Sawtooth Software

RESEARCH PAPER SERIES

Product Mapping with Perceptions and Preferences

Richard M. Johnson,
Sawtooth Software, Inc.
1998

530 W. Fir St.
Sequim, WA 98382
(360) 681-2300
www.sawtoothsoftware.com

Product Mapping with Perceptions and Preferences

Richard M. Johnson
Sawtooth Software, Inc.

Background

In the '50s and '60s, mathematical psychologists developed theories about how perceptions and preferences might be related. They considered objects to be arranged in some kind of perceptual space, determined either with respect to perceived similarities, or with respect to ratings on descriptive attributes. Each individual was also thought to have an ideal direction in the space and to prefer objects that were farther in that direction, or to have an ideal point in the space and to prefer objects closer to that point.

Market researchers have found these ideas very fruitful. Use of product maps became widespread in marketing research in the '60s and '70s, and they have proved to be useful aids for thinking about differences among products, customer desires, and ways in which products might be modified to become more successful.

Perceptual data have been used most often to create product spaces. In early years judgements about overall similarity of pairs of products were used with multidimensional scaling techniques. However, in later years attribute ratings have been used more widely, analyzed with factor analysis, discriminant analysis, or correspondence analysis.

Preference data have also been used to develop product spaces in marketing research. The first techniques for making maps based on preferences were developed in the early '60s: Coombs' Unfolding method (which assumed each individual had an ideal point) and Tucker's Points of View approach (which assumed each individual had a preferred direction in space). In an important contribution in 1970, Carroll and Chang showed that vector models can be regarded as special cases of generalized ideal point models, and they also provided the first practical method for estimating ideal points.

Most methods for making product maps have used either perceptual data or preference data, but seldom both. And there have been problems with maps of both types.

Maps based on **perceptions** are easy to interpret and good at conveying insights, but they are often less good at predicting individual preferences. One reason is that they may focus on differences that are easy to see but less important in determining preferences.

Maps based on **preferences** are better at accounting for preferences, but their dimensions are sometimes hard to interpret. For example, consider cups coffee with differences in temperature ranging from boiling to tepid. Most of us would probably prefer some middle temperature and reject both extremes. But if no perceptual information is available to establish their differences on the underlying temperature scale, the most extreme cups may be close together in a preference-based map, because their only recognized property is that they are both rejected by nearly everyone.

There is still another problem with **aggregate** maps of both types: a product's position on a map is based on the average of many individuals' perceptions or preferences. Because individuals differ, a single map can seldom describe different individuals' perceptions or preferences very precisely.

At the previous (1997) Sawtooth Software Conference, John Fiedler and Terry Elrod presented papers analyzing the same data but using different methods. John used a discriminant-based method which considered only **perceptual** data in the form of attribute ratings, and Terry used a technique he had developed which considered only **preference** data. Their maps were surprisingly similar. This reinforced the underlying theory relating perceptions and preferences, and suggested that even better maps might be produced if based on *both* perceptions and preferences. In the few instances where both types of data have been used, the usual practice has been first to use perceptual data to make the map, and then to fit preference data to it "after-the-fact." By contrast, because the methods described here use *both* perceptual and preference data simultaneously, we call them "composite" methods.

Composite Methods

We have developed both "vector" and "ideal point" models. Each model uses both **perceptual** data, consisting of product ratings on attributes, and **preference** data, consisting of paired-comparison preference ratings for products. These are the same types of data as provided by APM, a perceptual mapping product released by Sawtooth Software in 1985. In fact, both composite models use APM data files, although they make no use of product familiarities or explicit ideal point ratings.

These two new models share several characteristics:

> Every respondent has a unique perceptual space, determined by his/her own ratings of products on attributes.
>
> Each dimension in the individual's space is a weighted combination of his/her ratings of products on attributes.
>
> However, the attribute weights defining the dimensions are required to be identical for all respondents.
>
> Those attribute weights are determined by optimizing the fit (over all individuals) between actual preferences and the preferences inferred from the individual perceptual spaces.

The overall product map is just the average of the individual maps. It is also a weighted combination of average product ratings, so it is truly a ***perceptual*** map, although with dimensions chosen so as to account best for preferences. In this way we produce maps which are firmly grounded in descriptive attributes, but which better account for individual preferences.

Both models require the estimation of two sets of parameters. One set consists of **attribute weights**, identical for all individuals, to be applied to attribute ratings to obtain the dimensions of individuals' perceptual spaces. The other parameters are unique for each individual: either **individual importance weights** in the case of the vector model, or **individual ideal point coordinates** in the case of the ideal point model. For both models, estimation is done using alternating least squares.

The ***vector model*** assumes that each individual has some preferred direction in space, and prefers products that are "farther out" in that direction. It is appropriate for product spaces that have dimensions where "more (or less) is always better." An initial guess is made at the attribute weights. Based on the implied perceptual spaces, the best-fitting importance weights are estimated for each individual. Then, given those importance weights for all individuals, an improved set of attribute weights is estimated. The procedure alternates between estimation of individual importance weights and common attribute weights, continually improving the goodness of fit, as measured by an r-square value. To aid interpretability, the overall map is constrained to have orthogonal dimensions, and each dimension is scaled so that the sum of squared product coordinates is equal to the square of the number of products.

The ***ideal point model*** assumes that each individual's liking for products depends on products' perceived distances from an ideal point. Such models are more appropriate for product categories in which respondents may prefer combinations of attributes corresponding to interior regions of the space. Our approach is somewhat simpler than that of Carroll and Chang in PREFMAP. To keep things simple, we assume that individual preference contours are circular; in other words, we do not permit individuals to weight dimensions differently. Also, our implementation of the ideal point model assumes that each individual's ideal point is interior to the convex hull of his/her perceived product locations. As a result, our vector model is not a special case of our ideal point model.

Estimation of the ideal point model is done by minimizing a weighted sum of squared distances. From each respondent's preference data we obtain a set of positive weights that sum to unity, similar to shares of preference in conjoint analysis. We also compute the squared distances from each respondent's ideal point to each product in his/her perceptual space. Finally, we sum the products of those squared distances times the corresponding preference weights. We minimize this sum over all respondents, producing a solution in which more preferred products have smaller distances from ideal points. (This approach has some similarity to that of Desarbo and Carroll (1985), although they use weights based on preferences to scale discrepancies between observed and predicted distances, whereas we use the weights to scale the distances themselves.)

If the preference weights were all equal, then the algorithm would minimize the sum of squared distances from respondents' ideal points to all of the products, and all ideal points would be estimated to be at the center of the space. At the other extreme, if the preferred product had a weight of unity and all others had weights of zero, each individual's ideal point would be estimated as coincident with his or her preferred product. As with conjoint simulators, one can choose an scale factor to apply to the preference weights that produces any behavior between those two extremes. For our examples we use an scale factor of 10, which requires each individual's ideal point to be quite close to his or her preferred product. However, there can be a lot of heterogeneity in the way individuals perceive brands. Even if each person's ideal point were made to coincide perfectly with his or her preferred brand, ideal points could still be dispersed all over the aggregate map.

As with the vector model, we constrain the overall map to be orthogonal, with the sum of squared product coordinates for each dimension equal the square of the number of products. This avoids degenerate solutions in which all products and ideal points are coincident.

In what follows, we compare composite mapping results to those from discriminant analysis, using two data sets. The first data set is an artificial one, deliberately constructed to show the potential superiority of composite methods. The second data set consists of real data for motorcycles.

An Artificial Example

The first data set consists of perceptual and preference data for 300 artificial respondents, with three attributes and eight products. Imagine the product category to be busses used in rush hour commuting in a large city. The first attribute is Color of the bus, with levels red or blue. Color is deliberately chosen as an attribute on which people would agree which color a bus actually had, but which would have little impact on preference. The second attribute is Speed with levels fast and slow. The third attribute is Roominess with levels of roomy and cramped.

The eight busses to be rated have all combinations of levels of the three attributes. The perceptual data were made heterogeneous by adding an independent random variable to each product's design value for each individual. Heterogeneity for Color was only half as great as for Speed and Roominess.

Respondent ideal points were also random, with mean near the center of the scale for each attribute, and with a large amount of random preference heterogeneity. Respondents' preference data were generated by constructing constant sum paired comparison answers for 12 pairs of bus concepts. Preferences were deliberately constructed so as not to be affected by Color. The sum of squared differences was first computed between each respondent's ideal point and his/her perception of each bus, considering only Speed and Roominess. Reciprocals of those sums were then exponentiated and percentaged, to simulate paired comparison preference values between 0 and 100.

Perceptual and preference data are like those used by Sawtooth Software's APM system, which uses discriminant analysis to construct perceptual maps. Discriminant analysis has optimal mathematical properties, guaranteeing that its maps will contain the greatest amount of information about how products are seen to differ from one another for a given number of dimensions.

Although we have data on three attributes, we desire a map using only two dimensions.
The results for APM's perceptual map are given in Table 1 and Figure 1. The first dimension consists almost entirely of Color. This is expected, since discriminant analysis accounts for as much variance as possible with its first dimension, and our variables are orthogonal, with Color

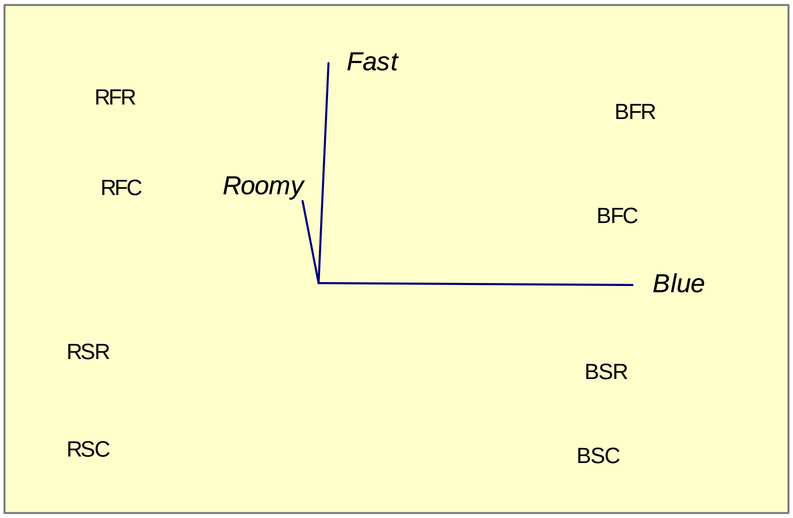
having the least amount of disagreement about which color each bus has. The second dimension consists mostly of a combination of Speed and Roominess. The third dimension, if we had included it, would have consisted of another combination of Speed and Roominess, which would account for the balance of the systematic differences among products.

Table 1
Discriminant Map for Synthetic Data

	1	2
Blue/Red	1.00	-0.01
Fast/Slow	0.03	0.95
Roomy/Cramped	-0.05	0.35
RSR	-8.61	-3.12
RFR	-7.73	7.90
RSC	-8.62	-7.30
RFC	-7.48	3.99
BSR	7.74	-3.91
BFR	8.94	7.34
BSC	7.46	-7.66
BFC	8.29	2.76

The second panel of Table 1 shows the coordinates of the 8 products in the discriminant space. Each product is identified by a three-letter string. The first letter is R or B indicating whether the bus is red or blue. The second letter is S or F, indicating whether it is slow or fast. The final letter is R or C, indicating whether it is roomy or cramped. As expected, the four red busses are all at one end of the first dimension and the four blue busses are all at the other end. The second dimension is characterized mostly by differences in Speed, with a small amount of information about differences in Roominess.

Figure 1
Discriminant Map



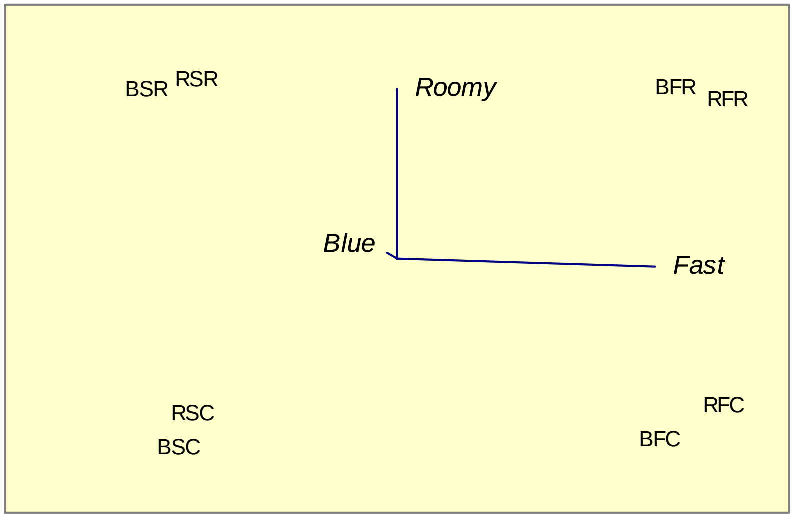
Since we constructed preferences to depend on Speed and Roominess but not Color, the first dimension of this space is useless for explaining preferences, and we have failed to capture important information about Roominess and Speed that would be required to account well for preferences.

We next consider maps of the same data produced using both perceptual and preference data. First, here are results for the vector map:

Table 2
Composite Vector Map for Synthetic Data

	1	2
Blue/Red	-0.04	0.04
Fast/Slow	0.99	-0.05
Roomy/Cramped	0.00	1.00
RSR	-0.92	1.05
RFR	1.11	0.97
RSC	-0.93	-0.92
RFC	1.10	-0.87
BSR	-1.13	0.99
BFR	0.92	0.98
BSC	-1.01	-1.12
BFC	0.86	-1.07

Figure 2
Composite Vector Map

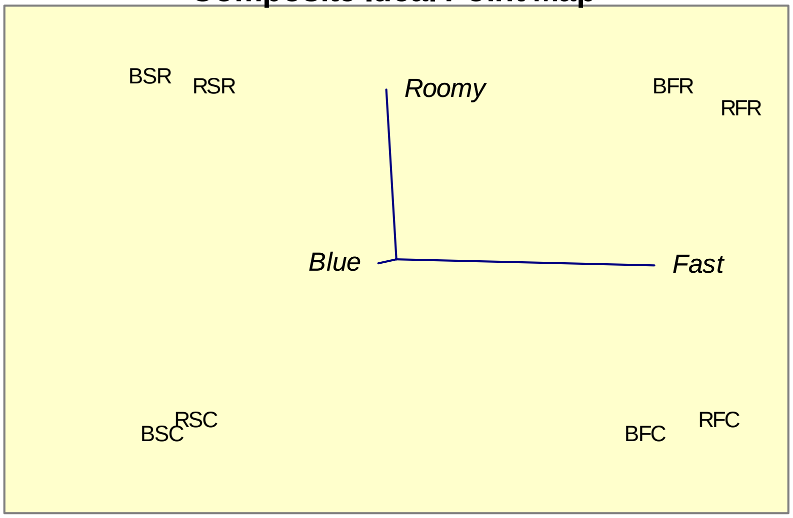


The first panel of Table 2 shows that Color scarcely enters into either dimension, and that both dimensions are concerned almost solely with Speed and Roominess, which we know are required to account for preferences in this example. The second panel shows that the corresponding red and blue products occupy similar positions in the space.

Composite maps may be subjected to any orthogonal rotation, so we have chosen to make the vectors for Speed and Roominess nearly horizontal and vertical. It is clear that the corresponding red and blue products occupy similar positions in the space. The small differences in location between corresponding products are due to the random heterogeneity of perception that was used in construction of the data file.

More important, this map accounts for preference more successfully than the discriminant map. If we find the direction in each map which best accounts for each respondent's preferences, we get an average r-squared between predicted and actual preferences of .59 for the discriminant map, vs. .74 for this map. We may also count the average number of correct orders for pairs of products when comparing actual rank orders of preference vs. predicted rank orders. For the discriminant map, 74% of the pairwise comparisons are correct, vs. 80% for this map. We should not expect either map to work perfectly because the vector model assumes that individual ideal points are infinitely far from the center of the space, and we constructed this data set so that a large proportion of the ideal points were near the center of the space.

Figure 3
Composite Ideal Point Map



We next consider a map of the same data produced using a composite ideal point map.

Table 3
Composite Ideal Point Map for Synthetic Data

	1	2
Blue/Red	-0.07	-0.02
Fast/Slow	0.99	-0.03
Roomy/Cramped	-0.04	1.00
RSR	-0.85	1.01
RFR	1.17	0.89
RSC	-0.92	-0.96
RFC	1.09	-0.96
BSR	-1.12	1.08
BFR	0.91	1.01
BSC	-1.07	-1.04
BFC	0.80	-1.04

The ideal point map, like the preceding vector map, ignores Color almost entirely, concentrating on Speed and Roominess. As with the vector map, the products that are identical except for Color are nearly superimposed, and the small differences between them are due to the random heterogeneity built into the perceptual data.

This map also accounts for preferences far more successfully than the discriminant map. We can use the preference data to estimate individual ideal points for both maps, and then compute distances from each product to each respondent's ideal. We can evaluate the performance of each map in accounting for preference by counting the number of product pairs for which the preferred product is closer to the respondent's ideal point. For the discriminant map, 79% of the pairs are correct, and for this composite ideal point map 92% of the product pairs are correct. This map does not provide perfect prediction because it restricts ideal points to lie within the convex hull of the product points, and the data set was constructed so that many of them actually lie outside that region.

In addition to the results shown, the composite vector mapping approach also estimates an ideal direction for each respondent, and the composite ideal point mapping approach estimates an ideal point for each respondent. We don't show those because the ideal points were constructed so as to comprise an undifferentiated "blob" of little interest. However, we shall consider individual preference information for the next data set.

Motorcycles

The second data set concerns road-going motorcycles, and was contributed by Tom Wittenschlaeger of Hughes Aircraft Company and John Fiedler of POPULUS, Inc. The data were collected in the United States in 1994 from a sample of 150 motorcycle riders. The project was methodological rather than substantive, so the data should not be used to develop marketing strategy, but serve nicely to illustrate mapping techniques.

The interview was typical of an APM questionnaire. Each respondent rated the importance of 10 attributes, and his/her familiarity with 11 motorcycle brands, and then rated the most familiar 5 motorcycles on the 5 most important attributes. The respondent's "ideal motorcycle" was also rated on the same scales. Finally, eight random pairs of motorcycles were presented and the respondent was asked to allocate 100 points between the members of each pair, indicating the relative likelihood of choosing each one in a purchase situation. Neither attribute importances ratings, product familiarity ratings, nor explicit ideal points were used in this analysis.

The attributes and products rated were as follows. Each has been given a short label to identify it on the maps:

Attributes:

Image	Has the image I prefer
Safe	Meets high standards of safety
Perform	Has high performance
Unique	Has a unique look and feel
Value	Offers good value for the money
Service	Has excellent service and support
Quality	Has high quality
Style	Has beautiful styling
Engin	Has excellent engineering
Fun	Is fun to ride

Products:

HON	Honda
KAW	Kawasaki
SUZ	Suzuki
YAM	Yamaha
DUC	Ducati
GUZ	Moto Guzzi
BIM	Bimota
BMW	BMW
TRI	Triumph
NOR	Norton
HAR	Harley Davidson

We have no market data with which to compare inferences from maps, but we do have preference data from the same respondents. Recall that each respondent selected the five products with which he or she was most familiar, and then answered paired comparison preference questions for eight random pairs of those products. We can accumulate the average preference proportions awarded to each product, which are shown in Table 4.

Table 4
Average Preference Percentages

HAR	71
HON	60
BMW	54
KAW	47
YAM	42
DUC	42
GUZ	41
SUZ	40
TRI	39
BIM	38
NOR	37

We should expect Harley Davidson, Honda, and BMW to have positions on the map indicating relative desirability.

The values in Table 4 do not reflect differences due to familiarity. Harley Davidson and the four Japanese brands were familiar to many respondents, while Bimota, Ducati, Moto Guzzi, and Bimota were familiar only to few. The more familiar brands will have more influence in determining the structure of the maps.

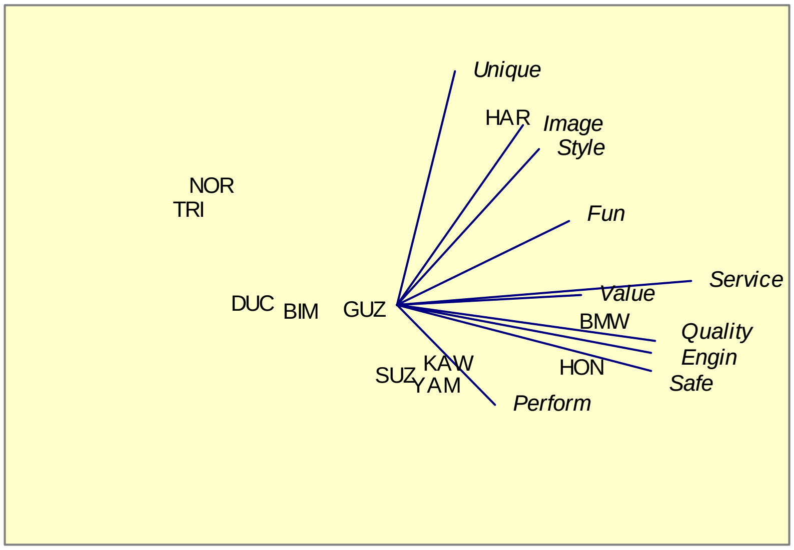
We now compare the maps produced by discriminant analysis with those produced by the two composite methods, using both vector and ideal point models.

Table 5
Discriminant Map for Motorcycle Data

	1	2
Image	0.32	0.60
Safe	0.65	-0.17
Perform	0.25	-0.33
Unique	0.15	0.78
Value	0.47	0.03
Service	0.75	0.08
Quality	0.66	-0.12
Style	0.36	0.52
Engin	0.65	-0.16
Fun	0.44	0.28
HON	0.37	-0.21
KAW	0.02	-0.20
SUZ	-0.10	-0.24
YAM	-0.01	-0.27
DUC	-0.47	-0.00
GUZ	-0.18	-0.02
BIM	-0.34	-0.05
BMW	0.42	-0.06
TRI	-0.64	0.32
NOR	-0.58	0.40
HAR	0.18	0.62

The correlations of the attributes with the two largest dimensions in the discriminant space are given in Table 5. This map has been rotated so all attribute vectors point toward the right side of the space. These attributes would all be regarded as favorable by most motorcycle riders, so one would expect the preferred brands to be toward the right side of the space. The graphical representation of these data is given in Figure 4. (The product coordinates have been scaled down by a factor of .25 to make their average absolute values approximately equal to those of the correlations.)

Figure 4
Discriminant Map



Harley Davidson is alone in the upper right quadrant, with large projections on Uniqueness, Image, Style, and Fun. BMW and Honda are in the lower right quadrant, with large projections on the Service, Value, Engineering, Quality, Safety, and Performance. BMW has larger projections on the attributes pointing upwards, and Honda is larger on Performance. The British brands Norton and Triumph are in the upper left quadrant, with large projections on Unique, Image, and Style. The three Italian brands are in the lower left quadrant, rather close to the center, and the Japanese brands Kawasaki, Suzuki, and Yamaha are toward the bottom of the map, with high projections on Performance.

In APM questionnaires the respondent is asked to describe his ideal product on the same attributes as he describes existing products. These explicit ideal points can also be incorporated into the map, although there is some question about the reasonableness of this procedure with attributes where the ideal levels might be at infinity. Rather than show all 150 respondents' explicit ideal points individually, we have done a cluster analysis, and show the locations of the centers of three clusters in Table 6.

Table 6
Centroids of Ideal Point Clusters in Discriminant Map

	1	2
Cluster A (39%)	1.16	-0.07
Cluster B (36%)	0.22	0.06
Cluster C (25%)	0.74	1.32

Cluster A, which contains 39% of the respondent sample, has an average position much farther to the right than any product, and slightly below the horizontal axis. Those respondents would be expected to favor BMW and Honda.

Cluster B, which contains 36% of the respondent sample, has an average position slightly to the right of and above the origin. Those respondents might be expected to prefer any of the products.

Cluster C, which contains 25% of the respondent sample, has a position far to the right, and very far above all the products, approximately in the direction of the Image vector. Those respondents would be expected to have very strong preferences for Harley Davidson.

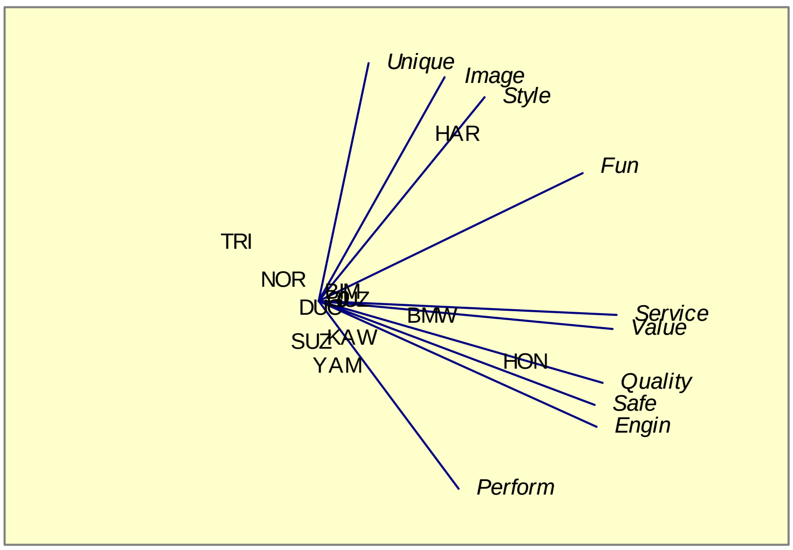
We turn now to the composite vector map, for which data are given in Table 7.

Table 7
Composite Vector Map for Motorcycle Data

	1	2
Image	0.40	0.91
Safe	0.88	-0.43
Perform	0.45	-0.77
Unique	0.16	0.97
Value	0.94	-0.12
Service	0.95	-0.06
Quality	0.91	-0.34
Style	0.53	0.83
Engin	0.89	-0.42
Fun	0.84	0.52
HON	2.11	-1.01
KAW	-0.11	-0.62
SUZ	-0.57	-0.70
YAM	-0.29	-1.01
DUC	-0.49	-0.13
GUZ	-0.13	-0.14
BIM	-0.15	-0.01
BMW	0.90	-0.27
TRI	-1.58	0.93
NOR	-0.96	0.32
HAR	1.26	2.69

This map has also been rotated so all attribute vectors point toward the right side of the space. It has strong similarities to the one produced by discriminant analysis. Harley Davidson is again alone in the first quadrant, with high projections on Unique, Image, Style, and Fun. BMW and Honda are again in the lower right quadrant, with strong projections on the remaining six attributes. The British products are again in the upper left quadrant, but this time much closer to the center of the space. The Italian products are very close to the center, and Suzuki, Kawasaki, and Yamaha are again in the lower left quadrant. The main apparent differences are that BMW is now much closer to the center than Honda, and the relatively low-rated products, which are those on the left side of the space, have moved toward the center.

Figure 5
Composite Vector Map



The fact that the discriminant map looks much like the composite map suggests that none of the attributes captures large but unimportant differences. This map is the average of 150 individual maps, on which respondents have different opinions about the positioning of products. The fact that the less popular products are closer to the center in the aggregate map suggests that they are in fact preferred by some respondents, who view them more favorably in terms of these attributes.

Since this is a vector map, individual preferences are given by “ideal direction” in space. We summarize those data by a count of the proportion of individuals whose ideal direction lies in each of eight “compass directions.”

Table 8
Summary of Ideal Vector Locations

Bearing	Percent
0 to 45 degrees	31
46 to 90 degrees	26
91 to 135 degrees	20
136 to 180 degrees	7
181 to 225 degrees	6
226 to 270 degrees	0
271 to 325 degrees	2
326 to 360 degrees	6

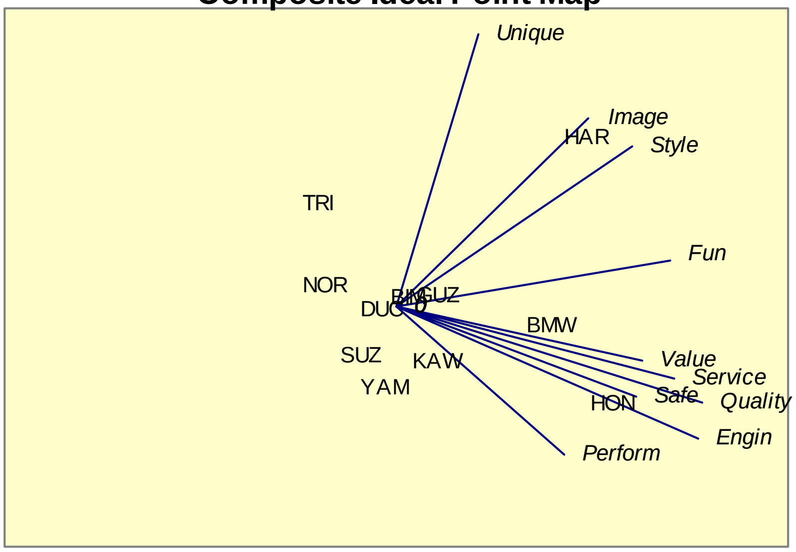
Nearly all respondents' ideal directions are in the right side of the space, with 57 percent in the upper-right quadrant and 27% in the lower right quadrant. These results are also similar to those for the discriminant map, which suggested that the strongest demand was for products in the upper right quadrant. In fact, this map and the discriminant map are very similar in terms of fitting preferences. If we find the direction in each map which best accounts for each respondent's preferences, and then count the percentage of correct orders for pairs of products when comparing actual rank orders of preference vs. predicted rank orders, we get 92% correct for the discriminant map vs. 93% correct for this map.

Finally, we consider the composite ideal-point map, for which data are given in Table 9.

Table 9
Composite Ideal Point Map for Motorcycle Data

	1	2
Image	0.49	0.63
Safe	0.61	-0.30
Perform	0.43	-0.49
Unique	0.21	0.91
Value	0.63	-0.18
Service	0.71	-0.24
Quality	0.78	-0.32
Style	0.60	0.54
Engin	0.77	-0.44
Fun	0.70	0.16
HON	1.80	-1.29
KAW	-0.02	-0.73
SUZ	-0.76	-0.67
YAM	-0.55	-1.09
DUC	-0.56	-0.05
GUZ	0.02	0.15
BIM	-0.24	0.02
BMW	1.15	-0.25
TRI	-1.23	1.37
NOR	-1.15	0.28
HAR	1.54	2.26

Figure 6
Composite Ideal Point Map



This map is very similar to the vector map of Figure 5. Both products and attribute vectors are in similar positions, and it seems doubtful that one would reach different conclusions from the two maps.

The mapping computation creates a file of individual ideal point estimates. We have subjected those to a cluster analysis, and have chosen to report three clusters. The coordinates of their centroids are given in Table 10.

Table 10
Centroids of Ideal Point Clusters in Ideal Point Map

	1	2
Cluster A (62%)	2.61	0.63
Cluster B (27%)	13.49	1.95
Cluster C (11%)	15.72	14.19

It may seem surprising that clusters B and C have locations far to the right of all of the products, when ideal points are required to be close to those individuals' preferred products. The reason for this is that there is considerable variation in individual product perceptions. Some individuals see their preferred product as very far to the right of its average location, and their ideal points are estimated to be near those locations.

As with the discriminant and vector maps, there is clear evidence of preference for products to the right of and above the horizontal axis. The largest cluster (62%) is centered to the right of and slightly above the origin. Since there is a lot of dispersion within each cluster, those individuals might prefer any of the products.

The second largest cluster (27%) is centered very far to the right, and moderately above the horizontal axis. Those respondents seem likely to favor Harley Davidson, Honda, and BMW.

The third largest cluster (11%) are extremely far to the right and extremely high, in the direction of Harley Davidson but much farther. They seem likely to be strong Harley preferrers.

This map also accounts for preferences slightly more successfully than the discriminant map. We can use the preference data to estimate individual ideal points for both this map and the discriminant map, and then compute distances from each product to each respondent's ideal. We can evaluate the performance of each map in accounting for preference by counting the number of product pairs for which the preferred product is closer to the respondent's ideal point. For the discriminant map, 89% of the pairs are correct, and for this composite ideal point map 92% of the product pairs are correct.

Estimating Demand

Since composite maps provide a relatively tight linkage between perceptions and preferences, it is tempting to consider ways of using composite maps to estimate demand for new products. There are at least two ways to do so.

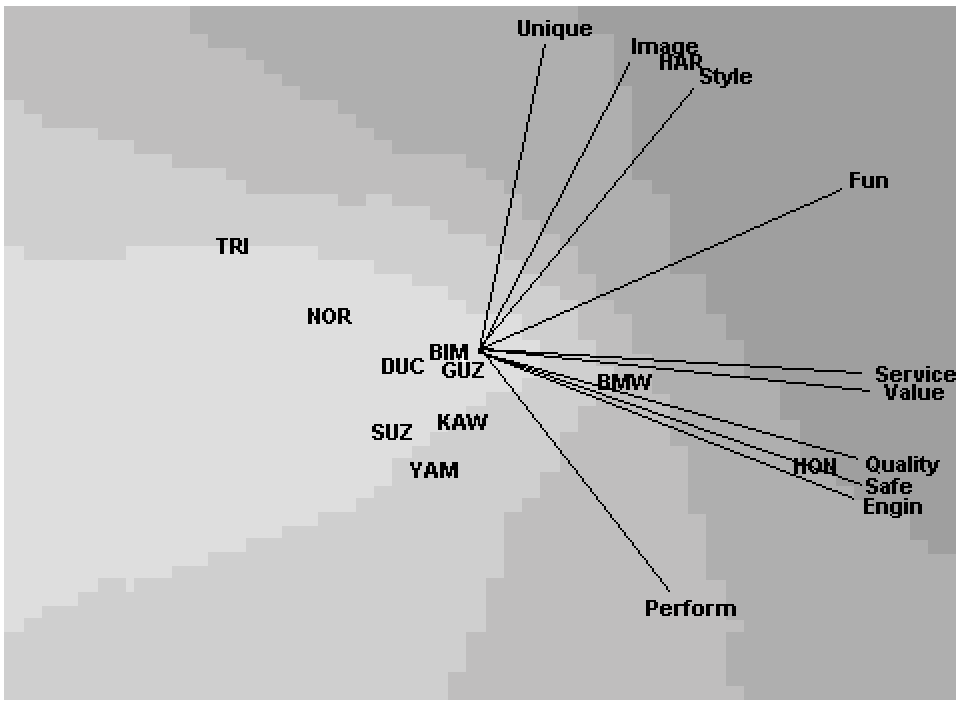
One way might be to construct a simulator to predict preferences for new products, or products modified on one or more attributes. That approach was used in Sawtooth Software's APM product, although the manual listed several reasons why it might not be completely successful. The basic problem is that there is a many-to-one mapping of attributes into dimensions. Products can have quite different levels on several attributes, and yet occupy the same point in space. Multicolinearity among attributes makes it difficult to infer the relative importance of each. Also, if a product is changed on one attribute without corresponding changes on others, then the space itself may change, and with it the capability of making inferences.

The other way to estimate demand for modified products is to consider products that differ in terms of locations in the existing space, rather than differing on specific attributes. That is the approach used in what follows.

Figures 7 and 8 present that information for the composite vector and ideal point maps, respectively. Relative demand is estimated using a "first choice rule." For the vector map, we see whether a product at each grid point would be the "farthest out" in each respondent's ideal direction (relative to other products), and score a 1 if so and a 0 if not. For the ideal point map we see whether a product at each grid point would be closest to each respondent's ideal point (relative to other products), and score a 1 if so and a 0 if not. The total number of hits is thus computed for a hypothetical new product at each grid point. The grid points are divided into quintiles in terms of their numbers of "hits," and each point is given background shading to indicate its quintile. The darker regions indicate higher demand.

As is evident in Figure 7, which provides a display for the composite vector model, the demand for new products would be greatest if they were positioned far to the right and in the upper half of the space.

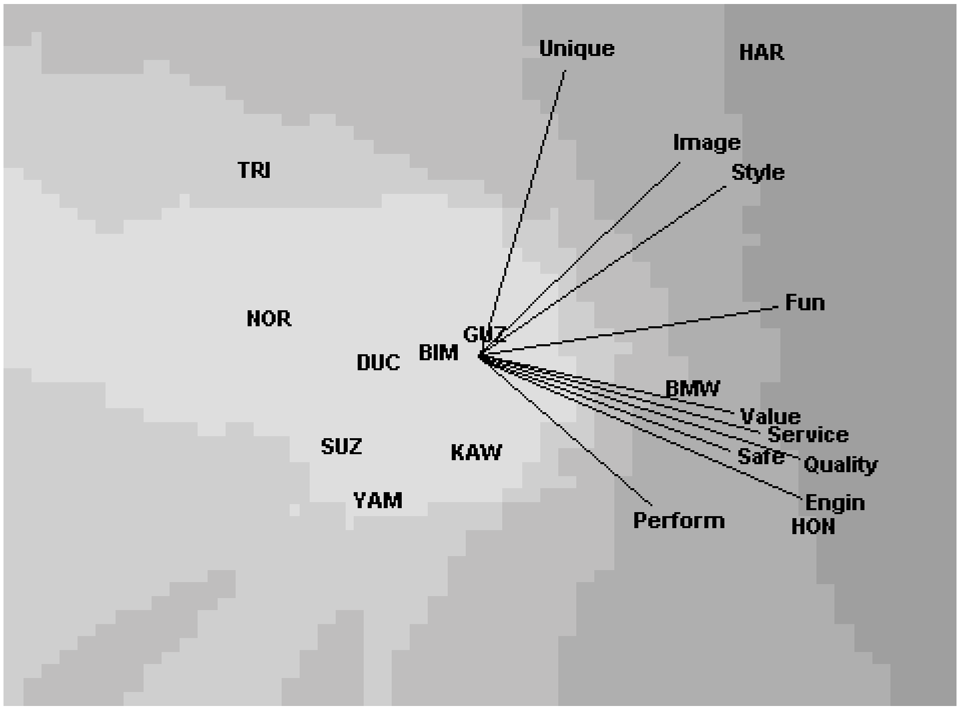
Figure 7
Density of Demand for Vector Model



Similar conclusions would be reached from Figure 8, which provides a similar display for the composite ideal point model, although in this map the region of highest demand seems to extend somewhat further downward, suggesting that there is more opportunity for the attributes pointing in that direction.

Their ability to portray relative demand for new or modified products is a major benefit of composite maps. Because composite maps are based on simultaneous analysis of perceptual and preference data from the same individuals, they have a strong advantage in this regard over other methods, which enhances their usefulness to managers. (However, we should repeat our earlier caveat, that the data for our examples were collected for methodological rather than substantive purposes, and these particular maps should not be used for marketing strategy purposes.)

Figure 8
Density of Demand for Ideal Point Model



Although the two maps produce similar information about likely demand for new products, it is reasonable to inquire which is better. The vector model accounts for preferences slightly more effectively, correctly fitting 93% of the pairwise judgements as opposed to 92% for the ideal point model. However, the product points seem to be more spread out in the ideal point map, suggesting that it supports finer distinctions among products. It would be premature to decide that one method is better than the other based just on this data set.

Summary and Conclusions

We have introduced techniques for Composite Mapping, which make use of both perceptual and preference data. The basic idea is that each respondent has an individual map in which product locations are weighted combinations of attribute ratings, and the weights are identical for all individuals. The aggregate map is the average of the individual maps. The weights used for all individuals are determined so as to maximize the correspondence between individuals' stated preferences and the preferences that would be inferred from the resulting maps. There are separate algorithms for vector maps and ideal-point maps.

Our results lead to these conclusions:

> Although mapping based on perceptual data alone can portray product images efficiently in maps of few dimensions, it can err by concentrating on differences among products that are easy to see but not important for preference.
>
> For an artificial data set in which two attributes were involved in preferences but a third had larger perceived differences among products, a perceptual map using discriminant

analysis failed to account for preferences, but both composite mapping methods reproduced the known preference structure of the data.

Although mapping based on preference data alone may be successful at explaining product preferences, the lack of perceptual information may lead to maps that are difficult to interpret.

For a real data set, the perceptual map using discriminant analysis predicted preferences quite well, and was visually very similar to both the composite vector map and the composite ideal point map. This should occur when the attributes are approximately equal in importance for predicting preferences.

The fact that the composite methods appear to produce better results when attributes differ strongly in their importances in affecting preference, but similar results when attributes are well chosen, suggests that composite maps can provide insurance against unfortunate choices of attributes

Since composite maps provide a relatively tight linkage between perceptions and preferences, they may be used for forecasting relative demand for new or modified products.

All in all, there seems to be no downside to using composite mapping methods, and the benefit of possibly improved interpretation and prediction can be great.

Appendix

Estimation of Composite Mapping Models

The vector model and the ideal point model both use the following definitions:

Let there be ***N*** respondents, ***n*** products, ***p*** attributes, ***d*** dimensions.

For the i-th individual, let:

$\mathbf{X_i}$ = an (n x p) matrix of product ratings, with column sums of zero.

$\mathbf{Y_i}$ = an (n) vector of product preference values. APM provides constant-sum preference information obtained by having respondents divide 100 points among members of each of several product pairs. We construct $\mathbf{Y_i}$ by taking logits of those preference percentages, and awarding half the logit value to the winning product in that pair, and penalizing the losing product with half of the logit value. This results in $\mathbf{Y_i}$ values similar to conjoint utilities and that sum to zero for each individual.

$\mathbf{T}$ = a (p x d) matrix of weights used to transform attribute ratings into dimensional coordinates. $\mathbf{T}$ is common to all respondents. We want to find a $\mathbf{T}$ which permits the best fit to all respondent's preferences. We start with an approximation obtained from a principal components analysis of attribute ratings, and then improve it iteratively.

$\mathbf{C_i}$ = an (n x d) matrix giving the configuration of products for the i-th individual.

$$\mathbf{C_i} = \mathbf{X_i} \mathbf{T} \qquad (1)$$

The Vector Model

In the vector model each individual is thought of as having an ideal direction in the space, represented by a vector, and should prefer products according to their projections onto that vector.

Let $\mathbf{W_i}$ = a (d) vector of importance weights to be applied to columns of the individual's configuration to best predict that individual's preferences. Our basic individual preference equation is

$$\mathbf{C_i} \mathbf{W_i} - \mathbf{Y_i} = \mathbf{E_i} \qquad (2)$$

where $\mathbf{E_i}$ is a vector of errors of fit. Equation 2 says that the individual's configuration of products in space $\mathbf{C_i}$ is weighted by the elements of $\mathbf{W_i}$ to get a prediction of $\mathbf{Y_i}$.

Substituting from (1) into (2), we get

$$\mathbf{X_i \; T \; W_i = Y_i + E_i} \qquad (3)$$

If we knew **T**, we could solve for $\mathbf{W_i}$, using ordinary least squares. We start with an initial approximation of **T** and improve it in subsequent iterations. After estimating a $\mathbf{W_i}$ for each individual, we then combine information from all individuals to find a **T** that fits individuals better on average. By alternating between re-estimation of the **W**'s and **T**, we eventually find estimates for the **W**'s and **T** that best fit the data.

An initial estimate of **T** is obtained either from random numbers or from the principal components of the sum of all individuals' **X** matrices. In each iteration we solve for weights for each individual using ordinary least squares, minimizing the sum of squared errors in $\mathbf{E_i}$:

$$\mathbf{\hat{W}_i = (T' X_i' X T)^{-1} T' X_i' Y_i} \qquad (4)$$

We also record the r-square for each individual as a measure of how well the model fits that individual's preferences.

To produce an improved estimate of **T**, we could use the partial derivatives of the sum of squared errors with respect to **T**. For the i-th individual, those partials are given in equation (5):

$$\frac{\partial \mathbf{(E_i' E_i)}}{\partial \mathbf{T'}} = 2\mathbf{X_i' X_i T W_i W_i'} - 2\mathbf{X_i' Y_i W_i'} \qquad (5)$$

The partial derivatives of the total sum of squared errors involves summing equation (5) over all respondents Setting the partial derivatives of that sum to zero yields an expression for **T** which minimizes the sum of squared errors.

$$\sum_{i}^{N} \mathbf{X_i' X_i T W_i W_i' - X_i' Y_i W_i' = 0} \qquad (6)$$

However, this equation appears to be intractable. In each term of the sum, **T** is premultiplied by $\mathbf{X_i'X_i}$ and postmultiplied by $\mathbf{W_iW_i'}$, so it is not clear how to solve for **T**.

One possibility would be separately to sum the products $\mathbf{X_i'X_i}$, $\mathbf{W_iW_i'}$, and $\mathbf{X_i'Y_iW_i'}$ and then to premultiply the sum of $\mathbf{X_i'Y_iW_i'}$ by the inverse of the sum of $\mathbf{X_i'X_i}$ and postmultiply by the inverse of the sum of $\mathbf{W_iW_i'}$. We have tried that, but find that the sum of squared errors obtained with that approximation does not decrease monotonically from iteration to iteration.

However, we have had success with a slightly different procedure. In equation (3) the weights $\mathbf{W_i}$ are applied to the product $\mathbf{X_i \; T}$ to predict $\mathbf{Y_i}$, but it is also true that the weights $\mathbf{TW_i}$ are applied to the matrix of attribute ratings, $\mathbf{X_i}$, to predict $\mathbf{Y_i}$. We may estimate weights ($\mathbf{T\,W_i}$) by premultiplying equation (3) by $(\mathbf{X'_i\,X_i})^{-1}\,\mathbf{X'_i}$:

$$\mathbf{T\,W_i} = (\mathbf{X_i'X_i})^{-1}\,\mathbf{X_i'Y_i} + (\mathbf{X_i'X_i})^{-1}\,\mathbf{X_i'E_i} \qquad (7)$$

The first term on the right hand side of (7) is the estimate of regression weights that would be obtained in trying to predict $\mathbf{Y_i}$ from the individual's ***entire*** set of attribute ratings, $\mathbf{X_i}$. The second term on the right hand side is an error term that we hope is small. If it were zero, then equation (7) would state that the individual's weights for predicting his/her preferences in his/her entire attribute space would be expressible as a weighted combination of the columns of **T**. If that were true, then there would be no loss of predictive ability from using a subspace of small dimensionality common to all respondents. Our method of estimation improves **T** in each iteration so as to minimize the sum of squares of the last term in equation (7).

To show this more clearly we define

$$\mathbf{V_i} = (\mathbf{X_i'X_i})^{-1}\,\mathbf{X_i'Y_i} \qquad (8)$$

$$\mathbf{F_i} = (\mathbf{X_i'X_i})^{-1}\,\mathbf{X_i'E_i} \qquad (9)$$

$\mathbf{V_i}$ is the vector of weights that would best predict $\mathbf{Y_i}$ from $\mathbf{X_i}$. Consider a new regression computation for each individual, fitting $\mathbf{V_i}$ as a weighted combination of the columns of **T.** Equation (10) is obtained by substituting from (8) and (9) into (7). Since the errors are different, (**F** rather than **E**), the estimated coefficients will be different as well. Call these coefficients $\mathbf{U_i}$ (rather than $\mathbf{W_i}$).

$$\mathbf{T\,U_i} = \mathbf{V_i} + \mathbf{F_i} \qquad (10)$$

The OLS solution for $\mathbf{U_i}$ is:

$$\mathbf{U_i} = (\mathbf{T'T})^{-1}\,\mathbf{T'\,V_i} \qquad (11)$$

Assemble the $\mathbf{U_i}$ = and $\mathbf{V_i}$ vectors as columns of matrices **U** and **V**:

$$\mathbf{U} = (\mathbf{U_1}\,,\,\mathbf{U_2},\ldots\mathbf{U_N})$$

$$\mathbf{V} = (\mathbf{V_1}\,,\,\mathbf{V_2},\ldots\mathbf{V_N})$$

Then the OLS estimate of **T** that best fits equation (10) for all individuals is:

$$\hat{\mathbf{T}} = \mathbf{VU'(UU')^{-1}} \qquad (12)$$

Each iteration consists of two steps. During the first step, individual r-squares are computed using the regression indicated in equation (3) to measure the goodness of fit to each individual's preferences with the current estimate of **T**. A second r-square value is also computed for each individual using equation (11), to indicate how successfully the individual's preference structure is captured by the **T** matrix. The $\mathbf{U_i}$ and $\mathbf{V_i}$ vectors are also saved for each individual, as determined in equations (8) and (11).

In the second step, the $\mathbf{U_i}$ and $\mathbf{V_i}$ vectors are assembled and used as in equation (12) to re-estimate **T**.

Because the same **T** is used for all individuals, the aggregate product configuration is obtained by averaging the individual $\mathbf{C_i}$ matrices. Since the goodness of fit to individual data is not affected by a linear transformation of the columns of **T**, we also adjust **T** in each iteration so that the columns of the aggregate configuration are orthogonal and each has sum of squares of *n*.

As iterations progress, the sum of goodness-of-fit r-squares from the regression of equation (3) tends to increase, but it is not required to do so, and it often fluctuates in later iterations. However, the sum of "preference structure capture" r-squares from the regression of equation (10) increases monotonically, and iterations are terminated when the increases fall to less than a small positive value.

There are some numerical problems that must be overcome. Some respondents produce little or no useful information in their attribute ratings, and there may be more attributes than products. In either case the inverse of $\mathbf{X_i'X_i}$ will not exist. Therefore all matrices to be inverted first have a small positive amount added to their diagonal elements. This "ridge regression" trick remedies both problems.

The individual weights $\mathbf{W_i}$ are saved in a file, together with the goodness-of-fit r-square value for that individual. Coordinates of a map are provided, showing the average positions of products and attribute vectors in a space. The product configuration is just the average of individuals' configurations. The attribute vector positions are indicated by correlations between average product attribute ratings and product coordinates on each dimension.

The Ideal Point Model

For convenience, we repeat definitions stated above. For the i-th individual, let:

$\mathbf{X_i}$ = an (*n* x *p*) matrix of product ratings, with column sums of zero, and $\mathbf{X} = {}^{1}/_{N} \sum \mathbf{X_i}$

$\mathbf{Y_i}$ = an (*n*) vector of product preference values. APM provides constant-sum preference information obtained by having respondents divide 100 points among members of each of several product pairs. We construct $\mathbf{Y_i}$ by taking logits of those preference percentages, and awarding half the logit value to the winning product in that pair, and penalizing the losing product with half of the logit value. This results in $\mathbf{Y_i}$ values similar to conjoint utilities and that sum to zero for each individual.

T = a (*p* x *d*) matrix of weights used to transform attribute ratings into dimensional coordinates. **T** is common to all respondents. We want to find a **T** which permits the best fit to all respondent's preferences. We start with an approximation obtained from a principal components analysis of attribute ratings, and then improve it iteratively.

$\mathbf{C_i}$ = an (*n* x *d*) matrix giving the configuration of products for the i-th individual.

$$\mathbf{C_i} = \mathbf{X_i}\, \mathbf{T} \qquad (1)$$

We start by exponentiating and then percentaging each individual's $\mathbf{Y_i}$ values to get a set of positive values that sum to unity, similar to "shares of preference" in conjoint analysis.
Call this vector of preference information $\mathbf{R_i}$.

We assume each individual to have an "ideal point" in his/her perceptual space, defined as the row vector $\mathbf{P_i'}$. The squared distance from each product to that ideal point is obtained by subtracting $\mathbf{P_i'}$ from each row of $\mathbf{C_i}$ and then summing the squared differences. Call the vector of squared distances $\Delta_\mathbf{i}^\mathbf{2}$.

We expect that the distances should be small for products most preferred, and larger for products less preferred. We can express this desired relationship between preferences and distances in terms of the sum over products of the individual's preference weights times his/her squared distances, which we wish were small. Call this value for the ith individual $\theta_\mathbf{i}$.

$$\theta_\mathbf{i} = \mathbf{R_i'}\, \Delta_\mathbf{i}^\mathbf{2} \qquad (13)$$

If $\theta_\mathbf{i}$ is small, then for the ith individual the products with large preferences must have small distances from the ideal. Our goal is to find an ideal point for each respondent ($\mathbf{P_i'}$) and a matrix of weights common to all respondent (**T**) that minimize the sum of the θ values for all respondents:

$$\theta = \sum \theta_\mathbf{i} \qquad (14)$$

To estimate the ideal point for the ith respondent, we differentiate $\theta_\mathbf{i}$ with respect to $\mathbf{P_i'}$ and set the result to zero. Observing that the sum of the **R**'s is unity, we get the equation:

$$\mathbf{P_i'} = \mathbf{R_i'}\, \mathbf{C_i} \qquad (15)$$

The estimate of the individual's ideal point ($\mathbf{P_i'}$) which minimizes $\theta_\mathbf{i}$ is simply the weighted average of the rows of his/her matrix of perceived product locations, where the weights are the $\mathbf{R_i}$ values. Recall that the $\mathbf{R_i}$ values are positive and sum to unity. If the respondent has such extreme preference for one product that its value is unity and the rest are all zero, then the ideal point will be estimated to be coincident with that product's location. If the respondent is indifferent among products so that all $\mathbf{R_i}$ values are equal, then the ideal point will be estimated to be at the center of the perceptual space. No matter what the respondent's preferences, ideal points estimated this way will always lie within the convex hull of that respondent's perceived product locations.

Given an estimate of the ideal points for each individual, an improved estimate of **T** can be obtained as follows. Let $\mathbf{D_i}$ be a diagonal matrix whose diagonal elements are corresponding elements of $\mathbf{R_i}$. Then, differentiating θ with respect to **T** and setting the partial derivatives to zero and summing over respondents gives the equation:

$$\sum \mathbf{X_i' D_i X_i \; T} = \sum \mathbf{X_i' R_i P_i'} \qquad (16)$$

It would seem that one way to estimate T would be to cumulate the two sums

$$\mathbf{A} = \sum \mathbf{X_i' D_i X_i} \qquad (17)$$

and

$$\mathbf{B} = \sum \mathbf{X_i' R_i P_i'} \qquad (18)$$

so that

$$\mathbf{A\,T = B}$$

and then simply estimate **T** as $\mathbf{A}^{-1}\mathbf{B}$.

However, the problem with this approach is that θ is minimized trivially by a **T** of zero, and an iterative process which estimates **T** in this way eventually converges to a **T** of zero. To avoid that, it is necessary to impose constraints on **T**. We choose to make columns of the overall configuration **XT** orthogonal, and for each column to have sum of squares equal to the number of products, ***n***. (Recall that **X** is the average of the $\mathbf{X_i}$ matrices.)

This is done with a symmetric matrix of Lagrange multipliers, following Schonemann (1965). We differentiate the sum of $\theta + \phi$ with respect to **T**, where

$$\phi = \text{trace}(\mathbf{S\,T'\,X'\,X\,T}) \qquad (19)$$

with **S** an unknown symmetric matrix.

Setting the sum of partial derivatives to zero yields:

$$\mathbf{AT = B + X'X\,T\,S} \qquad (20)$$

Premultiplying by **T'** and recalling that **T'X'X T** is constrained to equal the identity matrix, we get:

$$\mathbf{S = T'(AT - B)} \qquad (21)$$

Premultiplying (20) by the inverse of **A**,

$$\mathbf{T = A^{-1}B + A^{-1} X'X\ T\ S} \qquad (22)$$

Equation (20) cannot be solved explicitly for **T**, but does submit to an iterative solution consisting of the following steps:

1) Use the value of **T** from the previous iteration or some initial value, to compute **S** as in equation (21). Early estimates of **S** will not be symmetric, so force symmetry by averaging corresponding elements above and below the diagonal.

2) Obtain the product $\mathbf{A^{-1} X'X\ T\ S}$, using current estimates of **T** and **S**.

3) Determine a scalar α by which to multiply the product obtained in step 2) so that $\mathbf{XA^{-1}B} + \alpha\mathbf{XA^{-1} X'X\ T\ S}$ has sum of squares equal to ***n* d*** .

4) Use the sum of terms: $\mathbf{A^{-1}B} + \alpha\mathbf{A^{-1} X'X\ T\ S}$ as an interim estimate of **T** to obtain an interim (not necessarily orthogonal) estimate of the configuration of products in space, **X T**.

5) Find the matrix with orthogonal columns and column sums of squares equal to ***n*** which is closest in the least squares sense to **X T**, using the procedure of Johnson (1966), as well as the right-hand transformation matrix that performs that orthogonalization.

6) Finally, postmultiply the interim estimate of **T** from step 4) by the transformation matrix determined in step 5) to get the estimate of **T** for the current iteration.

Repeat steps 1-6 until estimates of **T** stabilize.

References

Carroll, J. D. and Chang, J. J. (1970), “Analysis of Individual Differences in Multidimensional Scaling Via an N-way Generalization of Eckart-Young Decomposition,” *Psychometrika*, 35, 283-319.

Coombs, Clyde H. (1964) *A Theory of Data*. New York: Wiley.

DeSarbo, Wayne S. & J. Douglas Carroll (1985), “Three-Way Metric Unfolding Via Alternating Weighted Least Squares,” *Psychometrika*, 50, September, 275-300.

Elrod, Terry (1997),”Obtaining Product-Market Maps from Preference Data,” Sawtooth Software Conference Proceedings, 273-288.

Johnson, Richard M. (1966) “The Minimal Transformation to Orthonormality,” Psychometrika, 31, March, 61-66.

Schonemann, Peter H. (1965) “On the Formal Differentiation of Traces and Determinants,” Research Memorandum No. 27, The Psychometric Laboratory, University of North Carolina.

Tucker, Ledyard R and Samuel Messick (1964) "An Individual Differences Model of Multi-Dimensional Scaling," *Psychometrika*, 333-367.

Wittenschlaeger, Thomas A. and J. A. Fiedler (1997), “Current Practices in Perceptual Mapping,” Sawtooth Software Conference Proceedings, 259-270.