

Louviere, Jordan J.; Pihlens, David; Carson, Richard

Article

Design of discrete choice experiments: A discussion of issues that matter in future applied research

Journal of Choice Modelling

Provided in Cooperation with:

Journal of Choice Modelling

Suggested Citation: Louviere, Jordan J.; Pihlens, David; Carson, Richard (2011) : Design of discrete choice experiments: A discussion of issues that matter in future applied research, Journal of Choice Modelling, ISSN 1755-5345, University of Leeds, Institute for Transport Studies, Leeds, Vol. 4, Iss. 1, pp. 1-8

This Version is available at:

<https://hdl.handle.net/10419/66807>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<http://creativecommons.org/licenses/by-nc/2.0/uk/>

Design of Discrete Choice Experiments: A Discussion of Issues That Matter in Future Applied Research

Jordan J. Louviere^{1,*} David Pihlens^{1, †} Richard Carson^{1,2, ‡}

¹ Centre for the Study of Choice (CenSoC)
The University of Technology, Sydney

² Department of Economics
University of California, San Diego

Received February 2009, revised version received September 2009, accepted June 2010

Abstract

Disagreement among researchers regarding types of optimal choice experiments is often best seen as resulting from differences in the set of assumptions researchers are willing to make about the underlying data generating process. Much of the current debate may have confused, rather than enlightened applied researchers because the underlying source of the debate lacks transparency. We argue that this debate would be better served if it were much more closely tied to the large existing literature on optimal design of experiments, where many of the issues currently being discussed have long been examined. We further argue that the current debate misses several key issues that are likely to be important to making progress in understanding the role played by experimental designs in applied settings of interest in economics, marketing and transportation research.

Keywords: Choice models, discrete choice experiments, experimental design criteria

* Corresponding author, T: +61 2 9514 9893, F: +61 2 9514 9897, Jordan.Louviere@uts.edu.au

† T: +9514 9795, F: +61 2 9514 9897, David.Pihlens@uts.edu.au

‡ T: + 01 (858) 534-3383, F: + 01 (858) 534-7040, rcarson@weber.ucsd.edu



1 Discussion

Discrete Choice Experiments (DCEs) began with Thurstone (1927) and experiments for pairs of choice options have been extensively studied since (e.g., see Bradley and Terry 1952; Davidson and Farquhar 1976; David 1988; Graßhoff and Schwabe 2008). While experiments and models for pairs of options can be useful, many real choice problems involve larger choice sets. For many years researchers in psychology and other fields used ad hoc combinations of pairs, triples, quadruples, etc, to study choices. Abelson and Bradley (1954) and Bradley and Helbawy (1976) provided optimal design results for the factorial design of preference experiments involving sequences of paired comparisons, and much earlier related work exists in the biometrics literature (e.g., Bliss 1935). Louviere and Woodworth (1983) provided design constructions for multiple choice options under the assumption of McFadden's (1974) conditional logit model.

Following this early work, only sporadic progress was made on designs for multiple choice problems until the early 2000s, when new developments began to proliferate. Recent work includes Sándor and Wedel (2002), Kanninen (2002), Burgess and Street (2003), Carlsson and Martinsson (2003), Street and Burgess (2004a, b), Lusk and Norwood (2005), Sándor and Wedel (2005), Street et al. (2005), Kessels et al. (2006), Raghavarao and Wiley (2006), Ferrini and Scarpa (2007), Street and Burgess (2007), and Scarpa and Rose (2008), and this list is by no means exhaustive. Clearly, interest in theory and application of DCEs is growing rapidly.

Despite this progress, debate has grown in applied economics, marketing, and transportation over issues related to the design of 'optimal' choice experiments. Much of this debate, however, can be seen as potentially misleading for applied researchers, particularly because it is largely about different assumptions that researchers make to derive design results. While such assumptions are often explicitly stated in technical terms, their implications for applied researchers are often not clearly communicated and applied researchers often lack the technical ability and inclination to wade through the fine details of the technical papers. The bottom line of these is frequently stated in terms of the efficiency of the design matrix or the sample size needed for a given level of precision in estimates of the model parameters if the assumptions hold. While this is the question for which applied researchers often think they want an answer, it rarely is the only question that they should seek to have answered, and often not the most important one.

There are two major problems with such statements of which we ourselves are guilty. First, little is known about when a set of assumptions may be empirically valid for a given application context¹ nor how robust the claims of optimality are to likely violations of the maintained assumptions.² Second, researchers are frequently not directly interested in model parameters and their associated design based covariance matrices but rather in quantities derived from them like willingness-to-pay for a change in attribute levels. The efficiency properties of such quantities may be very different, and it has long been clear in the statistics literature that the properties of the model parameters do not carry over to non-linear functions of multiple parameters.³ Thus, the debate has high potential to confuse and

¹ There are occasionally signs of interest in this topic starting with Kuhfeld (1994) and largely limited to the marketing literature. See Louviere et al. (2008) for a recent example.

² Some papers such as Sándor and Wedel (2001) and Ferrini and Scarpa (2007) examined the performance of their proposed designs in the face of different types of misspecification. More work along this line and, in particular, work that examines the performance of different designs across a range of likely situations would be helpful. See Scarpa and Rose (2008) for one such example.

³ It suffices here to note that a d-optimal design for model parameters that has maximal efficiency may result a much larger confidence interval for ratio of two of the model's parameters than would be

mislead applied researchers who seek only to identify the best way to approach a particular problem they wish to solve.

Similar debates have been evident for many years in the statistics literature, where there often is a split between researchers willing to make strong assumptions and those only willing to entertain the weakest of assumptions. A variant of the debate is a classicist versus Bayesian debate, which continues unabated with little resolution except for the development of computational techniques that make the latter approach more tractable. However, a key difference between the literature on optimal design theory in statistics (e.g., Silvey 1980; Atkinson and Donev 1992; Pukelsheim 2006) and the DCE literature, is that the assumptions made in order to derive optimal design results are usually more clearly stated as papers in statistics tend to start with a set of assumptions (not necessarily realistic) and derive results given that set of assumptions.⁴ Robustness issues to a range of misspecification issues also receive more attention (Sitter 1992; Adewale and Wiens 2009).⁵ We suggest that greater familiarity with this literature should lead to improvements in clarity of communication of optimal design results in the economics, marketing and transportation literature.⁶ Further, it will help researchers avoid reinventing the wheel and later having a strong sense of *déjà vu* over some claims being made.

Reinventing wheels is not always bad, especially if it gives fresh perspectives instead of repetition of long-acknowledged mistakes. Unfortunately, recent literature in marketing, economics and transportation research tends to suggest the latter, with some misleading claims being made, such as random designs being close to optimal (e.g., Lusk and Norwood 2005; see also the commentary by Carson et al. 2009). In this case, the properties of randomness render such claims dubious at best. Unfortunately, this is not an isolated incident. One of the best known papers in the marketing literature on design for choice experiments (Huber and Zwerina 1996) advances utility balance as a desirable criterion (making the alternatives equally attractive), which on the surface may seem like a useful thing to do. However, it is easy to show that this criterion for any fixed prior about the

obtained under a *c*-optimal design that directly minimized the ratio's confidence interval under the assumption (for both design criteria) that the true parameter values were used in deriving the designs. The *c*-optimal design though is known to be quite sensitive to the parameter estimates assumed while the *d*-optimal design is more robust. Alberini and Carson (1993) provide an extensive discussion of this issue in the context of a binary choice with a randomly assigned price variable. Scarpa and Rose (2008) recently examined the role of design on confidence intervals for WTP in a DCE context.

⁴ Again, though the implications of these assumptions may not be clearly recognized by applied researchers. For example, Street and Burgess (2007) use ortho-normal coding in deriving their designs rather than the more standard effects coding which has implications for efficiency unless the ortho-normal coding is also used in the actual estimation.

⁵ The difference with the statistical work along this line is not so much that it examines the robustness of the results to some alternative assumptions, as a number of design papers in applied microeconomic fields do, but rather that these papers start from the perspective that the assumptions behind the model to be estimated are likely to be wrong.

⁶ One sometimes hears the argument that the work being done by statisticians is irrelevant because that work is based on linear models. Nothing could be further from the truth. From the early days of R.A. Fisher, statisticians have looked at experimental design in models with discrete outcomes, the bioassay dose-response experiment being the classic example. What is true is that the terminology used is often different (see for instance the Stata 9 manual entry for *clogit* which notes: *clogit* fits what biostatisticians and epidemiologists call conditional logistic regression for matched case-control groups and what economists and other social scientists call fixed-effects logit for panel data. It also fits McFadden's choice model. Computationally, these models are exactly the same. Results in the statistics literature on non-linear models (e.g., Atkinson and Haines 1996) are more likely to be developed for classes of models and particular the class of (non-linear) generalized linear models that characterize many assumed choice models rather than specific models.

model parameters has the opposite effect if those parameters are true.⁷ The paper's results showing the usefulness of the utility balance approach are simply based on one prior being false and utility moving the design away from what would be optimal if that prior was true toward a design that would be more efficient under an alternative (true) prior. However, if one had faith in the alternative prior for the parameters, a much more efficient design could be obtained by using that prior to derive an optimal design (under the desired criteria) ignoring utility balance. Of course, superiority of any particular optimal design result can only be assessed relative to satisfaction of a set of maintained assumptions about associated empirical processes under study.

We believe that it would be desirable if more of the papers putting forth a particular approach to DCEs, and particularly those claiming to make a major advance, appeared in the traditional design literature in statistics as the advance is likely to be relevant to multiple applied fields.⁸ We also take this view because the number of applied researchers working with DCEs is sufficiently large that the statistics community (broadly defined) should now be interested in such problems and because it would be a sign of the larger acceptance of the importance of DCE design issues. At the same time, few choice modellers, including many with strong backgrounds in econometrics, seem sufficiently knowledgeable about the larger experimental design literature to have the broad perspective necessary to assess the correctness of claims made, and to determine if papers clearly advance the existing literature.⁹ While journal editors in applied economics, marketing, and transportation are probably not likely to follow our suggestion, we urge them to insist on a substantial fraction of referees on any optimal design paper being drawn from the broader statistical design community, while continuing to ensure that the papers they publish are useful to applied researchers in those fields. This is a difficult balancing act.

We do not mean to imply there is not a place in applied economics, marketing, and transportation journals for papers dealing with the design of choice experiments. Instead, we propose that it would be useful for these journals to recognize and encourage some streams of work of particular interest to applied researchers. For example, it would be helpful to have more research directed towards the empirical validity of different assumptions often made about underlying data generating processes, as these are field-specific (e.g., Louviere et al. 2008).

Comparisons of how *different* design construction approaches perform on real datasets obtained using otherwise comparable data collection procedures would be welcome. Likewise, more research on design approaches that can support comparisons of several models would be desirable because it is common for researchers to want to compare competing models estimated from the same DCE data source.¹⁰ Of particular note is the widespread practice of in applied work of constructing designs based on the assumption that a simple conditional logit model represents the choice data generating process, when this model form rarely seems to be empirically correct, as evidenced in many papers devoted to

⁷ This can be seen in the original paper as the design at the prior that all of the coefficients are zero becomes less efficient with utility balance (under a different prior about the coefficients). Interestingly, this paper makes important enduring contributions with respect to other aspects of DCE design.

⁸ The issue here is one of both helping to ensure that the major contributions are technically correct and that the potential usefulness of the contribution across a range of applied fields is realized. Hopefully, the *Journal of Choice Modeling* can serve both of these functions.

⁹ This is, of course, some truth in the opposite generalization: many theoretically oriented statisticians are more interested elegant general mathematical formulation than in helping to solve a specific, and even common, design problem in a specific field.

¹⁰ Again there is an old and now fairly large statistics literature on this topic (e.g., Box and Hill 1967; Atkinson and Federvov 1975; Federvov and Khabarov 1986).

random parameters and latent class choice models.¹¹ Ideally this will help move the field in the direction of more robust approaches. Quests for statistical efficiency also have overshadowed several issues related to design of DCEs that may matter more to applied researchers, including among others, model identification, statistical power, choice of levels to represent attribute variations and task complexity.

In the case of identification, researchers often want to estimate the parameters of a broader class of models than a simple main effects specification and to compare models, some that may be nested and some that may not. Many design approaches seem to satisfy identification for strictly additive indirect utility functions, but pay virtually no attention to other plausible specifications. This is problematic because it is unclear when strictly additive indirect utility models are empirically correct. For example, the literature on composition rules in psychology suggests the existence of distributive and multiplicative rules in addition to the common additive specifications (e.g., see Krantz and Tversky 1971). Further, it often is impossible to reject restrictive forms like strictly additive indirect utility specifications in favour of more general forms due to identification restrictions inherent in design strategies that confound included model effects with omitted higher order effects required for non-additive specifications.¹²

Statistical power is another under-appreciated aspect of DCEs. Studies are often published where one cannot reject large differences in competing models. Given our discussion on the need for more testing of the appropriateness of the assumptions underlying optimal design results, an improved ability to reject a false null hypothesis would be extremely useful for applied researchers. In the context of DCEs, power is an issue closely tied to the variability of underlying data generating processes, and hence, is an excellent topic for applied economics, marketing, and transportation researchers. Robustness of sample size calculations to violations of the underlying assumptions also is an area just starting to be studied in the DCE literature (e.g., Bliemer and Rose 2009), and is one where much more work could be fruitfully done.

In the case of attribute descriptions, the levels used in DCEs have a large bearing on policies and quantities of interest that can be studied. A common and often ignored issue is that due to perceptual or technical reasons, attribute levels often are correlated across attributes, which has substantive implications for types of designs that should be used. It is well-known in the biometrics dose-response and binary discrete choice contingent valuation literature (e.g., Alberini and Carson 1993) that choice of levels can have large impacts on precision of statistics of interest, such as willingness-to-pay. To minimize task complexity, researchers often omit attributes that are not important to most of the population but which may have large effects for some substantial segment or restrict task sizes in other ways, such as limiting numbers of choice options or choice sets. Applied researchers realize that “simplifying” choice tasks is not without consequence but often “feel” like this is the lesser of two evils. However, there is now much evidence accumulated over the past 20 years that most participants in DCEs can and will do larger and more complex tasks with little loss of reliability or validity (e.g., Louviere et al. 2008). We suggest that more research is needed to test the effects of different levels of task complexity on discrete choice responses. Broad statements about respondents reaching “cognitive limits” need more empirical tests within a DCE context rather than relying on tangentially related work from other fields.

¹¹ It is good to see some work on formal designs for non-MNL models (e.g., Sándor and Wedel 2002; 2005; Ferrini and Scarpa 2007; Bliemer et al. 2008; Bliemer and Rose 2010).

¹² While work on efficient designs in the face of interactions has long played a role in the large literature involving linear models, such work is in its infancy for DCE. See Yu et al. (2006) for one recent exploration.

In summary, we suggest that current debates about different optimal design results lack transparency from the perspective of applied researchers. Applied researchers are confused over what design they should use for a particular study they want to do. They are justifiably upset over finding contradictory advice and of having their papers rejected on the grounds of having used a faulty design strategy. Greater clarity in the assumptions made to derive theoretical results can improve applied researchers' ability to make informed choices of appropriate experimental designs for their problems and more effort needs to be devoted to showing how specific design results relate to classes of applied problems. Much progress can be made if there is general acknowledgement that greater efficiency always will be achieved if the underlying empirical assumptions are correct, but many optimal design results (read 'set of assumptions') are unlikely to be robust to plausible departures from these maintained empirical assumptions. Applied researchers need to be encouraged to make design choices that are both suitable to the typical assumptions made about the model they believe appropriate and robust to plausible alternatives.

It is extremely gratifying to see the current interest in DCE design issues and the amount of good work that is going on. We believe that there is a tremendous opportunity for the field to make a concerted effort to move forward by focusing on common problems such as ignorance of true, underlying data generation processes. In doing so, we need to be careful not to sow more confusion for the applied researchers that we are endeavouring to assist than we help to resolve.

References

- Abelson, R. M. and R. A. Bradley. 1954. A 2 x 2 factorial with paired comparisons. *Biometrics* 10, 487-502.
- Adewale, A. J. and D. P. Weins. 2009. Robust designs for misspecified logistic models. *Journal of Statistical Planning and Inference*, 139, 3-15.
- Alberini, A. and R. T., Carson. 1993. The efficiency of binary discrete choice estimation. Department of Economics, Department of Economics, University of California, San Diego, Discussion Paper 90-34R.
- Atkinson, A. C. and A.N., Donev. 1992. *Optimal Experimental Designs*. Chalendon Press, Oxford.
- Atkinson, A. C. and V. V. Federov. 1975. The design of experiments for discriminating between two rival models. *Biometrika*, 62 57-70.
- Atkinson, A.C. and L.M., Haines. 1996. Designs for nonlinear and generalized linear models. In S. Gosh and C.R. Rao, eds. North-Holland, Amsterdam.
- Bliemer, M.C.J. and J. M. and Rose. 2009. Efficiency and Sample Size Requirements For Stated Choice Experiments, *Transportation Research Board Annual Meeting*, Washington DC January.
- Bliemer, M. C. J. and J. M., Rose. 2010. Construction of experimental designs for mixed logit models allowing for correlation across choice observations. *Transportation Research Part B: Methodological*, 44: 720-734.
- Bliemer, M. C. J., J. M., Rose and D. A. Hensher. 2008. Efficient stated choice models for estimating nested logit models. *Transportation Research Part B: Methodology*, 43: 19-35.
- Bliss, C. J. 1935. The calculation of dosage mortality curve. *Annals of Applied Biology*, 22, 134-167.

- Box, G. E. P. and W. J., Hill. 1967. Discrimination among mechanistic models. *Technometrics*, 9, 57-71.
- Bradley, R. A. and T. El-Helbawy. 1976. Treatment contrasts in paired comparisons: basic procedures with application to factorials. *Biometrika*, 63, 255-262.
- Bradley, R.A. and M. E., Terry. 1952. The rank analysis of incomplete block designs. I. The method of paired comparisons. *Biometrika*, 39, 324-345.
- Burgess, L. and D. J., Street. 2003. Optimal designs for 2^k choice experiments. *Communications in Statistics: Theory and Methods*, 32, 11, 2185-2206.
- Carson, R.T., Louviere, J.J. and N. Wasi. 2009. "A Cautionary Note on Designing Discrete Choice Experiments: A Comment on Lusk and Norwood's 'Effect of Experiment Design on Choice-Based Conjoint Valuation Estimates,'" *American Journal of Agricultural Economics*, 91, 4, 1056-1063.
- Carlsson, F. and P., Martinsson. 2003. Design techniques for stated preference methods in health economics. *Health Economics*, 12, 281-294.
- David, H. A. 1988. *The Method of Paired Comparisons*, 2nd ed. Lubrecht & Cramer, Port Jervis, NJ.
- Davidson, R.R. and P. H., Farquhar. 1976. A bibliography on the method of paired comparisons, *Biometrika*, 32, 241-252
- Federov, V. V. and V. Kabarov. 1986. Duality of optimal designs for model discrimination and parameter estimation. *Biometrika*, 73, 183-190.
- Ferrini, S. and R. Scarpa. 2007. Designs with a priori information for nonmarket valuation with choice experiments: a Monte Carlo study. *Journal of Environmental Economics and Management*, 53, 342-363.
- Graßhoff, U. and R. Schwabe. 2008. Optimal design for the Bradley-Terry paired comparison model. *Statistical Methods and Applications*, 17, 275-289.
- Huber, J. and K., Zwerwina. 1996. The importance of utility balance in efficient choice designs. *Journal of Marketing Research*, 33, 307-317.
- Kanninen, B.J. 2002. Optimal design for multinomial choice experiments. *Journal of Marketing Research*, 39, 214-227.
- Kessels, R., P. Goos, and M. Vandebroek. 2006. A comparison of criteria to design efficient choice experiments. *Journal of Marketing Research*, 43, 409-419.
- Kuhfeld, W., R. D., Tobias, M., Garatt. 1994. Efficient experimental design with marketing research applications. *Journal of Marketing Research*, 31: 545-557.
- Krantz, D. H. and A. Tversky. 1971. Conjoint measurement analysis of composition rules in psychology. *Psychological Review*, 78, 151-169.
- Louviere, J. J. and G., Woodworth. 1983. Design and analysis of simulated simulated choice and allocation experiments. *Journal of Marketing Research*, 20, 350-367.
- Louviere, J. J., T., Islam., N., Wasi, D. Street and L., Burgess. 2008. Designing Discrete Choice Experiments: Do Optimal Designs Come at a Price? *Journal of Consumer Research*, 35, 2, 360-375.
- Lusk, J.L. and F. B., Norwood. 2005. Effect of experimental design on choice-based conjoint valuation estimates. *American Journal of Agricultural Economics*, 87, 771-785.
- Raghavarao, D. and J. B, Wiley. 2006. Design strategies for sequential choice experiments involving economic alternatives. *Journal of Statistical Planning and Inference*, 136, 3287-3306.
- McFadden, D. 1974. Conditional logit analysis of qualitative choice behaviour. In: Zarembka, P. (Ed.), *Frontiers in Econometrics*. Academic Press, New York.
- Pukelsheim, F. 2006. *Optimal Design of Experiments*. SIAM, Philadelphia.
- Sándor, Z. and M., Wedel. 2001. Designing conjoint experiments using manager's prior beliefs. *Journal of Marketing Research*, 38, 430-444.

- Sándor, Z. and M., Wedel. 2002. Profile Construction in Experimental Choice Designs for Mixed Logit Models, *Marketing Science*, 21, 4, 455-475.
- Sándor, Z. and M., Wedel. 2005. Heterogeneous Conjoint Choice Designs. *Journal of Marketing Research*, 42, 2, 210-218.
- Scarpa, R. And Rose, J. M. 2008. Experimental designs for environmental valuation with choice experiments: monte carlo investigation, *Australian Journal of Agricultural and Resource Economics*, 52: 253-282.
- Silvey, S.D. 1980. Optimal Design. Chapman-Hall, London.
- Sitter, R. 1992. Robust designs for binary data. *Biometrics*, 48, 1145-1155.
- Street, D. J. and L. Burgess, L. 2007. The Construction of Optimal Stated Choice Experiments: Theory and Methods. Wiley, New York.
- Street, D. J. and L., Burgess. 2004a. Optimal and near-optimal pairs for the estimation of effects in 2-level choice experiments. *Journal of Statistical Planning and Inference* 118, 185-199.
- Street, D.J. and L., Burgess. 2004b. Optimal stated preference choice experiments when all choice sets contain a specific option. *Statistical Methodology*, 1, 37-45.
- Street, D.J., L., Burgess, and J. J., Louviere. 2005. Quick and easy choice sets: Constructing optimal and nearly optimal stated choice experiments. *International Journal of Research in Marketing*, 22, 459-470.
- Thurstone, L.L. 1927. The law of comparative judgement. *Psychological Review* 34, 273-286.
- Yu, J., P., Goos, M., Vandeboek. 2006. The importance of attribute interactions in conjoint design and modelling. Working paper 0601, Department of Decision Sciences and Information Management, Katholieke Universiteit Louven.