

## FORD'S PROCEDURE FOR COMBINING MULTIPLE SETS OF PARTIALLY ORDERED DATA<sup>1</sup>

FRANK M. ANDREWS AND DONALD C. PELZ

Institute for Social Research  
University of Michigan

MORE than a decade ago, Lester R. Ford, Jr. published a mathematical procedure for combining multiple sets of partially ordered data (Ford, 1957). Although this procedure has received little attention from social scientists, it solves a methodological problem which is not uncommon in some fields of educational and psychological research. Furthermore, it permits certain flexibilities in the way data are collected which, if used, would probably increase measurement validity in some kinds of variables.

Ford's procedure is likely to be useful whenever one wants to achieve a single rank order by combining data obtained from multiple "judges" who have placed "objects" into ordered categories. Of course, the "judges" need not be people, but any procedure which orders objects. The procedure has special advantages when data are untidy in any of several ways: (a) some objects are unclassified by some judges (e.g., when some judges lack knowledge about some of the objects being ranked), (b) there is less than perfect agreement among the judges, (c) some objects are tied—i.e., placed in the same ranked categories by certain judges, (d) the judges differ in the number of ranked categories they use (e.g., when judges differ in their ability to discriminate along the dimension being considered), (e) the judges differ in their perception of the overall distribution along the dimension

---

<sup>1</sup> We are grateful to Allen Krebs, who assisted in getting Ford's procedure operational on computers at the University of Michigan, and Bernard Galler and Robert Hsieh, who did the computer programming. Some of the material presented here also appears in Appendix A of Pelz and Andrews (1966).

being considered (e.g., one judge might feel that highs were rather common, while another might feel they were rare). To our knowledge, Ford's procedure is the *only* one which appropriately handles situations involving substantial "missing data"—the first of the problems listed above.

Because of the difficulties of handling such situations, investigators usually try to avoid them. Often judges are asked to categorize all objects, to use exactly  $X$  categories, and to distribute their assignments in some pre-determined manner. It is clear, however, that in some research settings these requirements would be overly restrictive. Rather than impose an unrealistic task on the judges (with consequent losses in validity) or forego an interesting research opportunity, use of Ford's procedure may permit the researcher to utilize the data he *can* collect.

### *Research Examples*

A situation in which Ford's procedure proved helpful occurred when the authors wished to measure the quality of scientists' performance (Pelz and Andrews, 1966). A group of the scientists' peers and supervisors were asked to make qualitative discriminations (on a carefully defined dimension of technical contribution) among the members of a lab. They were instructed to judge only those members whose work they knew. And they were permitted to use as many qualitative categories as they chose, and to place as many people in each category as they thought appropriate.

In this case judges differed in the subsets of objects they evaluated, in the numbers of categories they used, in their distributions across these categories, and (sometimes) in the relative positions assigned to the scientists. Ford's procedure handled these data well and produced a single rank order among the scientists which could be used as a variable in subsequent analyses.

Other situations in which the procedure might be applied are easy to imagine. It might be used in assessing hierarchies among needs or motives, attractiveness of products to consumers, competitive skill among groups such as simulated business teams, and many more.

### *Application to Data*

The advantages of Ford's procedure over certain other ways of combining partially ordered data are shown in Table 1, which

TABLE 1  
Artificial Evaluations of Ten Objects by Six Judges

True order	Object	Evaluations by judges						Method I Mean rank	Method II Mean per-centile	Method III Win %	Method IV Order	Method by Order by Ford Program
		I	II	III	IV	V	VI					
1	A	1	1	1	1	1	1.0	89	100	1	1	
2.5	B	2	2	2	2	2	2.0	72	78	3	3	
2.5	C	1	1	3	2	2	1.7	78	86	2	2	
4	D	3	2	4	4	4	3.0	50	50	6	6	
5	E	4	4	5	5	5	4.5	28	22	9	8	
6	F	5	5	6	6	2	3.2	47	45	7	7	
7	G					1	2.0	63	61	4	4	
8	H	1		3	3	3	2.3	61	60	5	5	
9	I	2	3	3	5	4	3.5	30	20	8	9	
10	J	3	4	4	4	5	4.0	16	6	10	10	

Note.—Judges were free to rank only objects with which they were familiar, and were allowed flexibility in the number of categories used and the number of objects assigned to each. Methods I to III for combining evaluations give misleading results. Ford's procedure, which is the basis of Method IV, gives results close to the true order.

contains a realistic, but fictitious, situation. Ten objects comprise the set being considered (A, B, . . . J) and the table shows the categories assigned by each of six judges. (In terms of the research example cited previously, the reader may wish to think of the objects in Table 1 as being scientists, and the criterion of ranking as being their technical contribution.)

One may note that the judges showed high, but not perfect, agreement in their rankings. Judges I, V, and VI considered objects relatively low on the dimension being considered. Judges II and IV ranked only objects at the upper end. Judge III categorized some of each. The number of categories used by the judges ranged from three (Judge I) to six (Judge IV). Although Judge III used only four categories, he evaluated more objects than any other judge.

Given the simplicity of Table 1, one can identify the "true" order of the objects. This is shown at the left. Note that the order between objects B and C is undefined (Judge II placed C above B, but Judge IV placed B above C, and no other judge compared them). It is clear, however, that these two objects fall below A and above D. In addition, there was minor disagreement among the judges concerning the relative positions of F versus G (two judges out of three placed F ahead) and I versus J (three out of four placed I ahead).

### *Three Unsatisfactory Systems*

In some kinds of data the *average rank* assigned by several judges provides a satisfactory way of combining judgments. The fact that in the present data judges evaluated different sets of objects, and used different numbers of categories, however, makes this system inappropriate. Note in Table 1 that this first method gives results very different from the true rank order.

Allowance could be made for the judges' using different numbers of categories by converting to a "percentile equivalent" for the rank (if one were willing to assume equal distances between categories). A combined score could then be based upon an average of the "percentile equivalents." This second method was applied to the data of Table 1. Note that it also fails to reproduce the true rank order. The reason is that this method, like the first, neglects the fact that judges ranked different sets of objects.

A somewhat more sophisticated system is to set up a "win-loss

matrix," as shown in Table 2. For each possible pair of objects, this matrix indicates the number of times the first was ranked above the second, and vice versa. Such a matrix can be used to determine each object's "win percentage"—the number of times it scored above others relative to the total number of times it won or lost. The "win percentages" derived by this third method are also shown in Table 1. This third method gives an order somewhat different from either of the previous two, but still fails to reproduce the true order.

### *Ford's Procedure*

When Ford's procedure was applied to these same artificial data, the results (shown at the right of Table 1) were virtually identical with the true order.<sup>2</sup> The procedure works as follows.

Ford proposed that one assign to each object a number or "weight" ( $w$ ) which could be interpreted as odds, in the sense that the probability of object  $i$  being preferred to object  $j$  in a future comparison would be  $w_i / (w_i + w_j)$ . With these probabilities, one could compute the a priori probability of obtaining exactly the win-loss matrix actually obtained. Furthermore one could rank the objects in order according to these weights.

The problem, then, is to determine that set of weights which maximizes the likelihood of obtaining the given matrix.

Ford showed that, beginning with an arbitrary set of weights,<sup>3</sup> one can arrive at the desired set of weights by an iterative technique—if the data meet certain assumptions. His procedure is to solve the following equation for each object in the set until the resulting weight for each object stabilizes.<sup>4</sup>

<sup>2</sup> Our computer program embodying Ford's procedure produced a completely ordered set (that is, no ties), whereas in actuality the set contained a pair of tied objects. The discrepancy is probably attributable to the addition of a small constant to the cells of the win-loss matrix to ensure that all sets of data would meet Ford's assumptions (described below).

<sup>3</sup> Our computer program uses as the initial set of  $w_i$ 's the percentage of wins in the win-loss matrix, as illustrated in Table 2. We believe this will usually minimize the number of iterations required.

<sup>4</sup> Although the criterion of what constitutes "stability" is arbitrary, one criterion used by us has been that no weight should change by more than 0.5% from one iteration to the next: that is, for all objects

$$\frac{w_i^{n+1} - w_i^n}{w_i^n} \leq .005.$$

As described subsequently in the text, stability in *rank order* of the weights (even though their actual values may change from iteration to iteration) is an alternative criterion.

TABLE 2  
Win-loss Matrix Based on Evaluations Shown in Table 1

		Winning objects										Times lost	Win %	Rank
		A	B	C	D	E	F	G	H	I	J	(=b)	[=a/(a + b)]	
Losing objects	A	1										0	100%	1
	B	1	1									2	78%	3
	C	2	2	3								2	86%	2
	D	1	2	2	2							7	50%	6
	E	2	2	3	2	2						7	22%	8
	F	2	2	3	2	2						12	45%	7
	G	1		1	1			1				5	61%	4
	H						2	2				4	60%	5
	I	1		1	1		3	2	3		1	12	20%	9
	J	1		1	1		3	3	3	3		15	6%	10
Times won (=a)		10	7	12	7	2	10	8	6	3	1			

$$w_i^{n+1} = \frac{\sum_i a_{ij}}{\sum_i \frac{a_{ij} + a_{ji}}{w_i^n + w_j^n}}$$

where  $a_{ij}$  = number of times object  $i$  was preferred to object  $j$ ;  $a_{ji}$  = number of times object  $j$  was preferred to object  $i$ ; and,  $w_i^n$  = number (weight) assigned to object  $i$  on the  $n$ th iteration.

### *Assumptions*

For Ford's technique to yield a solution, the data must meet the following assumption: "In every possible partition of the objects into two non-empty subsets, some object in the second set has been preferred at least once to some object in the first set." (Ford, 1957, p. 29). Thus if the objects were all baseball teams, this assumption would be violated if it were possible to divide the teams into a major and minor league, where the major teams had always defeated the minor teams.

In practice, this basic assumption can be violated in four ways. (a) One object may be universally preferred by the judges. It heads all lists on which it appears. (In Table 1, A is such an object.) (b) An object might be at the bottom of all lists on which it appeared. (c) Some objects might be judged neither universally high nor universally low, but taken as a group, were simply not judged in relation to the other objects. (d) Some objects might fall in a subset such that the comparisons with another subset were all in one direction. (After removing object A from Table 1—see next paragraph—the subsets B-C, B-C-D, B-C-D-E, and B-C-D-E-F-G constitute violations of this type.)

In programming Ford's procedure for the computer, it proved relatively easy to identify violations of the first and second types. Our Ford Program identifies objects which were "universal highs" or "universal lows," and removes them before computing the weights. (They subsequently get ranks above or below the set of objects for which weights are computed, as appropriate.)

Violations of the third and fourth types, however, are often extremely difficult to identify. Our solution consisted of adding an extremely small constant (0.00001) to every cell of the win-loss matrix derived from the judges' evaluations. This constant

insured that the matrix would not contain violations of the third and fourth types.<sup>5</sup>

### *A Computer Program*

With the information provided above, a competent computer programmer could write a program to carry out the various steps described. We have substantial experience with one such program. This program, written in FORTRAN for an IBM 360/40, will handle up to 130 judges who may use up to 130 ranked categories in evaluating up to 130 objects.<sup>6</sup> In a wide variety of test cases, our Ford Program has yielded a combined rank order which closely approximated a known "true" order, as in Table 1.

However, experience with this program has shown that for some sets of data an exceedingly large number of iterations would be required before all weights met the stability criterion described above (see footnote 4). (In such cases the iterative procedure is usually stopped after a designated number of iterations.) It turned out, however, that the *rank order* of the weights usually did not change after 20 to 30 iterations. Since it is the rank order of the weights which we desired, the slowness to reach stability is usually not a problem.<sup>7</sup>

The data shown in Table 1 provided one example where the *rank order* of the weights stabilized after 25 iterations even though the actual values of some weights had not reached the stability criterion. Figure 1 shows how the weights of each object changed over 150 iterations. Note that no lines crossed after the 25th iteration.

### *Conclusion and Discussion*

Ford's procedure, virtually unknown to social scientists, provides a practical and appropriate way for combining several sets

---

<sup>5</sup> This solution was accepted only after experimentation with several other possibilities which included: adding 0.00001 only to the zero cells of the win-loss matrix, adding 1.0 to all cells, and adding 1.0 only to zero cells. In several trial sets the smaller constant produced more rapid stabilization of the weights, and whether the constant was added to all cells or only to zero cells proved to make little difference.

<sup>6</sup> A listing of the program (operational on IBM's 360/40) and instructions for data input are available from the authors.

<sup>7</sup> Where all weights did reach stability it might be possible to use the relative distances among them to scale the objects with more precision than a simple rank order.



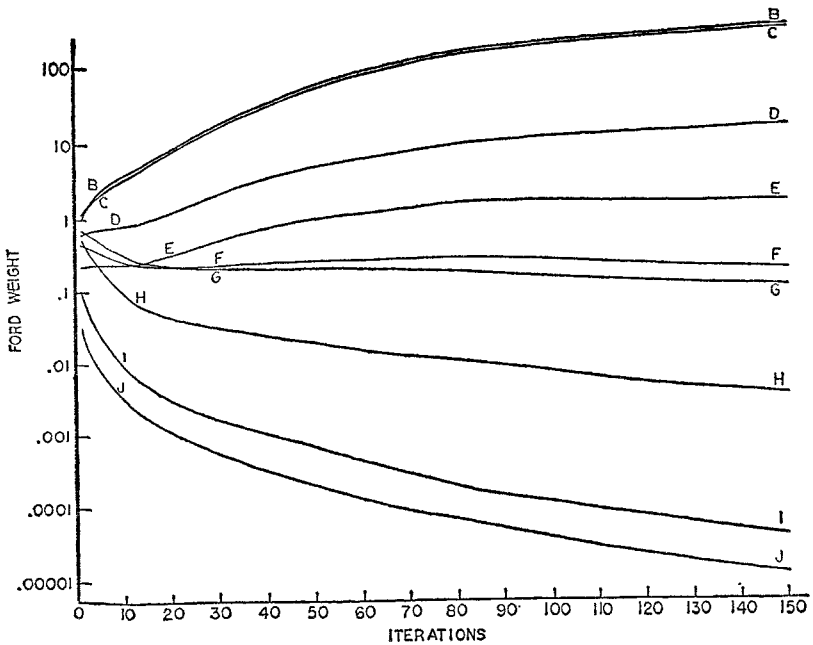


Figure 1. Weights assigned objects B-J on each of 150 iterations.

Note.—Based on data shown in Table 1. Object A was removed as a “universal high” prior to the first iteration..

of partially ordered data. Although such data are usually avoided because of the difficulties of analysis by traditional means, in many instances data of this kind could be exploited, in conjunction with Ford's method, to derive better measures of many variables of interest to social scientists. This article has described Ford's procedure, given a specific example of its use, and shown how it produced more valid results with these data than three more traditional techniques.

While the method, as embodied in a computer program described above, has already proved useful, several modifications might further improve its results. In working directly from the win-loss matrix, it takes no account of the fact that a judge might have placed two or more objects in the same category. Furthermore, it does not separately maintain the identity of each judge, nor examine the extent to which his rankings were consistent with those of other judges. One can imagine that a still more sophisticated approach might explicitly incorporate ties, and might

somehow weight each judge's decisions according to the extent they were consistent with information from other judges. Such developments await the attention of an appropriately inclined methodologist.

## REFERENCES

- Ford, L. R., Jr. Solution of a ranking problem from binary comparisons. *American Mathematics Monthly*, 1957, 64 (8,II), 28-33.
- Pelz, D. C. and Andrews, F. M. *Scientists in organizations: Productive climates for research and development*. New York: Wiley, 1966.