Computer interactive heuristics to assess preference indefinability and resolve discrepancies between preference ratings and rankings

MICHAEL E. MILLS

Loyola Marymount University, Los Angeles, California

Ipsative and absolute assessment of preferences are contrasted, with particular reference to the different quantitative and qualitative information provided by each. Although it may not be possible to mathematically combine the respective qualitative information provided by rankings and ratings, when used in conjunction, they may offer a valuable consistency check of a respondent's preference assessments. A set of heuristics, interactively implemented via computer, is proposed to (1) identify inconsistencies between preference rankings and ratings of the same set of items, and (2) attempt to resolve those inconsistencies. The number of iterations required to resolve inconsistencies might be used as an index of the respondent's lack of preference definition.

Suppose that we requested a respondent to both rate and rank order his or her preferences for several desserts—say, pie, cake, pudding, and cookies. If the respondent's rank ordering was consistent with his or her ratings, we might have some confidence that this individual had fairly well-clarified preferences for these desserts. But suppose that we found an inconsistency between the respondent's rankings and ratings—for example, if he or she *ranked* pie higher than cake, but *rated* pie lower than cake. We might interpret this inconsistency as evidence that the respondent had fairly weak or unclarified preferences for these desserts, or perhaps he or she had misunderstood the instructions. More generally, given this inconsistency, we might wonder whether the rankings or the ratings constituted the more accurate representation of the respondent's true preferences.

The illustration above exemplifies a significant and often misunderstood problem encountered in the assessment of preferences. Scaling methods to assess preferences fall into one of two general categories—ipsative or absolute measurement. Although some investigators have naively assumed that both assessment methods yield equivalent information or are interchangeable, the differences between the quantitative and the qualitative sets of information provided by these two measures are profound.

Ratings (absolute measurement) assume an interval level of measurement, but they may be severely distorted by the respondent's response biases and styles. However, when compared to a normative sample, ratings may provide valuable information about interindividual differ-

ences. In contrast, rankings (an instance of ipsative measurement) only achieve an ordinal level of measurement (and thus may "throw away" information), and may force an indication of difference where no appreciable difference in preferences in fact exists. Valid inferences regarding between-person differences in preferences are virtually impossible when an ipsative assessment method is used (Hicks, 1970)—only *intraindividual* differences can be compared via ipsative assessment methods. Unfortunately, many investigators have inappropriately used ipsative measures to make *interindividual* comparisons, in an attempt to obviate the problems associated with response biases and styles (Hicks, 1970).

In addition, the qualitative information provided by ratings and rankings is different. Ratings typically assess preferences as if all options were equally attainable; rankings can provide information about relative priorities—which may be appropriate when all options are not in fact equally attainable and/or are mutually exclusive.

Thus, rankings and ratings each have their own unique liabilities—either method alone may often be suboptimal for the accurate assessment of preferences. Indeed, psychologists have attempted to understand the respective strengths and weaknesses of "absolute" and "ipsative" measurement, and a few aborted attempts have been made to combine them (Bauernfeind, 1962; Hicks, 1970). Moreover, neither of these methods alone assesses the reliability, or "clarity," of a respondent's preferences—the degree to which a respondent clearly *has* preferences and knows what they are.

Detecting Inconsistencies Between Rankings and Ratings

Although it may not be possible to mathematically combine the qualitatively different types of information pro-

Correspondence should be addressed to Michael E. Mills, Department of Psychology, 7101 W. 80th St., Loyola Marymount University, Los Angeles, CA 90045.

vided by rankings and ratings, when used in conjunction, they may offer a valuable consistency check of a respondent's preference assessments. Below, I will describe a set of heuristics that may be used to: (1) identify discrepancies (which hereafter will be termed *displacements*) between a respondent's preference ratings and rankings, (2) attempt to resolve such displacements by restructuring the respondent's rankings and/or ratings, and (3) assess the overall degree of preference indefiniteness (confusion or lack of preferences) exhibited by a particular respondent. A respondent's set of ratings for a list of items will be called the *ratings vector*. The same respondent's rank ordering of this set of items will be called the *rank vector*. The heuristics proposed to assess the consistency of these two vectors follow.

1. First, the respondent's ratings and rank vectors are sorted. Items in rank vector are sorted in ascending order (from lowest to highest). The ratings vector is sorted in descending order, since higher ratings typically indicate a stronger preference. The rank order of tied items in the ratings vector is determined according to which of the items has a higher (more preferred) position in the rank vector.

2. Given the sorted rank and ratings vectors, we may assess the displacement between the ranking and rating of a particular item. For example, consider the rank and ratings vectors below for a respondent's preferences for four desserts—pie, cake, pudding, and cookies. (The ratings are on a scale of 1–7, with 7 indicating *very high preference*.)

Rank Vector		Ratings Vector	
Rank	Dessert	Dessert	Rating
1	pie	cookies	7
2	cake	cake	6
3	pudding	pudding	4
4	cookies	pie	3

Note that the rating of pudding places it, after sorting, in the same rank position in the ratings vector that it has in the rank vector. That is, the item is rated relative to ratings of the other items in a way that is consistent with its ranking. There is no inconsistency, or "displacement," between the ranking and rating given to pudding.

However, we can observe a different situation for the item "pie." This dessert was assigned the highest ranking, but given the lowest rating. The respondent apparently is unsure about how he or she values this item, or, the ranking and/or rating instructions were not well understood. We can represent this confusion mathematically by calculating the differences in relative rank position of the item in the ratings and rank vectors. Specifically, this displacement may be represented by the absolute value of the difference of the rankings of the item in the ratings and rank vectors. Thus, where x is an item: $\text{Displacement}(x) = |J - I|$, where J = ranking of item x in the rank vector, and I = ranking of item x in the ratings vector. In other words, the displacement of item x is simply the absolute value of the difference, or offset, of the sorted rank position of x in the ratings and rank vectors.

We can determine the overall amount of inconsistency between the ratings vector and rank vector by summing the displacement scores of each of the items. Thus, overall displacement score = $\Sigma(x)$. The higher the overall displacement score, the greater the inconsistency between the rank and ratings vectors.

Some Heuristics to Resolve Rank/Ratings Conflicts

A simple but unrealistic scheme for resolving ranking/rating conflicts might be to repeatedly force the respondent to rerank and rerate all of the items until all inconsistencies between the rank and ratings vectors have been resolved. Unfortunately, this would rapidly exhaust the patience of even the most enthusiastic respondent.

A proposed alternative strategy is to have the respondent rerate only items with high displacement scores, and to rely on the collection of heuristics described in this paper to decide whether, and how, to modify the rank vector to better correspond to the new rating(s). Specifically, after each rerating of an item with a high displacement score, a simple heuristic is applied to determine whether, and how, the rank vector should be modified. One of two possible ways of changing the rank vector will be selected, a *swap* or a *scoot* (both of which are described below), based on which of the two operations results in a lower overall displacement score. This iterative process continues until the overall displacement score reaches some tolerable level.

Note that it is the heuristics, rather than the respondent, that determine whether the rank vector should be modified, and if so, how. This process is transparent to the respondent—he or she is modifying only the ratings of specific items.

Heuristic 1: Was the Rank or the Ratings Vector More Accurate?

Given an inconsistency between the rank and ratings vectors, how can we determine which of the two is a more accurate representation of the respondent's true preferences? As mentioned earlier, it is preferable to have the respondent rerate a few item(s) with the highest displacement score, instead of rerank all of the items, because it is simpler and requires less time. But if the resolution scheme only involves rerating certain items, how are errors in the rank vector to be corrected if the rank vector is in error? Let us make the assumption that the respondent's *rerating* of an item is more accurate than its original rating. This may be a reasonable assumption, since the respondent has already considered his or her preferences of all of the other items, in terms of both ranking and rating them, by the time the second rating of the item is performed.

When the respondent's rerating is different from the original rating of the item, there are two possible explanations of the original displacements between the items in the two vectors. Either the rank vector was originally in error (and it should be changed to better correspond

to the new ratings vector) or the old ratings vector was in error and has now been improved by the respondent's new rating; thus the rank vector should be left unchanged.

To make the determination of which one of the original vectors (rank or ratings) was more correct, we can compare the total displacement score of (1) the new and old ratings vectors and (2) the new ratings vector and the rank vector. The original vector with the higher displacement score is presumed to be a less accurate reflection of the respondent's preferences. If the rank vector has the higher displacement score, it will be restructured, as will be explained shortly. Otherwise, it is assumed that the new rating vector (based on the respondent's new rating[s] of an item[s]) is more accurate than the old rating vector, and the rank vector is left unchanged. These procedures are performed as follows.

When the ratings vector was in error. Consider the following rank and ratings vectors:

Rank Vector		Ratings Vector	
Rank	Dessert	Dessert	Rating
1	cookies	pie	5
2	cake	cookies	4
3	pie	cake	3
4	pudding	pudding	2

Note above that the item "pie" has the highest displacement score: 2. Thus, the respondent is requested to rerate his or her preference for pie (without reference to his or her previous rating). Suppose that the respondent gives pie a lower rating this time—say, a rating of 1. Let us compare the old and new rating vectors, with particular reference to the position of pie in the rank vector, as well as its position in the new and old rating vectors:

Rank Vector		New Ratings Vector		Old Ratings Vector	
1	cookies	cookies	4	pie	5
2	cake	cake	3	cookies	4
3	pie	pudding	2	cake	3
4	pudding	pie	1	pudding	2

Note that the new rating of pie is closer (has a lower displacement score) to its rank vector value than it is to its old ratings vector value. This suggests that the rank vector is a more accurate reflection of the respondent's preferences than the old ratings vector is. We assume that the old ratings vector was in error, and that the new ratings vector is more accurate. The old rank vector is left unchanged.

When the rank vector is in error. We should restructure the rank vector if the respondent's rerating suggests that the old ratings vector is more accurate than the rank vector. This would be the case if (1) the new rating given is identical to the old rating, or (2) if the displacement score between the new rating and the ranking of the item is greater than its displacement score between the new rating and the old rating. Suppose, for example, that the respondent had given pie a rerating of 7, instead of 1. Now the new rating of pie is more consistent (has a lower

displacement score) with its position in the old ratings vector than it is with its position in the rank vector. We assume that the rank vector is in error and that it needs to be restructured to better correspond to the old ratings vector.

Heuristic 2: Two Ways to Restructure the Rank Vector—Swapping or Scooting

Given that it is assumed that the rank vector is less accurate than the old ratings vector, there are at least two reasonable ways that the rank vector might be modified to achieve greater consistency with the old ratings vector.

1. We could *swap* the item with another item in that rank vector. For example:

Old Rank Vector		New Ratings Vector	
1	pie	pie	7
2	cake	cookies	4
3	pudding	cake	3
4	cookies	pudding	2

New Rank Vector			
1	pie		
2	cookies		
3	pudding		
4	cake		

Note above that the item "cookies" has been swapped with the item "cake" in the rank vector.

2. Alternately, instead of swapping positions with the item "cake," we could move the item "cookies" into the position occupied by the item "cake" and *scoot* the item "cake" down, along with all the items below it. This would result in the new rank vector below:

New Rank Vector	
1	pie
2	cookies
3	cake
4	pudding

If "cookies" had been rated lower than its original rating, rather than higher (as in the above example) items would have been scooted up instead of down. For example:

Old Rank Vector		Old Ratings Vector	
1	pie	pie	5
2	cookies	pudding	4
3	cake	cake	3
4	pudding	cookies	2

New Rank Vector			
1	pie		
2	cake		
3	pudding		
4	cookies		

Which operation should we execute, a swap or a scoot? To make this determination, both operations are performed. The resulting total displacement score for each

operation is calculated. The operation that produces a lower displacement score is the one that is used to restructure the rank vector.

Again it should be noted that the modification of the rank vector is performed without the knowledge of the respondent. We assume that since the rank vector is now more consistent with the ratings vector it should be a more accurate reflection of the respondent's true preferences. This, of course, may not be a correct assumption. If we were to continue from this point, the operations in the next iteration of the heuristics would be based on an inaccurate rank vector. Thus, after any modification to the rank vector, the rank vector is again presented to the respondent. If the respondent wishes to do so, he or she is given an opportunity to reconsider the sorting of the items, and rearrange the rank vector to result in what he or she believes is a more accurate ranking.

Looping to Resolution

The above operations are repeated until the total displacement score reaches zero, or until an acceptable tolerance level is reached. When the number of items is large, it becomes progressively more difficult to achieve a total displacement score of zero, particularly if the respondent does not have well-defined preferences. Thus, we can say that the inconsistencies are "acceptable" if the overall displacement is less than or equal to some predefined tolerance level. For example, given 10 items, one might rather arbitrarily define this tolerance level to be, say, 3. Once the total displacement score has reached this tolerance level, further attempts at resolution are terminated.

There may be rare instances when, either because the respondent's preferences are truly undefined or because he or she did not understand the instructions, the overall displacement score is not reduced with each iteration, but may stabilize, or even increase. This would tax the patience of even the most cooperative respondent, since the program would loop infinitely. A mechanism is needed to detect this situation, and to permit a graceful exit. If the total displacement score has not improved after some specified number of iterations, then exit. Note that this should not necessarily be considered a failure of the heuristics described here.

Assessing Preference Indefinition

Indeed, the fact that we have determined that the respondent's preferences are essentially undefined is valuable information in itself. We might otherwise have used the respondent's ratings as valid data under the false assumption that they represented clear preference differentials. Measures of preference indefinition include: (1) the total displacement score at the outset, (2) the number of iterations required to achieve "resolution," and/or (3) some combination of the two.

Such measures of preference indefinition may themselves prove to constitute a useful index. For example, individuals with little prior experience of the activities or objects that they are asked to rate and/or rank are likely to have relatively undefined preferences, in comparison with individuals with extensive experience of them. A wine connoisseur is likely to have more clearly defined preferences for varieties of wine than will individuals with less epicurean tastes. Students with little work experience are less likely to have well-defined preferences for specific occupational rewards than mature employees with years of experience. One might argue that to use such student ratings and/or rankings of occupational rewards as a basis to recommend certain occupations to them may not be justified. In preference and opinion surveys, the inclusion of data from individuals with undefined or weak preferences for the items may erroneously skew the data (or at least add "noise"). Thus, an index of degree of preference indefinition may be an important variable in certain studies, and may deserve more attention from researchers.

Summary of the Heuristics

1. Sort the rank and ratings vectors.
2. Identify the item (x) with the largest displacement score.
3. If the displacement of x is less than or equal to some predefined tolerance level, then stop. If the total displacement score has not improved after a predefined number of iterations, then assume that client is unable to resolve discrepancies, and exit. Otherwise, continue to the next step.
4. Ask the respondent to rerate x . Substitute the new rating for the old rating of x . Compare the new ratings vector to the old ratings vector and the ranking vector. If the rating is "closer" to the ranking vector, then return to Step 1. Otherwise, continue to the next step.
5. The rank vector will be restructured to make it more consistent with the ratings vector. To do this, either a scoot or a swap operation will be used. We calculate which of the two operations produces the lowest total displacement score, and we use that operation to restructure the rank vector. We allow the respondent to modify the rank vector after this operation, and then we return to Step 1.

Program Availability

An interactive computer program to demonstrate these heuristics is available. Simply send a formatted 3.5-in. disk in a returnable mailer to the author.

REFERENCES

- BAUERNFEIND, R. H. (1962). The matter of "ipsative" scores. *Personnel & Guidance Journal*, **41**, 210-217.
- HICKS, L. E. (1970). Some properties of ipsative, normative, and forced-choice normative measures. *Psychological Bulletin*, **74**, 167-184.