BEST PRACTICE IN SP DESIGN

Nobuhiro Sanko
Nagoya University[i]
Andrew Daly and Eric Kroes
RAND Europe

1. INTRODUCTION

In recent years, widespread use has been made of Stated Preference (SP) data, but consensus concerning appropriate approaches has not been achieved. Instead, a number of competing 'schools' – in the US, Australia and Japan as well as in Europe – have each developed their own methodology and although there is argument between them there has been little attempt to review the alternative approaches and select the best of each.

Disagreement between the schools extends to many areas of SP practice but is particularly marked in the design of experiments. For this reason, a study of SP experimental design was undertaken (Sanko 2001), which is summarised in the present paper. An attractive approach is to review as many of the competing approaches as can be understood from the published literature, to review these critically in the light of the alternatives adopted by other groups and to identify the most appropriate for specific circumstances. The literature is diverse, ranging from papers in academic journals to commercial advertising, and another aim of the study is to help researchers confronted by this mass of information.

In this study, the focus was on SP using choice experiments. Other methods of elucidating preference require different design procedures and a consideration of these was beyond the scope of the present study. Nevertheless, choice experiments are a widely-used SP procedure and a resolution of the design issues for this procedure would be useful. Special attention is given to statistical aspects of the design.

The attention throughout is on deriving recommendations for practical work. In some cases it is possible to lay out clear alternatives for use in different circumstances, but some areas are clearly in need of further work. The paper gives an overview of a number of relevant design approaches, associated problems and possible solutions. It ends by providing a practical step-by-step framework for experimental design for SP choice experiments; the examples in this paper are for binary choice but the principles extend to multiple choices.

2. TERMINOLOGY

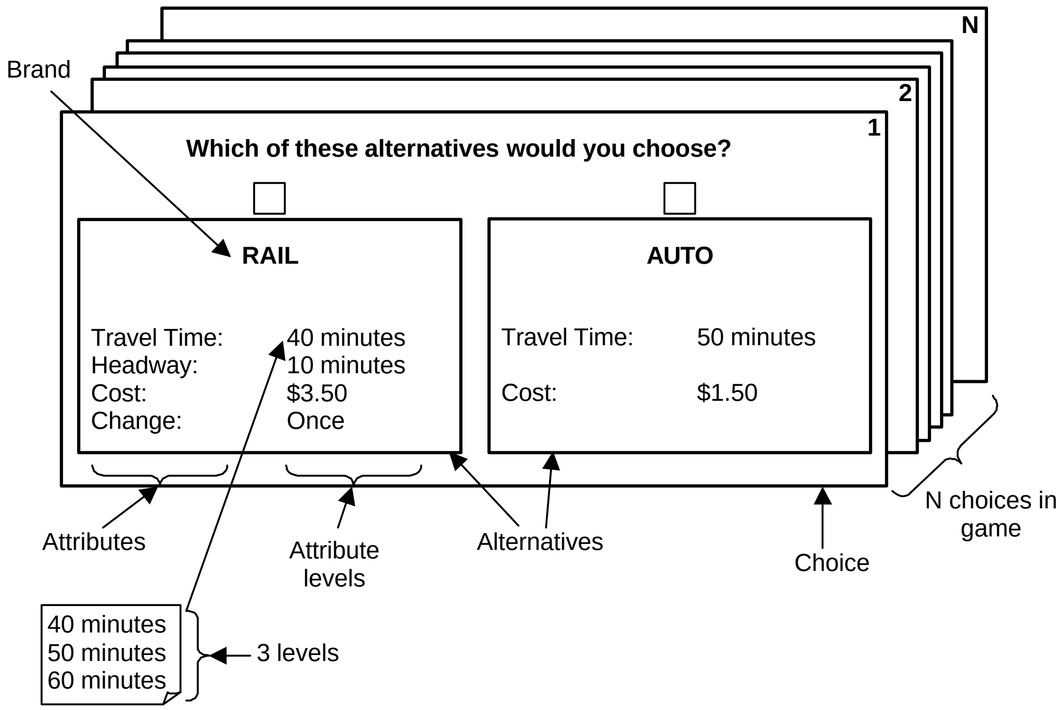
To clarify the discussion it is necessary to define some of the terms used to describe choice experiments – usage differs among the various researchers working in the field. Figure 1 illustrates the main features of a typical SP experiment. A respondent is asked to make a series of *choices*. In this case, the Figure shows N such choices, which are organised into a *game* or *experiment*. In a practical survey, one, two or sometimes more games may be presented to each respondent.

In each choice, the respondent must select one of the *alternatives* in the *choice set*. In this example, the choice set has two alternatives which are identified by a name or *brand*, in this case RAIL and AUTO (an 'opt-out' response, i.e. "neither of the above" is sometimes also presented). In this paper, attention is given only to the most common case, that of the *fixed choice set*. Variation in the choice set can be handled in a separate game but we do not consider issues of design **between** games.

An alternative is characterised by its *attributes* and *attribute levels*. In the example, RAIL has four attributes: Travel Time, Headway, Cost and Change. The attributes have specific levels in each choice, for example, Travel Time has three levels: 40, 50 and 60 minutes. The combination of levels presented in a specific choice is called a *scenario*.

The SP design issue, as considered in this paper, is exactly how Figure 1 should be constructed for each respondent.

Figure 1: Choice-Based SP Game



3. FUNDAMENTAL DESIGNS

In principle, the simplest organisation of a game, once attribute levels have been fixed, is to present all the possible combinations to each respondent. This is called a full factorial design and forms the basis from which other designs can be derived.

3.1 Full Factorial Design

The specification of a full factorial design is straightforward: an example is illustrated in Table 1, which shows the design for a game with three attributes, each of two levels. The design can also be presented numerically, as shown in the right side of the table, and in that form it is transferable to any other context of three attributes with two levels.

Table 1: Full Factorial Design: three attributes, two levels each

Scenario	Attributes		
	Fare	Time	Freq
1	High	Slow	Low
2	High	Slow	High
3	High	Fast	Low
4	High	Fast	High
5	Low	Slow	Low
6	Low	Slow	High
7	Low	Fast	Low
8	Low	Fast	High

Scenario	Attributes		
	At.1	At.2	At.3
1	0	0	0
2	0	0	1
3	0	1	0
4	0	1	1
5	1	0	0
6	1	0	1
7	1	1	0
8	1	1	1

An important characteristic of the full factorial design is that of *orthogonality*, that is, the values of the attributes are independent. In the full factorial design, orthogonality applies also to *interactions* between the attributes, so that if the value of one attribute depends on the value of another, this can be fully identified from the experiment.

However, it is clear that when the number of attributes or the number of levels increases, the number of scenarios required for a full factorial design can become very large. For example, if there are four attributes, each with four levels, the full factorial design requires $4^4 = 256$ scenarios.

3.2 Fractional Factorial Design

A fractional factorial design is based on a systematic selection of a number of rows from a full factorial design. For example, using the shaded rows in Table 1 would represent a fractional factorial design. The obvious advantage of a fractional factorial design is that it reduces the number of scenarios that are presented to respondents.

A potential problem with fractional factorial designs is that orthogonality is lost in at least some respects. This may affect the main effects, i.e. the values of the attributes themselves become correlated, but it is not always the case, for example in the shaded rows of Table 1. It is very likely that it will affect interactions, i.e. it may become difficult (or impossible) to identify interactions between some or all the attributes when smaller fractional factorial designs are used. In the shaded rows of Table 1, interactions are no longer orthogonal.

The full factorial and fractional factorial designs were developed for experiments in which a single scenario is presented at one time. In experiments based on choices, at least two scenarios are presented so that the respondent can choose between them. This extension adds another layer of complexity to the design.

3.3 Creation of Choice Sets

The methods applied in the literature (e.g. Louviere *et al.* 2000) for combining scenarios to obtain choice sets fall into three typical groups.

a. Simultaneous choice set creation

The typical method of this type is the 'L^{MN}' procedure, giving choice sets of N alternatives with M attributes, each of L levels. In the example of Table 1, we would have L=2, M=3 and N=2, giving 64 choices which cover all possible combinations of the levels. This design can be represented in a single table by using separate columns for the attributes of each alternative, thus obtaining N*M=6 columns, and in this form a fractional factorial design can be developed by selecting rows as was indicated above. For this particular problem, the smallest design that maintains orthogonality for the main effects has 8 rows, i.e. 8 choices.

b. Sequential choice set creation

The *fold-over* design procedure creates the choice sets by creating scenarios for subsequent alternatives based on the scenario for the first alternative. The procedure starts from a full or fractional factorial design, which forms the scenarios for alternative A. The scenarios for alternative B are then formed by a systematic transformation of the attributes, changing some to the next higher value, leaving others unchanged, etc.. A simple change of all the attributes to their next level is called *shifting*; more complicated transformations, often involving a random element (*shuffle*) are called *fold-over* designs.

For example, Table 2 shows an application of these procedures to the simple problem of Table 1. Alternative A uses the scenarios of the fractional factorial design (i.e. the shaded rows in Table 1). Alternative B on the left side uses a fold-over, changing 0's to 1's and 1's to 0's for attributes 1 and 2 and not changing attribute 3. On the right side a shifting design is implemented in which all attributes change.

Table 2: Two Fold-overs based on Fractional Factorial Design
three attributes, two levels each, as in Table 1

Fold-over without Shuffle						
Ch.	Alt. A			Alt. B		
	At.1	At.2	At.3	At.1	At.2	At.3
1	0	0	0	1	1	0
2	0	1	1	1	0	1
3	1	0	1	0	1	1
4	1	1	0	0	0	0

Shifting						
Ch.	Alt. A			Alt. B		
	At.1	At.2	At.3	At.1	At.2	At.3
1	0	0	0	1	1	1
2	0	1	1	1	0	0
3	1	0	1	0	1	0
4	1	1	0	0	0	1

c. Randomised choice sets

Randomised choice sets are generally based on a full or fractional factorial design. Then alternatives are randomly selected (with or without replacement) from these to present to the respondent. When both alternatives relate to the same brand, selections are generally made from a single set; when alternatives relate to different brands, as in the example, selections are generally made from two independently created factorial designs.

4. PROBLEMS OF FACTORIAL DESIGNS

As has already been indicated, the key problem with the full factorial design is that it generates too many scenarios. In the context of a survey in the field, this is not acceptable because of the impact on respondents. A fractional factorial design obviously reduces the problem, but in the context of choice experiments there may still be too many choices to be made.

Another problem is dominance. In factorial designs, there are often choices in which one alternative 'dominates' the other, i.e. it is better on all the attributes. In Table 1, for example, scenario 8 dominates all of the others, so any choice in which it appears is 'trivial', i.e. the response is in principle known in advance and does not, in principle, add any useful information. In some cases, it is not clear whether an attribute is of positive or negative value (e.g. 'smoking is permitted') but often, as in the example, it can be expected that all respondents ought to take the same view. Even when ambiguously valued attributes are included, dominance can apply between alternatives for which this attribute is the same.

Assuming transitivity, the notion of dominance can be extended. If, say, in choice 1, scenario A is preferred to scenario B and if, in choice 2, scenario B is preferred to scenario C, then in choice 3 comparing scenarios A and C ought to produce an obvious answer. This is simply the application of transitivity. Going further, if a respondent in the game of Table 1 chooses scenario 2 over scenario 3, then (because scenario 6 dominates scenario 2) we can avoid asking for the comparison of scenarios 6 and 3 in later choices. This effect can be termed transitivity + dominance.

In some cases, contextual constraints apply which constitute another potential problem. This may mean that certain combinations of attributes are not reasonable. For example, if driving time and distance were attributes of a car journey, we would do well to ensure that the speed implied by each combination of attributes was reasonable before presenting it to respondents.

Finally, the issue of orthogonality is more complicated than it may appear. A simple point made by Hensher (1994) is that in choice modelling orthogonality needs to apply to attribute differences rather than to the base values themselves, because it is the differences between the attributes that drive a choice model. However, in modelling it is often necessary to change the formulation of the variables, e.g. by non-linear transformations, or to introduce interaction terms with socio-economic variables. Orthogonality with respect to these transformed variables is effectively impossible to achieve in advance. In any case, the direct relationship between design attribute orthogonality and parameter estimate orthogonality, which applies for linear models, does not apply for non-linear models used to represent choice.

The avoidance of these problems and the maintenance of efficiency in estimation is the key task of developing practical designs which build on the factorial designs and basic choice set creation procedures discussed above.

5. ASSESSMENT OF FACTORIAL DESIGNS

The design methods presented in Section 3 can be assessed in the light of the problems discussed in Section 4.

a. Simultaneous choice set creation (L^{MN})

The advantage of this approach is its simplicity and orthogonality, achieved at the expense of a large number of choices. However, a further problem with this approach is that a large fraction of the choices are trivial since one alternative dominates the other. This is illustrated in Table 3, which applies to the example introduced in Table 1. In this case, no fewer than 46 of the 64 choices are trivial, if we assume that attribute level 1 is always preferable to level 0.

Table 3: L^{MN} Method for Full Factorial Design (Binary, Three Attributes, Two Levels Each)

Choice	Alternative A			Alternative B			Trivial
	Fare	Time	Freq	Fare	Time	Freq	
1	0	0	0	0	0	0	Trivial
2	0	0	0	0	0	1	Trivial
…	…	…	…	…	…	…	…
13	0	0	1	1	0	0	OK
14	0	0	1	1	0	1	Trivial
15	0	0	1	1	1	0	OK
…	…	…	…	…	…	…	…
64	1	1	1	1	1	1	Trivial

b. Sequential choice set creation

This procedure greatly reduces the number of choices relative to simultaneous creation of choice sets. Moreover, the frequency of trivial choices depends on the way in which attribute shifting is implemented and the efficiency of the design can work out substantially better. For example, in the minimal shifting design (mentioned above) for the example of Table 1, there are just 2 trivial choices of the 8 that are needed.

c. Randomised choice sets

In this case, the number of choices for each individual is controlled by the analyst. In principle, if homogeneity is assumed, there is no loss in the identification of effects relative to the design on which the random experiments are based.

In the standard example, 28 different choice sets (8 * 7 / 2) can be created, a full set, and sampling can be done within this set. However, 19 of the 28 choices are trivial.

6. OTHER METHODS

We have illustrated in sections 4 and 5 some of the main problems associated with factorial designs. Below we present a number of practical methods to overcome these problems. The starting point or benchmark for the comparison is the full factorial design. For each method we briefly describe:

- the purpose, the problem it aims to address,
- the main approach to solving the problem,
- the underlying assumptions that are implicit in its use,
- the trade-off involved, what is sacrificed and
- the justification for its use.

6.1 Fractional Factorial Design

As we have seen already, the main aim of the fractional factorial design is to reduce the number of scenarios offered to each respondent. This is achieved by carefully selecting a specific subset of scenarios from all the scenarios included in the full factorial design, in such a way that the main effects can still be properly estimated. The trade-off is that some or all of the interactions can not be estimated and hence the implicit underlying assumption is that these are not significant in explaining preferences. In practical research it has often been observed that indeed the main effects account for most of the observed variance in the preferences, and the interactions add little to that, although in some cases specific interactions may be of importance (Louviere, 1988).

6.2 Removing Trivial Choices

This is another method to reduce the number of choices offered to the respondents, while it also avoids asking questions to the respondents that yield no meaningful information (and may indeed generate irritation among the respondents). Trivial choices can be identified using information about the sign of the utility associated with the attribute, together with dominance and transitivity assumptions. By removing trivial choices, however, orthogonality is normally lost, and estimation problems may arise. One approach that has been used to overcome such problems at estimation stage has been to re-insert the removed trivial choices prior to estimation, together with inferred preferences (based on the assumed transitivity and dominance rules). However, this approach cannot be recommended as it simply adds arbitrary 'information' to the data.

6.3 Contextual Constraints

By removing scenarios that are technically impossible or highly unlikely, the SP exercise preserves credibility for the respondents, and at the same time the number of scenarios is reduced. However, this normally leads to correlations between the attributes, and hence loss of orthogonality. Note that here it would be impossible to re-insert the removed scenarios prior to estimation (see 6.2) even if we wanted to, as the preferences cannot be inferred. The application of contextual constraints to eliminate scenarios requires prior knowledge about what is technically impossible and seen by the respondents as unlikely or unreasonable.

An alternative approach to eliminating certain impossible or unlikely scenarios is retaining the scenarios but with modification of one (or more) attribute levels relative to the experimental design. In this way the correlation effect (and hence the estimation problem) can be reduced, but not always avoided. The number of scenarios remains the same.

6.4 Use a Block Design

Another approach to reducing the number of choices offered to a single respondent is by dividing the full set of choices of a single experiment into two (or more) subsets (separate experiments or "blocks"), each of which is offered to different respondents. For example a full factorial design may be subdivided into four fractional factorials (a), or a fractional factorial may be subdivided into four partial subsets (b). In case (a) individual level estimation of the main effects is still possible, but the interactions can only be estimated by pooling data of multiple respondents. In case (b) even estimation of the main effects only requires pooling of observations. A key requirement for this is that the underlying assumption, homogeneity of the preferences across the respondents, is a valid one.

6.5 Use Common Attributes in Multiple Experiments

If the number of attributes to be included in a SP experiment becomes large (normally more than say 5 or 6) it is no longer possible to include them all in one single experiment. This is because the number of required scenarios becomes too large, but also because it leads to information overload for the respondents (too many attributes to consider simultaneously). In such cases, the very large single experiment can be divided into multiple sub-experiments, each containing only some of the attributes, but which all have one attribute in common (typically price). This design makes it impossible to identify interactions between attributes which are in different games, so the subdivision requires some careful structuring. It also requires that the common attribute is estimated with reasonably high accuracy in all the sub-experiments, so that the estimation results can be related to each other by using the estimates for the common attribute.

6.6 Define Attributes in Terms of Differences between Alternatives

As we have seen in section 4, choice modelling requires that the differences between the attributes in the choice pairs are orthogonal rather than the absolute levels. One way to ensure this is by explicitly specifying the attributes in terms of differences between the alternatives, rather than in absolute levels for each of the alternatives. A key requirement for this is that all attributes are "generic', i.e. they apply to all alternatives in the same way. Attributes which are specific to a single alternative only can not be used. This method is useable when the SP experiment investigates preferences for different versions of the same type of alternative, also indicated as "within-mode" experiments.

6.7 Show One Design Differently

This technique uses a single basic statistical design, but in mapping the attribute levels onto the experimental design, random variation is introduced. As a consequence the same experimental design actually looks different to different respondents. This approach may be used when the base fractional factorial design enables only some of the interactions to be estimated or even none. By randomly showing the design differently to the respondents, all the interactions may be estimated by pooling the observations. As in 6.4 this presupposes homogeneity of preferences. A side benefit is that the randomisation will avoid any possible influence of the order in which the attributes are presented, and may increase the efficiency of the estimation.

6.8 Random Selection

A final technique to mention here, which is sometimes used to reduce the number of choices to be evaluated by each respondent, is the random selection of choices from a factorial design. In a way this is similar to using a block design, but here the selection is done randomly rather than by the "block" logic. As a consequence, individual level estimation is not possible, and orthogonality will normally be lost. But by pooling all observations and assuming homogeneity, all the parameters can be estimated.

7. SETTING ATTRIBUTES AND ATTRIBUTE LEVELS

We have discussed the need to limit the number of scenarios which is offered to the respondents for evaluation for practical reasons. Given the nature of experimental designs, this in turn limits the number of attributes and levels that may be presented in any one experiment.

Normally the number of attributes to be included in a single experiment should be limited to a maximum of 6 or 7 (in most cases 4 or 5 attributes works well). In case valuations for more than 7 attributes are required, the solution described in 6.5 may be used.

Concerning the attribute levels, the need to ensure that competitive trade-off decisions are presented requires:

1. that attribute levels presented to the respondents cover a sufficiently wide range to include likely boundary values (implicit trade-off switching points) between attributes, and

2. that attribute levels are sufficiently evenly spread to allow sufficiently accurate estimates of the parameters, whatever their actual values may be within the range.

This, in combination with the limited *a priori* information that may exist with regard to the boundary values, tends to increase the number of attribute levels. One practical way of allowing a large number of attribute levels across all responses while limiting the number to three or so for each individual respondent is by randomly selecting say 3 levels for each respondent from a larger total number of levels. This concept is similar to what was outlined in 6.4 and 6.8, but applied to attribute levels rather than scenarios.

When good *a priori* information is available about parameter values, designs can be highly optimised (see 8.2), but this will not often be the case.

8. DEPARTURES FROM THE ORTHOGONAL DESIGN

When the objective is to estimate mean coefficient values with minimal standard error, then an orthogonal design (the closest approximation of estimation data orthogonality) provides, for a given sample size and when a linear model is used, the best result. In other cases, however, this is not necessarily the case. Toner *et al.* (1999) found that when non-linear models are used (and nearly all choice models are non-linear) the orthogonal design does not always minimise the variance. Further, Fowkes *et al.* (1993) demonstrated that in some specific cases important reductions of error variance of the estimates could be obtained by using non-orthogonal designs.

8.1 Ratio Estimators

When the main objective of a SP experiment is to provide estimates of the monetary value-of-time (or a similar trade-off ratio), the analyst needs to minimise the error variance of the ratio estimators (time coefficient divided by cost coefficient), given approximately by

$$\text{c.o.v.}^2 (\beta_1 / \beta_2) \approx \text{c.o.v.}^2 (\beta_1) + \text{c.o.v.}^2 (\beta_2) - 2.\, \rho_{12}.\, \text{c.o.v.} (\beta_1) \,.\, \text{c.o.v.} (\beta_2)$$

where c.o.v. is the coefficient of variation, the standard error divided by the estimate of the parameter, the β's represent the coefficients to be estimated and ρ gives the correlation of the estimates.

If the values of the two coefficients were known, the level of correlation that maximises the accuracy of the ratio estimator could be determined in advance. In reality, of course, the analyst does not have exact knowledge about the values of the coefficients before the experiment, so one cannot really optimise. But by using some prior expected values, a reasonable estimate can be obtained of the amount of correlation that would be approximate to improve the estimation (see Fowkes *et al.* (1993)).

8.2 The "Magic" Choice Probability

Another interesting design issue is the identification of what Toner *et al.* (1998) called "the magic P", based on much earlier work by Gunn (1983). They demonstrated that, for a two variable binary logit model and generic coefficients, the necessary condition for the variance of each of the coefficients to be minimised is to specify the two alternatives so that their utility difference is about $\pm$ 2.4. This leads to the "optimal" choice probabilities for the two alternatives of 0.917 and 0.083, the magic P. This is in conflict with the intuitive belief held by many researchers (e.g. Bates, 1994) that the most useful information is obtained where respondents are on the borderline between one alternative and another (i.e. P= 0.50).

9. SP DESIGN FRAMEWORK

Having covered some of the main issues involved in the design of SP choice experiments, we now turn to providing practical researchers some help in designing their SP choice experiments. In order to do this we have designed a systematic approach to SP design. The starting point for the framework is a full factorial design, derived from the specification of the attributes and attribute levels that need to be included in the experiment (Step 1). Then we ask 8 questions to address the problems involved in the use of the full factorial, to arrive at an appropriate practical SP design (Steps 2 through 9). This process is summarised in Figure 2. In the different steps we give recommendations, where possible, of recommended "default" strategies, based on the discussion above and our practical experience; these are marked with an asterisk in the figure.

Step 1: Setting attributes and attribute levels (see Section 7)

Here you need to specify which attributes you want to include in the experiment, and how many levels you set for each attribute. Generally more than 2 attributes are suggested to be included in the experiment, and more than 2 levels are suggested to be included in the important attributes. If you want to investigate non-linear effects you need at least 3 levels. You are advised to check the boundary values, and set attribute levels in order to obtain reasonable trade-off values. When you are interested in "define attributes in terms of differences between alternatives" (see 6.6) you need to consider this here.

Step 2: Is it possible to treat all the attributes in one SP exercise?

An upper limit of the number of attributes in one experiment is 6 or 7; in practice 4 or 5 attributes are often used. When you treat more than 5 attributes it is worth considering assigning the attributes to more than one experiment, while using (at least) one common attribute (see 6.5). Since it is impossible to estimate interaction effects between attributes included in different experiments, you need to think which interactions you are interested in, and put these attributes in the same experiment.

Step 3: Should you use an orthogonal design?

If you are primarily interested in an analysis of the ratio between two estimates (e.g. value-of-time) and have *a priori* knowledge of the estimated parameters, then you can develop a correlated design along the lines Fowkes (1993) proposed for this (see 8.1). Or you may also want to consider the 'Magic choice probability' if you have *a priori* knowledge of the parameters to be estimated (see 8.2). In both cases you can skip steps 4 to 9. However, these two methods are relatively new and require advanced statistical knowledge and knowledge of the parameter values. If you don't have this knowledge, we do not recommend these methods. In those cases, or if you are not interested in the departure from orthogonal design anyway, go to Step 4.

Step 4: Are you interested in interactions?

In many cases it is possible to concentrate in the research on the main effects only, as important interactions between attributes are not known or expected to exist. In such cases we recommend ignoring all interactions, following stream C in Figure 2. Here you use a fractional factorial design with the minimum number of scenarios.

In other cases, however, interactions are known to be of importance, and this needs to be taken into account in the experimental design. If <u>all</u> interactions are to be investigated, a full factorial design is needed which brings you to Stream A in Figure 2. If <u>some</u> interactions are to be considered, an appropriate fractional factorial design will be needed. In case you are able to accommodate a fractional factorial design which includes all interactions you are interested in, we call this an “enough fractional”, which we classify again under stream A.

If you are interested in some interactions, but can not accommodate these (e.g. because the number of scenarios of the enough fractional becomes too large), we suggest Stream B in Figure 2. Here you will need a small fractional factorial design designed to measure only some key interactions (or even none at all) combined with the “show one design differently” method (see 6.7). If you can assume homogeneity this allows you to estimate some or even all of the interactions across the sample, provided the sample size is adequate. The difference from the stream A and C designs is related to the way in which the interaction effects are treated:

- in stream B you rely on the “show one design differently” method to be able to estimate the interaction effects across the sample;
- in stream A the design is set up to measure the interaction effects for each respondent separately;
- in stream C interaction effects are ignored.

Within streams A, B and C the usual approaches can be followed to arrive at suitable choice pair designs (see 3.3): simultaneous, sequential or randomised choice set creation.

We want to emphasize that it would be, in our view, highly useful to further investigate the issue of the use of the different methods to estimate interaction effects within and across respondents. Using Monte Carlo simulation methods more insights should be obtained into the practical differences between the stream A and stream B approaches, and the associated requirements in terms of sample sizes. We recommend more research in this area, particularly focused on SP choice designs. Until such research has been carried out, it would be prudent to follow the stream A approach to measure key interaction effects.

Step 5: Do you want to show one design differently?

In the process of reducing the number of scenarios to be evaluated, some of the advantages of the original factorial designs are lost. As we have seen already in Step 4 one possible solution is showing one design differently for each respondent. This is particularly advantageous for designs created through stream B of Step 4, but is possibly also useful for stream C designs. Further research would be desirable here.

Step 6: Are you concerned about contextual constraints?

If you are interested in maintaining as much as possible reality in your experiment, which should normally be your aim in any SP experiment, you should remove (or modify) choice pairs which contain scenarios which are in conflict with contextual constraints (see 6.3). In this process, you lose orthogonality.

Step 7: Do you care about trivial questions?

The use of any experimental design will lead to a number of trivial questions, due to dominance (see 6.2). Often these can be identified in advance. We would recommend that the questions which are certainly trivial should be removed. If they are removed, orthogonality is lost, but the respondent is less likely to lose interest in the choice experiment. It can be useful to keep at least one trivial choice pair in the experiment as a means to check the reliability of the responses given by the respondents. In identifying and removing trivial questions you need to be very careful not to make strong assumptions on preference.

Step 8: Do you need to make a special allocation of tasks to respondents?

It is generally recommended to limit the number of choice pairs for each individual. Pearmain *et al.* (1991) suggested a maximum of 9 to 16 choices per respondent (for a single experiment) and Bradley and Daly (1991) showed that the quality of responses declined as the number increased in this range. If you need more scenarios, a task allocation strategy is recommended. There are two techniques to do this:

- Use of a Block design (see 6.4);
- Use of Random selection (see 6.8).

Step 9: Can you assume transitivity?

Another strategy to reduce the number of choices to be made by each respondent, is to use transitivity to remove trivial scenarios based on former responses (see 4 and 6.2). In this case orthogonality is lost, but unnecessary questions are avoided. Unlike trivial questions related to dominance, these cannot be identified before the interview, but need to be established during the interview itself. This is difficult or impossible using conventional questionnaires, but can be done effectively when computer based interviewing takes place using appropriate software, e.g. WinMINT.

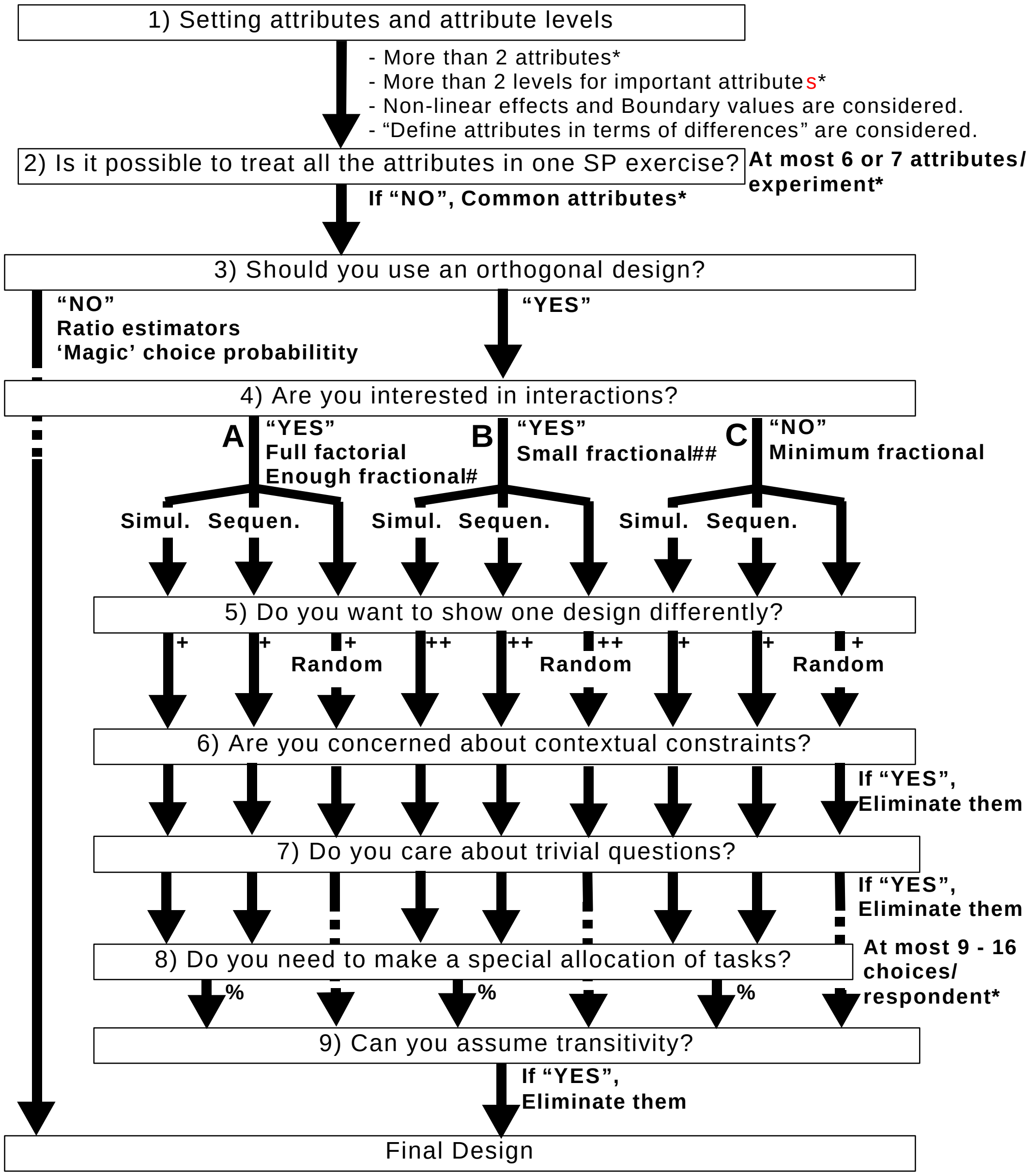
Concluding remarks

Having described the Steps 1 to 9 which we feel are a suitable process to arrive at an appropriate design for SP choice experiments, we want to make a few concluding remarks.

Firstly the description here has concentrated on alternatives of the same brand, also indicated as “within mode” alternatives. For “between mode” alternatives the same logic applies, but due to differences in the number of attributes or attribute-levels, some complications may arise.

Secondly we want to emphasise that every SP experiment, even the best designed one, should always be carefully piloted in the field, followed by a proper analysis of the pilot data. This is really the only way to establish the

Figure 2: SP Design Framework



***: Recommended and generally accepted Strategy/ Criteria**
+: If "YES", do fold-over for each respondent
++: The answer "YES" is highly recommended, do fold-over for each respondent
%: If "YES", Block design, or Random selection
#: Fractional factorial design which considers all interactions you are interested in
##: Fractional factorial design which doesn't consider all interactions you are interested in
(This includes minimum-sized fractional factorial design.)

quality of the measurement instrument, and the practical ability to obtain all the desired estimates.

Thirdly we want to argue for “virtual piloting” prior to the real pilot, as part of the design stage, by means of Monte Carlo simulation. Specialised software is available for this (e.g. SPEED developed by Hague Consulting Group), even the use of simple spreadsheet software enables the serious analyst to easily investigate the quality of SP designs. Simulation is a powerful tool, and an excellent means to assess elements of uncertainty in the design, although the results may depend on the assumptions made.

10. CONCLUSION

In this paper we have discussed some of the main issues involved in developing an appropriate experimental design for SP choice experiments. We have brought together a number of SP design elements from different “schools”, and recommended a practical step-by-step approach. It is clear, however, that more research needs to be done to investigate the merits of some of the proposed approaches, so that their potential benefits can be assessed and hopefully made available to a larger audience.

In particular we would like to recommend further research in the following specific areas:

- the use of the “show one design differently” (randomisation) method to estimate interaction effects across the sample which cannot be identified at individual level, and their homogeneity and sample size requirements;
- the use of “removing trivial choices” and “contextual constraints”, and their impact on the error variance of the estimates;
- the use of correlated designs and the “magic P” in practical multi-attribute experiments to obtain minimum error-variance estimates, preferably without requiring exact *a priori* knowledge of the parameters;
- the more general issue of how to create designs that minimise the error variance of the estimates when non-linear models such as logit are used.

In our view, these issues concerning SP choice design for efficient estimation should be systematically investigated using both Monte Carlo simulation and empirical SP data.

In the meantime, we hope that our paper will be of help for practical users in selecting the right, or at least an appropriate mix of more conventional methods to the design of SP choice experiments.

REFERENCES

- Bradley and Daly (1991): Estimation of Logit Choice Models Usung Mixed Stated Preference and Revealed Preference Information, 6th International Conference on Travel Behaviour, Quebec.
- Fowkes and Wardman (1993): Non-orthogonal Stated Preference Design, PTRC SAM.

- Hensher (1994): Stated Preference Analysis of Travel Choices: the State of Practice, Transportation 21, pp. 107 – 133
- Louviere (1988): Conjoint Analysis Modelling of Stated Preferences, Journal of Transport Economics and Policy, Vol 22, no 1.
- Louviere, Hensher and Swait (2000): Stated Choice Methods, Cambridge University Press.
- Pearmain, Swanson, Kroes, and Bradley (1991): Stated Preference Technique – A Guide to Practice (2nd Ed.), Steer Davies Gleave and Hague Consulting Group
- Sanko (2001): Guidelines for Stated Preference Experiment Design, MBA Dissertation, Ecole Nationale des Ponts et Chaussées, Paris.
- Toner, Clark, Grant-Muller and Fowkes (1998): Anything you can do, we can do better: A provocative introduction to a new approach to stated preference design, WCTR 8 at Antwerp
- Toner, Wardman and Whelan (1999): Testing Recent Advances in Stated Preference Design, European Transport Conference, Stream F, Cambridge.

---

[i] The work described in this paper was done during a student internship of the first author at RAND Europe which formed part of a Master's Thesis at Ecole Nationale des Ponts et Chaussées. Space here does not allow the full literature studied in that work to be listed.