Background: Among market research practitioners, there has recently been interest in scaling product or category characteristics (such as possible benefits) based on responses indicating the items with the greatest and least magnitude among a subset of possible items. The basis for this choice could be liking, purchase interest, importance or even a sensory characteristic such as sweetness. For example, considering the characteristics of plug-in air care products, items to consider might include “low cost,” “does not fade over time,” “has a use-up cue,” and “has a fresh scent.” A respondent may be instructed to choose the attribute of most importance and the one of least importance in making a purchase decision. From a large collection of items, subsets of equal size are chosen and presented in a balanced design. The typical number of items used per respondent is four. The analytic task is to develop a scale on which each attribute can be placed so that scale values for all of the items from most to least can be obtained.

Scenario: Your company markets a variety of air care products including instant action air fresheners and plug-in products that last for as long as 60 days. In order to improve the new product development process by identifying key features of interest to your consumers, you seek to prioritize ten possible features of plug-in air fresheners. These features are presented in Table 1. Among a target group of interest, you obtain responses to questions as shown in Figure 1 concerning the importance of the features listed. Sets of four features are presented in a balanced design to each of 200 consumers and each consumer evaluates fifteen sets of four features.

Table 1. Ten possible benefit features of plug-in air fresheners.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low cost</td>
<td>Comes in a wide variety of possible scents</td>
</tr>
<tr>
<td>Has a use-up cue</td>
<td>Has an attractive appearance</td>
</tr>
<tr>
<td>Does not fade over time</td>
<td>Fills the entire room with fragrance</td>
</tr>
<tr>
<td>Lasts at least 60 days</td>
<td>Made by an environmentally sensitive company</td>
</tr>
<tr>
<td>Has a fresh scent</td>
<td>Is the best selling brand</td>
</tr>
</tbody>
</table>

Of the following four features of plug-in products, which is the most and which is the least important to you when making a purchase?

**Figure 1.** Typical responses in a first-last choice task.
A first choice Thurstonian model was published by Ennis and O’Mahony\(^5\) to account for sequential effects in product tests. Recently, a Thurstonian model for first-last choice has been generalized to include any number of items in the subset. The model is different from a commonly used alternative based on the logit\(^6\), as the new model assumes that the items are distributed normally on the unidimensional utility or hedonic continuum. A great advantage of this approach is that it allows the prediction of results from other psychophysical tasks, including paired preference, rating and ranking tasks, using the already developed Thurstonian family of related models, and thus provides a basis for validation and comparison. This model also allows the determination of the variances and covariances of the parameter estimates at the aggregate level so that statistical comparisons can be made.

**Modeling First-Last Data:** In order to approach the development of a first-last Thurstonian model, we assume that the perceptual representations of the items can be modeled as univariate normal distributions. For simplicity, we assume these distributions have unit standard deviations. The difference between the means of the distributions are called \(\delta\) values and the units of the \(\delta\) values are perceptual standard deviations. These are the same assumptions that we used to model the results of difference tests, ratings and rankings in previous papers and technical reports\(^7,8,9\), and these analyses can be conducted in *IFPrograms*\(^\text{TM}\). In addition, we assume that a respondent considers the features listed in Figure 1 and uses the same perceptual values when making the first and last decisions. This last assumption is called dependent sampling. Independent sampling would occur if the respondent recorded their first choice and then considered a separate random sample to decide their last choice. We assume dependent sampling, although independent sampling may occur in other settings, such as when a respondent is re-tasting or re-smelling food items between their choice of the first and last items. A complicated issue that arises with independent sampling is that the same item may appear to be both first and last. Although here we assume dependent samples, this issue has been discussed previously for Richardson’s method of triads\(^4\). The first-last choice model connects \(\delta\) values with probabilities of choosing particular items first or last.

**Results from Modeling the Plug-In Data:** Figure 2 shows the \(d'\) values (estimates of \(\delta\) values), or scale means, for the ten plug-in features. It can be seen that for this particular group, you conclude that the benefit “does not fade over time” is the most important item and “made by an environmentally sensitive company” is the least important. All of the benefits are scaled relative to the last one which is set at zero. Figure 2 also shows standard error bars for the scale means. Since the last benefit is assumed to have a \(\delta\) value of zero, it has no error bar. Using these results, we can now predict the outcomes of other methods for which Thurstonian models have been developed.

"![Bar chart showing scale means (d' values) and their standard errors for ten features.](Figure 2)"

**Conclusion:** A Thurstonian model with dependent sampling has been developed and, in the scenario, was applied to first-last data for plug-in fragrance products. The resulting \(\delta\) value estimates were used to prioritize the benefits for this category. The psychophysical task involving first and last choices has been referred to historically as MaxDiff and Best-Worst testing. Neither of these descriptors is appropriate as the method involves more than finding the two items that exhibit maximum difference and is not limited to hedonic or utility continua.

**References:**