table III.2 gives, for the pure problem set defined by $\underline{p} = (1, 1, 5, 7)$, the apportionments generated by all five of Huntington's methods and by their quota analogs. (Because of Theorems 10 and 11, we need only specify apportionments for house sizes between 0 and $\sum p_i = 14$ inclusive.) We do not distinguish between the ascending and descending quota methods, because we have no example where the resulting apportionments are different. All the results of this table assume that ties are broken by selecting the "first among equals" as the gaining state and the "last among equals" as the losing state.

Although the apportionments $(0, 0, 4, 5)$ and $(1, 1, 1, 2)$ both satisfy quota, and can each be reached by either an a.q. method or a d.q. method but not both, neither of them can occur in a monotone quota solution. We may ask whether there are apportionments satisfying quota which cannot be reached by any quota algorithm. The answer is "yes"; there are, in fact, sets of apportionments ("enclaves") which cannot be reached either from h_* or from h^* . A simple example is the apportionment $(0, 0, 2, 2)$ in the pure problem set with $\underline{p} = (3, 3, 4, 4)$; when $h = 3$, the exact quotas are $(9/14, 9/14, 6/7, 6/7)$, so that the 4'th seat would have to be withdrawn from both states 3 and 4, and when $h = 5$, the exact quotas are $(15/14, 15/14, 10/7, 10/7)$, so that the 5'th seat would have to be given to both states 1 and 2.

It is interesting to speculate on a possible maximum size of enclave -- presumably depending on the number s of states and the population-vector $\underline{p}$ for a pure problem, and also on the bounds $\underline{r}$ and $\underline{b}$ for a constrained problem. Theorems to this effect could avoid the need to always begin an a.q. algorithm with h_* and a d.q. algorithm with h^* .

III.9 Non-Integral Populations

Many of the above theorems depend on the assumption that the population vectors are integral, and would not be true if the p_i were relatively irrational. Actual census populations, of course, are always integers -- even the original "three-fifths of a man" in the U.S. Constitution resulting in rational population vectors -- but relative priorities, which need not be rational fractions, may be used in some manpower-allocation problems as a basis for apportionment of available men or spaces.

Those manpower problems, and other conceivable applications of these apportionment methods, will almost always possess natural constraints $\underline{r}$ and $\underline{b}$, and the properties specified by those theorems will thus be of lesser importance in those cases. Theorem 14 (which applies that the ascending MFQ and the descending MFQ methods are identical) does depend on the integrality of $\underline{p}$, but the loss of that theorem seems to be the only practical consequence of using non-integral values in place of populations $\underline{p}$.

(From the mathematical standpoint, a pure problem with irrational $\underline{p}$ would be almost-periodic in a well-defined sense, but that fact seems to be irrelevant to any contemplated applications.)

III.10 Recommendations

Because the periodicity seems an attractive and elegant notion, I favor the use of a linear-divisor quota method for manpower allocation; among these, MFQ would be my first choice, because it is well-defined, because it avoids the inconvenient ties when a distributable community is slightly over-manned or slightly under-manned, and because the middle seems like a good compromise. However, as stated in Section II above, DQ ought to be used for the qualitative Phase IV.

Primarily on the grounds of mathematical elegance, I would favor the same method for legislative apportionment. On the other hand, existing legislation specifies EP -- so that the Congress might be more receptive to EPQ than to MFQ. This would be especially true if it could be proven that EPQ always agrees with EP, whenever the latter satisfies quota. (That seems very likely, but has been so difficult to prove that I now suspect the existence of rare counter-examples.) The difference between EPQ and MFQ is likely to be miniscule; every state must have one representative anyway, and even at $a = 1$ (i.e., when considering whether or not a state gets its second seat), we find that EP gives $d(1) \approx 1.414$, MF gives $d(1) = 1.500$ -- a difference of only 6%. Such small differences as do exist would tend to make EP favor the small states -- i.e., a doubtful state is more likely to get its second seat with EP than with MF.

REFERENCES

- (1) (NMCA) Balinski, M. L., and Young, H. P., "A New Method of Congressional Apportionment," P.N.A.S., v. 71 (1974), pp. 4602-4606.
- (2) (QMA) Balinski, M. L., and Young, H. P., "The Quota Method of Apportionment," A.M. Monthly, v. 82 (1975), pp. 701-730.
- (3) (ASQM) Balinski, M. L., and Young, H. P., "Apportionment Schemes and the Quota Method," A.M. Monthly, v. 84 (1977), pp. 450-455.
- (3a) (HMA) Balinski, M. L., and Young, H. P., "On Huntington Methods of Apportionment", SIAM Jour. Appl. Math., v. 33 (1977), pp. 607-618.
- (4) (MTAR) Huntington, E. V., "The Mathematical Theory of the Apportionment of Representatives," Proc. N.A.S., v. 7 (1921) pp. 123-127.
- (5) (ARC) Huntington, E. V., "The Apportionment of Representatives to Congress," Trans. A.M.S., v. 30 (1928), pp. 85-110.
- (6) (QMNU) Mayberry, J. P., "Quota Methods for Congressional Apportionment are Still Non-Unique," submitted to P.N.A.S.