Bridging Designs for Conjoint Analysis: The Issue of Attribute Importance

Pierre Francois

Katholieke Universiteit Leuven, Belgium

and

Douglas L. MacLachlan

University of Washington, Seattle, USA

September 1997

This is a draft. Please do not quote without the authors' permission. Contact the authors at pierre.francois@econ.kuleuven.ac.be or macl@u.washington.edu.

BRIDG97.doc

Bridging Designs for Conjoint Analysis: The Issue of Attribute Importance

(Key words: Market Research; Conjoint Analysis)

Abstract

Conjoint analysis studies involving many attributes and attribute levels often occur in practice. Because such studies can cause respondent fatigue and lack of cooperation, it is important to design data collection tasks that reduce those problems. Bridging designs, incorporating two or more task subsets with overlapping attributes, can presumably lower task difficulty in such cases. In this paper, we present results of a study examining the effects on predictive validity of bridging design decisions involving important or unimportant attributes as links (bridges) between card-sort tasks and the degree of balance and consistency in estimated attribute importance across tasks. We also propose a new symmetric procedure, Symbridge, to scale the bridged conjoint solutions.

Introduction

Conjoint analysis has emerged in recent years as being among the most important and useful methods in marketing research (Cattin & Wittink, 1982; Wittink & Cattin, 1989; Green & Srinivasan, 1990). Its value to marketing decision makers stems from its ability to provide realistic customer evaluations of the attributes of products, services, and so forth, as well as to allow the simulation of choice shares in hypothetical competitive scenarios. Academics have explored the usefulness and applicability of conjoint analysis and sought understanding of its limitations. For example, Reibstein, Bateson, and Boulding (1988) assessed the reliability of conjoint analysis under a variety of conditions, finding, among other things, that the type of data collection procedure has an impact on the reliability of the results.

Among the limitations of the method is the typically difficult task respondents face in providing data suitable for conjoint analysis. Frankly speaking, the task is usually boring, complex, and frustrating for even the most highly motivated respondents. Although different data collection methods used in studies (e.g., full-profile, trade-off, and paired-comparisons, with rating or ranking of stimuli) have different degrees of task difficulty, it is apparent that the task is never easy. Although it is probable that hybrid designs (Green 1984) and adaptive methods employing interactive computers (Johnson 1987) help to reduce the task difficulty and boredom factors, it is likely that full-profile designs will continue to be used for some time in marketing applications of conjoint analysis. Even with fractional factorial designs, full profile data collection procedures involve comparing a number of stimuli with multiple attributes and levels. The respondent's decision problem is exacerbated when the number of attributes and levels becomes large. Various studies (e.g., Miller 1956 and Wright 1975) have shown that people have difficulty processing more than a relatively few pieces of information at one time. Since the conjoint task requests people to keep in mind multiple attributes and levels simultaneously, their task becomes much more difficult as the number of attributes/levels (and hence tradeoffs) increases.

Possible strategies to reduce the task difficulty in conjoint studies.

Looking at the literature about Conjoint Analysis published in the nineties, the technique is still very popular, and many topics are still being addressed today. In this section we try to give an overview of ways to handle the number of attributes problem. Some of them have been specifically been designed for this purpose, while others were proposed with another objective in mind, but with the potential side benefit of reducing the information overload for the respondent (typically the number of stimuli to be evaluated).

1. Ad hoc reduction of the number of attributes and/or levels.

The easiest and most subjective approach, and the one probably used most, is to prune the set of attributes and levels, until a reasonable number of stimuli (cards) has been obtained. It is still open to debate what this optimal number could be, but in our own experience, conjoint designs requiring complete rank ordering of stimuli should not contain more than 12 to 16 cards.

Obviously this approach entails the risk of eliminating relevant attributes, although a good pretest can reveal which attributes are used more often or are more important to representative users (depending on how one measures ‘importance’).

At the same time, one can try to reduce the number of levels for some attributes.

This may not be possible in some cases (think about brand names), and will have the undesirable effect of reducing the importance of attributes with a smaller number of levels (Wittink et al (1990)).

Another approach would be to reduce the number of parameters to be estimated, by imposing certain functional forms (vector, ideal point) on the partial utility functions. In this way one would gain degrees-of-freedom, and in principle be able to reduce the number of stimuli accordingly. This approach to the problem has not been extensively researched, and known references studied the impact on predictive validity instead (Pekelman and Sen 1979). Although this is an empirical matter, one could argue that imposing specific forms could lead to misspecifications, and this would explain why the generalized form is so popular. Also, it would only apply to metrically scaled attributes (e.g. price).

2. Combination of full-profile and self-explicated data.

These approaches do not try to reduce the number of attributes, but reduce the number of judgments by respondents, compared to full-profile methods. Among the more popular methods, one can mention ACA (Johnson (1987)) and hybrid conjoint models (Green (1984)).

An extension of ACA combines unacceptable levels with the basic ACA approach (Mehta et al 1992)

A more recent approach and alternative to ACA is CCA (Srinivasan and Park 1997).

Self-explicated methods have proven to be very popular with practitioners, probably due to their appealing data collection procedure, ability of handling large problems, and relative mathematical simplicity. Whether they perform as well or better than traditional full-profile methods, has been the subject of some very involved debates, and the issue is not resolved yet (Carroll and Green, 1995).

The hierarchical integration method (Oppewal et al 1994) also belongs to this class, and tries to define superattributes, which in turn can be conjoint modelled as a function of attributes only relevant for a particular superattribute. The overall preference function would then only contain the superattributes, which are smaller in number than all the attributes that compose them.

3. Adaptation of the design matrix.

These strategies aim to keep the number of attributes, and reduce the number of judgments by respondent, by trying to adapt the full fledged orthogonal design, typically by incorporating available information (Huber and Zwerina 1996).

4. Optimal segmentation methods.

This family of techniques is the object of much research lately. From our point of view, these methods solve the problem by abandoning the idea of estimating individual utility functions (and reducing the workload of the respondent), but retaining information about the individual respondents as sources of heterogeneity. Methods include latent class models (Kamakura et al (1994)), optimal scaling (Hagerty (1985), Green et al (1993), Desarbo et al (1992)), Hierarchical Bayes methods (Allenby et al (1995),Lenk et al (1996)).

Some of these methods have however not been positioned to reduce the respondents'workload, but more as alternative segmentation models (compared to clustering of individual part-worths).

This reduction in the number of stimuli comes however at the price of having to interview more respondents.

5. Bridging

This variant of conjoint, which is the subject of our research, tries to cope with the overload problem, by splitting the set of attributes in two, and building a scaling ‘bridge’ between the two sets of derived utilities.

6. Constrained estimation procedures.

This extension of conjoint has also received some attention lately (Allenby and Ginter 1995),Green and Krieger (1995). These researchers do not mention reduction of number of stimuli as an objective but rather the efficiency of the estimates or quality of prediction. However, imposing constraints might lead to savings in degrees of freedom.

Bridging Designs

A way to reduce respondent difficulty handling full-profile designs having large numbers of attributes/levels is to decompose the data collection into simpler subtasks that are later linked analytically. Hybrid designs are one type of such design, where self-explicated attribute weights are combined analytically with data coming from card-sort (or other tradeoff) tasks. Another approach is to divide the card-sort (or other stimulus comparison) task into multiple designs containing two or more subsets of attributes with one or more attributes common to both subdesigns. In such case, the linking or "bridging" attributes are used to scale the partworths from the two subdesigns into one overall set of partworths. It is presumed that people will have an easier time working with these subdesign tasks and therefore be more cooperative and provide more thoughtful responses. The general class of these latter designs are called bridging designs. Although the total number of stimuli needed to be compared in bridging designs may not be reduced, the task is made simpler for respondents by asking them to evaluate stimuli on fewer attributes within each subdesign. (An extreme form of bridging design is the pairwise tradeoff grid approach, where only two attributes are compared at a time.)

Suppose one wished to develop a bridging design. What decisions should be made? What issues should be considered? The research problem in the current study was to investigate the conditions involved in bridging designs that might lead to better or worse estimates of attribute level partworths. Although there is a commercially available software package for scaling bridging designs (Bretton-Clark's Bridger, 1988), we were unable to uncover any published studies that examined these conditions or offered advice about how the designs should be developed.

The issues that emerge immediately are:

(1) How many bridging attributes should be used?

(2) Which attributes should be used as bridges?

a. How "important" should they be relative to other attributes?

b. How balanced should the designs be in terms of attribute importance?

(3) What scaling procedure should be used to combine solutions from the subdesigns into overall partworth estimates?

In the present study we decided to investigate issues (2) and (3). Logically, the more attributes employed as bridges, the better will be the results ("better" will be described and operationally defined later). For the present study, we decided to focus on two bridging attributes, because it creates more stimulating issues than with one bridging attribute and not much more understanding is gained from using more attributes as bridges. The more interesting practical problem is deciding which attributes should be employed as bridges and what are the consequences of the choice of bridging attributes. Our research hypotheses investigate this issue. A secondary issue to be addressed in this study is how the subdesign solutions should be analytically combined.

Research Hypotheses

In the following, we state our hypotheses in the way we anticipate them to be verified. Obviously, the null hypotheses to be tested in each case will be that of no effect.

Although it will not be possible, in advance of a conjoint study, to know with assurance the level of importance of all attributes for all respondents, there is often some a priori information that could be used to develop the designs. Because the bridging attributes are key to the linkage of subdesign partworth estimation into a single overall solution, it is reasonable to expect that their relative importance will have an impact on the quality of the estimates. This comes about in a couple of ways. First, the more important the attributes (i.e., the larger the range of its partworths), the more leverage is given to the estimation. This is analogous to regression, where the wider the range of the independent variable, the more stable is the assessment of slope and intercept (and the lower is the standard error of estimation). Hence, our first hypothesis is:

H1: The higher the relative importance of the bridging attributes, the better is the overall partworth solution.

Secondly, since the nonbridging attributes can be distributed in different ways across the subdesigns, one would expect solutions to be better to the extent that the subdesigns are "balanced" in terms of their attributes' relative importances. If one subdesign were to contain most of the important attributes, one would expect that design to have partworth estimates of higher quality than the other subdesign. This should lead to lower overall estimate quality than if the designs are well-balanced in terms of attribute importance. Thus the second hypothesis is:

H2: The more balanced the designs are with respect to attribute importance, the better is the overall partworth solution.

We will define two kinds of attribute imbalance to test this hypothesis.

Thirdly, with respect to the resulting estimates of bridging attribute importances, since they are separately estimated in each design, it is possible to question the effect on overall solution quality of the degree of inconsistency in their measured relative importance. If an attribute is estimated to have low importance in one subdesign and high importance in the other design relative to the other bridging attribute, this should affect the ability of any linking procedure to provide good estimates of overall partworth across both designs. If bridging attributes are measured to be consistent in importance, therefore, we hypothesize that the solution will be better. Thus,

H3: The more consistent the bridging attribute relative importance between subdesigns, the better is the overall partworth solution.

Finally, it seems useful to investigate whether overall attribute importance spread over the two subdesigns has an impact on the quality of the partworth solution. That is, since total attribute importance indicates the extent to which a respondent holds strong opinions regarding the extent to which different levels of the various attributes have different utility, we hypothesize that respondents with stronger-held opinions will provide higher quality data yielding better solutions. This last hypothesis does not directly depend on which particular attributes are used as bridges (i.e., it is more a hypothesis regarding respondent quality than design quality).

H4: The higher the total attribute importance over subdesigns, the better is the overall partworth solution.

Although these hypotheses could conceivably be addressed with mathematical analysis or simulation, practical application of bridging designs involves dealing with people, so we decided to examine them with real subjects and realistic situations. We developed the following study design to test the hypotheses empirically.

Study Design

The study employed a household durable product with a fairly high number of important attributes: vacuum cleaners. A pretest was conducted wherein 240 housewife subjects were interviewed on three separate occasions using different procedures to elicit conjoint-analysis based importance scores for a subset of the attributes used in the current study. For the present study, some additional attributes were employed, for which we subjectively assessed relative importance. The attributes and ranges are shown in Table 1.

[Table 1]

In order to test the hypotheses, we needed to assure ourselves that sufficient ranges of bridging attribute importance would exist in our sample. In this case, therefore, we had a priori indications of the relative importance of the attributes (a combination of pretest results and researcher subjective judgments) that could be used in generating bridging designs having (on average) different levels of importance in the bridging attributes. The attributes either had three or two levels. Three conditions were specified: (1) bridging attributes both important, (2) one important and one unimportant bridging attribute, and (3) both unimportant bridging attributes. We carefully selected the remaining attributes in each design to try to achieve balance in overall importance across the two subdesigns. Moreover, we set up the subdesigns so that each half contained the same number of attributes with two and three attributes. This also meant we could use the same fractional-factorial design in preparing the stimuli for each half and each subdesign would contain the same number of stimuli. The resulting designs contained attributes as indicated in Table 2. Although we fully realized that individual respondents might assign quite different relative importances to the various attributes in each design, it was felt useful to employ the a priori information in creating the study to assure a wide range of conditions across respondents. The testing of our hypotheses, however, does not depend on the a priori conditions being confirmed on average.

[Table 2]

Main-effects only stimulus configuration designs were obtained for the subdesigns, each containing five attributes (using Bretton-Clark's Conjoint Designer, 1987). The fractional-factorial designs each resulted in a set of 16 stimuli. Attributes and levels were assigned to the subdesigns in such a way that no stimulus was completely dominated by (or completely dominated) the others; i.e., no stimulus had the "best" or "worst" values on all attributes.

A sample of 120 housewives in Belgium was selected and randomly assigned to the three attribute bridge conditions. Student interviewers were employed and the data collected as part of their thesis projects. Before doing the bridging task, the respondents were asked to rank order a set of six holdout stimuli containing realistic combinations of attribute levels across all eight attributes (these were selected to be sufficiently different to provide an adequate range of responses). Before or after the bridging task (randomly assigned to half the respondents), respondents were asked to complete a complete full-profile task with all eight attributes. This latter task required sorting 27 cards, a minimum main-effects design. It is important to note that the overall task for the two halves of the bridging design involved more stimuli than the full design. It is apparent on its face, though, that the latter task requires more effort and should be more difficult for the respondent than the former. Each subdesign task had 16 cards with 5 attributes (requiring subjects to process simultaneously 80 bits of information), whereas the full design had 27 cards with 8 attributes each (requiring simultaneous processing of 216 bits of information). As a final task, respondents were asked to indicate the importance of the attributes in a self-explicated fashion (6-point scale).

Measures

Assume that we have five attributes in each of two subdesigns with two overlapping attributes (bridging attributes). For each subdesign we may estimate partworths and hence the range of partworths for each attribute. These latter ranges we define to be the importances I_{ij} for design i (i=1,2) and attribute j. Next, we define the relative importances RI_{ij} corresponding to the I_{ij} divided by the sum of the I_{ij} within each subdesign.

For the example of the current study, we have:

Design 1	Design 2	Overall	
RI_{15}		RI_5	
RI_{14}		RI_4	
RI_{13}		RI_3	
RI_{12}	RI_{22}	RI_2	] Bridging
RI_{11}	RI_{21}	RI_1	] attributes
	RI_{26}	RI_6	
	RI_{27}	RI_7	
	RI_{28}	RI_8	

where RI_j is the relative importance of attribute j in the final solution (to be discussed later).

Measures of Solution Quality

Two measures of the general quality of the bridged solutions are used in this study, solution inconsistency (a measure of the squared differences between original relative importances and the rescaled relative importances) and correlation (between original scaled relative importances and rescaled relative importances). Thus:

<u>Solution inconsistency</u>: This measure is not used to test any hypotheses. It is a measure of solution internal inconsistency, i.e., a measure of inconsistency between the original bridging estimates and the "bridged" solution.

$$\text{Solution inconsistency} = \text{sum}(\text{diff}_j^2) \tag{1}$$

where diff_j is $(RI_{1j} - RI_j)$ or $(RI_{2j} - RI_j)$.

The measure is summed over 10 differences in the above example.

Correlation: This is the simple Pearson product-moment correlation between the original estimates of relative importance and the rescaled relative importances.

$$\text{Correlation} = \text{Corr}(RI_{ij}, RI_j) \tag{2}$$

where the replicates are the 10 relative importances derived in each experimental condition.

Both measures are helpful for determining how well the particular bridging algorithm works in comparison with other bridging algorithms. Obviously, the lower the measured value for solution inconsistency and the higher the value for correlation, the higher is the solution quality.

Independent variables

A total of five independent variables were developed to test the research hypotheses.

Importance: The first independent variable, used to test H1, is the importance of the bridging variables. This is simply measured as the sum of the relative importances of the bridging variables across both subdesigns. Thus,

$$\text{Importance} = RI_{11} + RI_{12} + RI_{21} + RI_{22} \tag{3}$$

Imbalance: Two imbalance measures are developed in this study. The first is a measure of imbalance across designs where total bridge-attribute importance in one design is compared with the similar value in the other design. Thus,

$$\text{Imbalance}_1 = \text{abs}((RI_{11} + RI_{12})\text{-}(RI_{21} + RI_{22})) \tag{4}$$

This is really a measure of the extent to which the non-bridging attributes are balanced in total attribute importance across subdesigns. That is, if the left subdesign has higher attribute relative importance estimated for bridging attributes than the right, it would imply that the left design had lower importance for the non-bridging attributes than the right subdesign.

The second measure is imbalance between attributes over the two designs. Thus,

$$\text{Imbalance}_2 = \text{abs}((RI_{11} + RI_{21})\text{-}(RI_{12} + RI_{22})) \tag{5}$$

Imbalance measure 2 focuses on the relationship between the two bridging attributes, as opposed to the non-bridging attributes. A solution can be balanced by measure 1 but imbalanced by measure 2. An example would be where, in each subdesign, the first bridging attribute was twice as important as the second attribute. We do not expect this to have as large an effect on predictive validity as measure 1, since so long as the sum of bridging attribute importances in a given subdesign is sizeable, the scaling of the overall solution should be stable.

Inconsistency: A measure is defined to indicate the degree to which there is inconsistency in the magnitude of the relative importances estimated for each bridging attribute across designs. The larger this measure, the more disparity exists in the relative size of the importances for any bridging attribute. Thus,

$$\text{Inconsistency} = \text{abs}(RI_{11} * RI_{22} - RI_{12} * RI_{21}) \qquad (6)$$

The measure is used to test H3.

Total Importance: In order to assess the strength of opinion of the respondent over all attributes and levels, we create a measure that is simply the sum of the ranges of all attributes in both subdesigns. For the above example, this sum is over 10 values. This measure is used to test H4, and we expect it to have a positive influence on predictive validity.

Dependent variables

The quality of the partworth solutions was investigated using standard measures of predictive validity, comparing predicted utility of hold-out stimuli with the reported ranking of those stimuli.

<u>Correlation</u>. The average Spearman rank correlation between predicted and actual ranks of hold-out cards.

<u>Percentage of first-choice hits</u>. Two cases: (1) "hit" only if prediction of card was first choice, and (2) "hit" if prediction for card was first or second choice.

The above dependent variables are progressively less stringent measures of solution quality.

Analysis and Results

The partworth solutions for each subdesign and the full design were obtained for each respondent subject. These were unconstrained OLS solutions using dummy-variable regression (i.e., using Bretton-Clark's <u>Conjoint Analyzer</u>, 1987). In order to combine the solutions for each of the subdesigns into a single vector of partworths for all attributes, a number of options were available. Since each subdesign task was performed independently, each provides estimates of the bridging attribute importances (ranges of partworth levels). Logically, the solution algorithm should depend on the ratio of attribute importances for the bridging attributes across subdesigns. However, the solution will differ depending on which ratio (1st over 2nd or 2nd over 1st) is used to rescale which subdesign's partworths (a problem of nonsymmetry) and how the importances for the two bridging attributes are combined.

To solve the second problem, we experimented with two types of ratio for the bridge algorithm. The first was a sum of ratios of bridging attribute ranges, i.e., (R_{11}/R_{21} +

R_{12}/R_{22}), where R_{ij} is the range of partworths of bridging attribute j in subdesign i. However, this tended to be unstable, because some of the ranges were near zero. A more stable alternative was the following:

$$B = (R_{11}+R_{12})/(R_{21}+R_{22}) \tag{7}$$

To solve the nonsymmetry problem, we applied B (a scalar value) to rescale the 2nd subdesign partworths and B^{-1} to rescale the 1st subdesign partworths, then added the resulting partworths. Thus, for our study, the algorithm was as follows, where u_{ij} represents a vector of partworths of all levels of attribute j in design i:

Stage 1	Stage 2	Final Partworths
u_{15}	$B^{-1}{*}u_{15}$	$u_{15}+B^{-1}{*}u_{15}$
u_{14}	$B^{-1}{*}u_{14}$	$u_{14}+B^{-1}{*}u_{14}$
u_{13}	$B^{-1}{*}u_{13}$	$u_{13}+B^{-1}{*}u_{13}$
$B{*}u_{22}$	$B^{-1}{*}u_{12}$	$B{*}u_{22}+B^{-1}{*}u_{12}$
$B{*}u_{21}$	$B^{-1}{*}u_{11}$	$B{*}u_{21}+B^{-1}{*}u_{11}$
$B{*}u_{26}$	u_{26}	$B{*}u_{26}+u_{26}$
$B{*}u_{27}$	u_{27}	$B{*}u_{27}+u_{27}$
$B{*}u_{28}$	u_{28}	$B{*}u_{28}+u_{27}$

We call our method Symbridge, to denote symmetrical bridging. Table 3 shows the measures of solution quality (defined above) for Symbridge in contrast with the solutions provided by Bridger. Symbridge provided an average solution inconsistency measure of 0.120 and correlation of 0.90 for the subjects. By contrast, the commercial package Bridger yielded an average solution inconsistency measure of 0.077 and correlation of 0.877 for the same input data. Thus the two measures show mixed performance for the two bridging methods. Given the asymmetric nature of Bridger solutions, we would have expected it to perform worse relative to Symbridge. Our understanding, from informal conversations with other conjoint-experienced researchers, is that Bridger uses a least-squares procedure for scaling the two solutions, which would tend to minimize the criterion we call "solution inconsistency". Symbridge provides somewhat more stable results (i.e., the distributions of solution quality measures over subjects are less skewed; e.g., skewness coefficient for solution inconsistency was 2.2 for Bridger, but only 0.6 for Symbridge).

[Table 3]

Table 3 also shows the average results for predictive validity for Symbridge and Bridger versus the full design on the various criteria. It is not surprising that the full design outperforms the other methods, on average, since the holdout sample used stimuli that included all attributes. Symbridge and Bridger performed approximately equally on all predictive validity criteria (certainly the differences are not statistically significant).

Table 4 shows the average values of attribute importances estimated from the sample by the different methods, augmented by the self-explicated importances and our a priori assessments of the attribute importances. Notice that our a priori indications for attribute importance were not confirmed by the analysis of this sample of respondents. Symbridge and Bridger yielded a high degree of correlation in their results, although one can observe some differences. Bridger average importances have a higher correlation with full-design average importances than does Symbridge. Symbridge average importances have a higher correlation with self-explicated average importances than do either Bridger or full design average importances.

[Table 4]

Table 5 shows the basic descriptive statistics and intercorrelations for the independent variables to be used to test hypotheses H1-H4. Although there is an indication of skewness for all of the independent variables except for "total importance", there is sufficient variability in all variables to provide some explanatory power. There is fairly high correlation between Importance and Imbalance_2, which is understandable given the variable definitions being based on the same subcomponents. Also, there is a moderate correlation between Importance and Inconsistency. Otherwise, the variables do not exhibit very high intercorrelation.

[Table 5]

Table 6 examines hypotheses H1-H4 using Spearman correlation of predicted versus actual holdout stimuli ranks as the predictive validity criterion. It is disappointing to observe that, except for H4, none of the hypotheses was supported (i.e., the null hypotheses could not be rejected) in all cases. This turns out the same when using "hit ratio" as the "goodness" criterion. H4 was not dependent on the type of design and could be tested as well for the full design. Interestingly, we found support for H4 in the bridged designs, but not the full design. It appears that using a bridge design causes total attribute importance to become more critical to predictive validity than when using a full design.

[Table 6]

At this point we began investigating the data for subjects with particularly weak predictive validity that might be skewing the results. However, even comparing highest and lowest quartile respondents on independent variables involved in H1-H3 turned up no significant differences for predictive validity, except for the peculiar result for both Symbridge and Bridger that high quartile respondents on the measure Imbalance_1 yielded significantly higher predictive validity than low quartile respondents (two-tailed p-values were .06 and .004, respectively).

Discussion

We began this investigation favorably disposed to bridging as a means of reducing respondent task difficulty, hopefully yielding higher quality results. Our conclusion is much less optimistic for bridging. Nevertheless, this is the way science progresses.

Despite the disappointing results, we feel we have addressed some worthwhile issues and have described ways of examining them. The method we developed for scaling bridging design solutions, Symbridge, while seemingly better on logical grounds, appears not to outperform the one existing commercial method, Bridger (although one should be careful in generalizing from a single study with only one product). Symbridge has face validity, since it does not depend on the order in which the bridging designs are entered into the algorithm. Although somewhat mixed as we saw in Table 3, the Symbridge did tend to provide more stable (less skewed) distributions of solution quality than did Bridger in this instance.

Our various hypotheses are reasonable on their face, as well. A finding consistent with the null hypothesis is not the same as proving it. Although the number of subjects in the study would seem to have provided statistical power sufficient to detect even small effects, the particular study conditions may have worked against rejecting the null hypotheses. For example, having two bridging attributes reduces the probability that two unimportant attributes will be used as bridges (and presence of one important attribute may be sufficient for reasonable solution quality and predictive validity). We continue to be convinced that choice of bridging attributes is a relevant issue for researchers planning to use bridging designs. To the extent it is possible, researchers using bridge designs should probably try to make the bridging attributes the most important ones for the majority of their respondents. Also, they should try to balance the subdesigns in terms of their overall attribute importance. On the other hand, we have evidence in this study of the inability of researchers to really judge the importance of attributes on an a priori basis. This would argue for creating designs "on the fly", so to speak, that would be potentially different for each respondent. That would be almost like using the adaptive methods (e.g., ACA, Johnson, 1987).

Although our study does not provide evidence for the proposition, it may be that any bridging design will be troublesome if any of the subdesigns omits an attribute that is extremely important to the choice decision. An example is the price attribute, which tends to be important in many conjoint analysis studies. There is thus an argument for using price as a bridging attribute in any bridging study.

One could use our results to support the idea that full designs should be used instead of bridging designs (i.e., that bridging should never be used). The full design outperformed bridging in every instance. More stimuli had to be compared in total in the bridging tasks, and actually took about the same time overall (the full-design task averaged 11.1 minutes, while the bridging task averaged 11.7 minutes in total). This is despite the "easier" tasks of sorting the stimuli in the subdesign conditions. (We note parenthetically that there was no significant correlation between predictive validity and time for the task for any of the solution methods, suggesting that more time taken has little to do with quality of "effort".) It is possible that subjects are able to simplify the full-design tasks by simply ignoring the levels of the unimportant attributes. And the full design stimuli have greater realism (the more complete description could provide better context for respondent decisions, even if the unimportant attributes are ignored).

Suggestions for Future Research

Given the inability to reject the null hypotheses corresponding to H1-H3, we need to replicate this study with different products, different situations, and possibly more motivated subjects, either paying them or developing some other procedures for getting them to take the task seriously. We should also examine the impact of using designs with larger numbers of attributes and attribute levels. Additionally, one could consider expanding the number of bridging attributes and/or the numbers of bridging subdesigns.

A different approach would be to attempt a solution to the problem with mathematical analysis. The number of variables and the complexity of the situation suggests that this approach would be very difficult.

Finally, the hypotheses might be investigated with Monte Carlo simulation. Of course, in such a case, the researcher must be careful to ground the various assumptions in reality to avoid merely playing games.

Additionally, new hypotheses might be investigated. For example, does bridging increase the importance of an attribute (i.e., are the relative importances of bridging attributes higher than they would be under full designs)? What happens if there are shape reversals for utility functions of attributes estimated from bridging subdesigns? The latter would be particularly upsetting if it occurred for important attributes.

A much more fundamental question, which has not been asked often, and certainly not answered, is whether we need to estimate individual utility functions at all. It is this desire that leads to all those problems of information overload. Supposing we could reduce this overload by submitting to every respondent a set of stimuli too small to estimate his own individual utility function, and estimate utilities at the group level, what are the positive and negative consequences ?

On the positive side, respondents would be less taxed, and perhaps provide data of better quality, and if the response rate is a function of the number of stimuli, the cost of research could decrease.

On the negative side, one would be faced with the potential problem of having only at best group or segment utility functions, which would be aggregates of individual preferences. In theory one can expect that the use of segment utility functions should lead to a decrease in the quality of conjoint results. The size of this decrease could be reduced by using optimal segmentation methods discussed earlier or any good clustering algorithm on whatever individual information we have.

Whether these effects will cancel out, is probably again an empirical matter, which should be tested with some urgency, because if aggregation is not a major problem, it opens up new perspectives about the use of conjoint analysis.

Another fundamental question which has not been answered thoroughly is about the size of the information overload problem in conjoint studies. Does a larger number of attributes and levels lead to less reliable data and thus to predictions of lower quality ?

We can distinguish between two sources of overload : simultaneous processing of large amounts of information (as in a ranking task of 16 or more cards containing 5 or more attributes) with the result that some information will not be used (properly), or consecutive processing (comparing a large number of choice sets containing e.g. 2 stimuli) leading to respondent fatigue and a decrease in response quality.

How much do we gain by reducing the information overload in terms of predictive validity and perhaps also respondent cooperation ?

As Carroll and Green (1995) have already pointed out, conjoint researchers and users are in need of studies comparing the different approaches to the different problems plaguing conjoint instead of continuing to propose new variants on an old theme.

References

Allenby, G. M., Arora N. & Ginter, J. L. (1995). Incorporating Prior Knowledge into the Analysis of Conjoint Studies. *Journal of Marketing Research* 32 (May), 152-16.

Allenby, G. M. & Ginter, J. L.(1995). Using Extremes to Design Products and Segments Markets. *Journal of Marketing Research*, vol 32 (november), 392-403.

Bridger (1988), New York: Bretton-Clark.

Carroll, J. D. & Green, P. E. (1995). Psychometric Methods in Marketing Research : Part I, Conjoint Analysis. *Journal of Marketing Research*, Vol. 32 (November), 385-391.

Cattin, P.& Wittink, D.R. (1982). Commercial Use of Conjoint Analysis: A Survey. *Journal of Marketing*, 46 (Summer), 44-53.

Conjoint Analyzer (1987), New York: Bretton-Clark.

Desarbo, W. S., Carroll, J.D., Lehmann, D.R. & O'Shaughnessy , J. (1982), "Three-way Multivariate Conjoint Analysis", *Marketing Science*, 1 (Fall), 323-50.

Green, P. E. (1984). Hybrid Models for Conjoint Analysis: An Expository Review. *Journal of Marketing Research* 21 (May), 155-159.

Green, P. E. & Krieger, A.M. (1995). Attribute Importance Weights Modification in Assessing a Brand's Competitive Potential. *Marketing Science* , vol 14, 253-270.

Green, P. E., Krieger, A.M. & Schaffer, C.M. (1993). A Hybrid Conjoint Model with Individual Level Interaction. *Advances in Consumer Research*, 20, 1-6.

Green, P. E. & Srinivasan, V. (1990). Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice. *Journal of Marketing* 54 (October), 3-19.

Hagerty, M. R. (1985). Improving the predictive power of conjoint analysis : the use of factor analysis and cluster analysis. *Journal of Marketing Research*, 22 (May), 168-84.

Huber, J.& Zwerina, K. (1996). The Importance of Utility Balance in Efficient Choice Designs. *Journal of Marketing Research* 33 (August), 307-317.

Johnson, R. M. (1987). Adaptive Conjoint Analysis. *Sawtooth Conference on Perceptual Mapping, Conjoint Analysis, and Computer Interviewing*, Ketchum, ID: Sawtooth Software, 253-266.

Kamakura, W., Wedel, M.& Agrawal, J. (1994). Concomitant Variable Latent Class Models for Conjoint Analysis. *International Journal of Research in Marketing*, 11, 451-64.

Lenk, P. J., Desarbo, W.S., Green, P.E. & Young, M.R. (1996). Hierarchical Bayes Conjoint Analysis : Recovery of Partworth Heterogeneity from Reduced Experimental Designs. *Marketing Science,* vol 15, N° 2, 173-191.

Mehta, R., Moore, W.L. & Pavia, T.M. (1992). An Examination of the Use of Unacceptable Levels in Conjoint Analysis. *Journal of Consumer Research*, 19 (december), 470-476.

Miller, G. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *Psychological Review*, 63, 81-97.

Oppewal, H., Louviere, J.J. & Timmermans, H.J.P. (1994). Modeling Hierarchical Conjoint Processes with Integrated Choice Experiments. *Journal of Marketing Research* 31 (February), 92-105.

Pekelman, D. & Sen, S.L. (1979). Improving Predictions in Conjoint Analysis. *Journal of the American Statistical Association*, 75 (december), 801-16.

Reibstein, D., Bateson, J.E.G. & Boulding, W. (1988). Conjoint Analysis Reliability: Empirical Findings. *Marketing Science* 7 (Summer), 271-286.

Srinivasan, V. & Park, C.S. (1997). Surprising Robustness of the Self-Explicated Approach to Customer Preference Structure Measurement. *Journal of Marketing Research*, vol 34 (May), 286-291.

Wittink, D. R., Krishnamurthi, L. & Reibstein, D.J. (1990). The Effects of Differences in the Number of Attribute Levels on Conjoint Results. *Marketing Letters.*

Wittink, D. R. & Cattin, P. (1989). Commercial Use of Conjoint Analysis: An Update. *Journal of Marketing*, 53, 91-96.

Wright, P. (1975). Consumer Choice Strategies: Simplifying vs Optimizing. *Journal of Marketing Research*, 12 (February), 151-175.

Table 1

Attributes and Levels for Vacuum Cleaners

Attributes	Levels		
Power (watt)	700	1000	1300
Price (Belgian francs)	3999	6999	9999
Warranty (years)	1	2	3
Power control (positions)	none	2	4
Type of cleaner	drum	box	
Accessories	yes	no	
Indicator Light	yes	no	
Cord	yes	no	

Table 2

Bridging Design for Vacuum Cleaners

Attribute importance	2 important attributes		1 important, 1 unimportant		2 unimportant attributes	
	set 1	set 2	set 1	set 2	set 1	set 2
Bridging attributes	price power	price power	price warr.	price warr.	warr. contr.	warr. contr.
Other attributes	warr. type access.	contr. indic. cord	contr. type access.	power indic. cord	price type access.	power indic. cord
Number of stimuli	16	16	16	16	16	16
Sample size	40		40		40	

Table 3

Measures of Solution Quality and Predictive Validity

of the Bridging Methods and Full Design

	Symbridge	Bridger	Full Design
Solution Quality (means, n=120)			
Inconsistency {sum($diff_j^2$)}	.120	.077	
Correlation	.860	.840	
Predictive Validity			
Spearman correlation	.51	.49	.61
1st place hits	38.3%	37.5%	40.0%
1st or 2nd place hits	55.0%	53.3%	60.0%

Note: Chance criterion for 1st place hits is 16.7% and for 1st or 2nd place hits is 33.3%; all results are significantly above chance ($p < .001$).

Table 4

Attribute Relative Importances Estimated by Different Methods

Attribute	Self-Explicated		Symbridge		Bridger		Full Design		A Priori
Means (SDs)									
Power	81	(23)	72	(36)	59	(38)	48	(34)	100
Price	58	(32)	38	(32)	26	(28)	27	(29)	50
Warranty	57	(33)	31	(30)	27	(29)	22	(23)	20
Power control	55	(34)	32	(28)	26	(27)	28	(28)	30
Type	69	(35)	49	(38)	49	(42)	65	(44)	50
Accessories	24	(28)	18	(22)	19	(26)	9	(15)	40
Indicator Light	30	(28)	19	(23)	15	(19)	13	(20)	40
Cord	54	(30)	21	(27)	16	(23)	13	(19)	30
Correlations									
Self-Explicated	1.00								
Symbridge	.88		1.00						
Bridger	.83		.97		1.00				
Full Design	.79		.83		.90		1.00		
A priori	.55		.84		.78		.55	1.00	

Note: The scales of the relative attribute importances were made equivalent to each other as nearly as possible. The correlations in the lower half of the table are based on the averages in the top half of the table.

Table 5

Descriptive Statistics and Intercorrelations for Independent Variables

Descriptive Statistics

	Mean	Std. Dev.	Median	Min.	Max.
Importance (bridging attrib.)	.93	.39	.88	.19	1.88
Imbalance 1	.16	.13	.14	.00	.59
Imbalance 2	.45	.36	.41	.00	1.24
Inconsistency	.03	.04	.02	.00	.21
Total Importance (all attrib.)	62.79	19.00	62.98	30.25	122.44

Correlations (n=120)

	Import.	Imbal.1	Imbal.2	Incons.	Tot.Imp.
Importance	1.00				
Imbalance 1	.13	1.00			
Imbalance 2	.82	.16	1.00		
Inconsistency	.47	.05	.28	1.00	
Total Importance	-.24	.32	-.17	-.18	1.00

Table 6

Effect of Independent Variables on Predictive Validity

Independent Variable	Symbridge	Bridger
Correlations (p-values)		
Importance	-.05 (.598)	-.03 (.718)
Imbalance 1	.14 (.140)	.13 (.158)
Imbalance 2	.03 (.709)	.07 (.471)
Inconsistency	-.06 (.510)	-.09 (.350)
Total Importance	.32 (.000)	.28 (.002)

Note: The measure of predictive validity used here is the Spearman correlation between predicted and actual holdout stimuli rankings. Thus, the tabled values are Pearson correlations between the importance measure and the Spearman correlations, where the replicates are the 120 respondents. All p-values are for two-tailed tests.

Authors’name and affiliation.

Pierre François is Associate Professor of Marketing at Catholic University of Leuven, Naamsestraat 69, B-3000 Leuven, Belgium.

Douglas L. MacLachlan is Professor of Marketing at University of Washington, Mackenzie Hall, Box 353200, Seattle, WA 98195-3200.

Authors’ vitae

Pierre François holds an MBA and PhD from Catholic University of Leuven, Department of Applied Economics, where he teaches Marketing Research and Industrial Marketing. His main areas of research are presently Pricing of industrial products, Measurement of trade show effects, Geomarketing and Choice and conjoint modelling.

Douglas L. MacLachlan is Associate Dean for Academic Programs and Professor of Marketing, **University of Washington Business School**. He is also Executive Director of University of Washington’s Center for International Business Education and Research. His research has focused primarily on methodological issues in marketing research and he has published in most of the major marketing and management journals. Work on this project was initiated while he was a visiting professor at Catholic University of Leuven 1991-92.