

9 Things Clients Get Wrong about Conjoint Analysis

Chris Chapman
Google
cchapman@google.com

Reprint from:

Chapman, C. (2013). 9 things clients get wrong about conjoint analysis. In B. Orme, ed. (2013). Proceedings of the 2013 Sawtooth Software Conference, Dana Point, CA, October 2013.

Abstract

This paper reflects on observations from over 100 conjoint analysis projects across the industry and multiple companies that I have observed, conducted, or informed. I suggest that clients often misunderstand the results of conjoint analysis (CA) and that the many successes of CA may have created unrealistic expectations about what it can deliver in a single study. I describe some common points of misunderstanding about preference share, feature assessment, average utilities, and pricing. Then I suggest how we might make better use of distribution information from hierarchical Bayes (HB) estimation and how we might use multiple samples and studies to inform client needs.

Introduction

Decades of results from the marketing research community demonstrate that conjoint analysis (CA) is an effective tool to inform strategic and tactical marketing decisions. CA can be used to gauge consumer interest in products and to inform estimates of feature interest, brand equity, product demand, and price sensitivity. In many well-conducted studies, analysts have demonstrated success using CA to predict market share and to determine strategic product line needs¹.

However, the successes of CA also raise clients' expectations to levels that can be excessively optimistic. CA is widely taught in MBA courses and a new marketer in industry is likely soon to encounter CA success stories and business questions where CA seems appropriate. This is great news ... if CA is practiced appropriately. The apparent ease of designing, fielding, and analyzing a CA study presents many opportunities for analysts and clients to make mistakes.

In this paper, I describe some misunderstandings that I've observed in conducting and consulting on more than 100 CA projects. Some of these come from projects I've fielded while others have been observed in consultation with others; none is exemplary of any particular firm. Rather, the set of cases reflects my observations of the field. For each one I describe the problem and how I suggest to rectify it in clients' understanding.

All data presented here are fictional. The data primarily concern an imaginary “designer USB drive” that comprises nominal attributes such as size (e.g., Nano, Full-length), design style, and ordinal

¹ There are too many published successes for CA to list them comprehensively. For a start, see papers in this and other volumes of the *Proceedings of the Sawtooth Software Conference*. Published cases where this author contributed used CA to inform strategic analysis using game theory (Chapman & Love, 2012), to search for optimum product portfolios (Chapman & Alford, 2010), and to predict market share (Chapman, Alford, Johnson, Lahav, & Weidemann, 2009). This author also helped compile evidence of CA reliability and validity (Chapman, Alford, & Love, 2009).

attributes of capacity (e.g., 32 GB) and price. The data were derived by designing a choice-based conjoint analysis survey, having simulated respondents make choices, and estimating the utilities using hierarchical Bayes multinomial logit estimation. For full details, refer to the source of the data: simulation and example code given in the R code “Rcbc” (Chapman, Alford, and Ellis, 2013; available from this author).

The data here were *not* designed to illustrate problems; rather, they come from didactic R code. It just happens that those data – like data in most CA projects – are misinterpretable in all the common ways.

Mistake #1: Conjoint Analysis Directly Tells Us How Many People Will Buy This Product

A simple client misunderstanding is that CA directly estimates how many consumers will purchase a product. It is simple to use part worth utilities to estimate preference share and interpret this as “market share.” Table 1 demonstrates this using the multinomial logit formula for aggregate share between two products. In practice, one might use individual-level utilities in a market simulator such as Sawtooth Software SMRT, but the result is conceptually the same.

Table 1: Example Preference Share Calculation

	Sum of utilities	Exponentiated	Share of total
Product 1	1.0	2.72	62%
Product 2	0.5	1.65	38%
Total	--	4.37	

As most research practitioners know but many clients don't (or forget), the problem is this: preference share is only partially indicative of real market results. Preference share is an important input to a marketing model, yet is only one input among many. Analysts and clients need to determine that the CA model is complete and appropriate (i.e., valid for the market) and that other influences are modeled, such as awareness, promotion, channel effects, competitive response, and perhaps most importantly, the impact of the outside good (in other words, that customers could choose *none of the above* and spend money elsewhere).

I suspect this misunderstanding arises from three sources. First, clients very much want CA to predict share! Second, CA is often given credit for predicting market share even when CA was in fact just one part of a more complex model that mapped CA preference to the market. Third, analysts' standard practice is to talk about “*market simulation*” instead of “*relative preference simulation*.”

Instead of claiming to predict market share, I tell clients this: *conjoint analysis assesses how many respondents prefer each product, relative to the tested alternatives*. If we iterate studies, know that we're assessing the right things, calibrate to the market, and include other effects, we will get progressively better estimates of the likely market response. CA is a fundamental part of that, yet only one part. Yes, we can predict market share (sometimes)! But an isolated, single-shot CA is not likely to do so very well.

Mistake #2: CA Assesses How Good or Bad a Feature (or Product) Is

The second misunderstanding is similar to the first: clients often believe that the highest part worth indicates a good feature while negative part worths indicate bad ones. Of course, all utilities really tell us is that, given the set of features and levels presented, this is the best fit to a set of observed choices. Utilities don't indicate absolute worth; inclusion of different levels likely would change the utilities.

A related issue is that part worths are relative within a single attribute. We can compare levels of an attribute to one another – for instance, to say that one memory size is preferable to another memory size – but should not directly compare the utilities of levels across attributes (for instance, to say that some memory size level is more or less preferred than some level of color or brand or processor). A common approach to compare levels across attributes is to apply rescaling that standardizes them (such as zero-centered differences in Sawtooth Software). Ultimately, product preference involves full specification across multiple attributes and is tested in a market simulator (I say more about that below).

I tell clients this: *CA assesses tradeoffs among features to be more or less preferred*. It does not assess absolute worth or say anything about untested features.

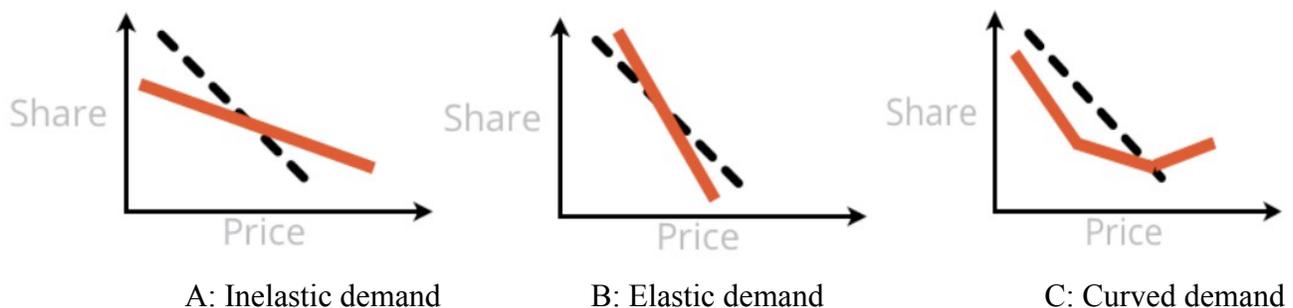
Mistake #3: CA Directly Tells Us Where to Set Prices

Clients and analysts commonly select a CA as a way to assess pricing. What is the right price? How will such-and-such feature affect price? How price sensitive is our audience? All too often, I've seen clients inspect the average part worths for price – often estimated without constraints and as piecewise utilities – and interpret them at face value.

Figure 1 shows three common patterns in price utilities; the dashed line shows scaling in exact inverse proportion to price, while the solid line plots the preference that we might observe from CA (assuming a linear function for patterns A and B, and a piecewise estimation in pattern C, although A and B could just as well be piecewise functions that are monotonically decreasing).

In pattern A, estimated preference share declines more slowly than price (or log price) increases. Clients love this: the implication is to price at the maximum (presumably not to infinity). Unfortunately, real markets rarely work that way; this pattern more likely reflects a method effect where CA underestimates price elasticity.

Figure 1: Common Patterns in Price Utilities



In pattern B, the implication is to price at the minimum. The problem here is that relative preference implies range dependency. This may simply reflect the price range tested, or reflect that respondents are using the survey for communication purposes (“price low!”) rather than to express product preferences.

Pattern C seems to say that some respondents like low prices while others prefer high prices. Clients love this, too! They often ask “how do we reach the price-insensitive customers?” The problem is that there is no good theory as to why price should show such an effect. It is more likely that the CA task was poorly designed or confusing, or that respondents had different goals such as picking their favorite brand or heuristically simplifying the task in order to complete it quickly. Observation of a price reversal as we see here (i.e., preference going up as price goes up in some part of the curve) is more likely an indication of a *problem* than an observation about actual respondent preference!

If pattern C truly does reflect a mixture of populations (elastic and inelastic respondents) then there are higher-order questions about the sample validity and the appropriateness of using pooled data to estimate a single model. In short: pattern C is seductive! Don't believe it unless you have assessed carefully and ruled out the confounds and the more theoretically sound constrained (declining) price utilities.

What I tell clients about price is: *CA provides insight into stated price sensitivity, not exact price points or demand estimates* without a lot more work and careful consideration of models, potentially including assessments that attempt more realistic incentives, such as incentive-aligned conjoint analysis (Ding, 2007). When assessing price, it's advantageous to use multiple methods and/or studies to confirm that answers are consistent.

Mistake #4: The Average Utility is the Best Measure of Interest

I often see – and yes, sometimes even produce – client deliverables with tables or charts of “average utilities” by level. This unfortunately reinforces a common cognitive error: that the *average* is the *best estimate*. Mathematically, of course, the mean of a distribution minimizes some kinds of residuals – but that is rarely how a client interprets an average!

Consider Table 2. Clients interpret this as saying that Black is a much better feature than Tie-dye. Sophisticated ones might ask whether it is statistically significant (“yes”) or compute the preference share for Black (84%). None of that answers the real question: *which is better for the decision at hand?*

Table 2: Average Feature Utilities

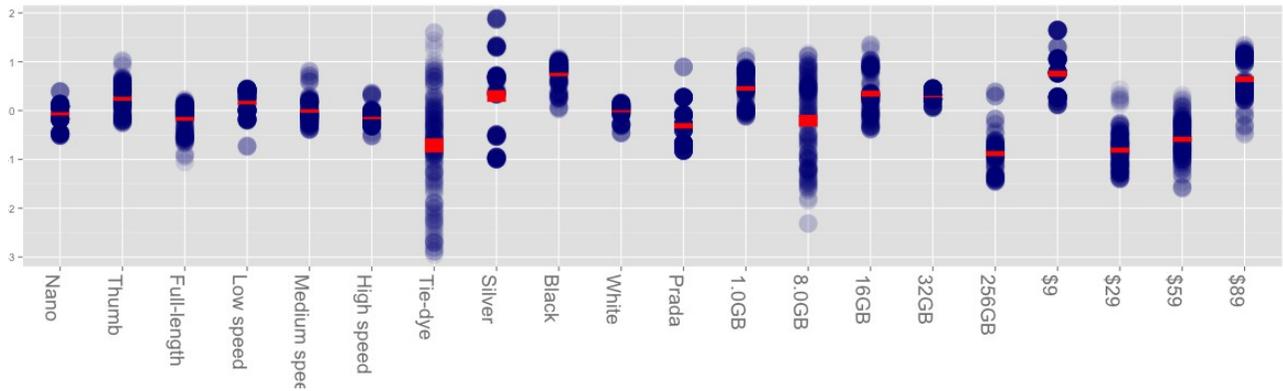
Feature	Average Utility
Black	0.79
Tie-dye	-0.85
...	...

Figure 3 is what I prefer to show clients and presents a very different picture. In examining Black vs. Tie-dye, we see that the individual-level estimates for Black have low variance while Tie-dye has high variance. Black is broadly acceptable, relative to other choices, while Tie-dye is polarizing.

Is one better? That depends on the goal. If we can only make a single product, we might choose Black.

If we want a diverse portfolio with differently appealing products, Tie-dye might fit. If we have a way to reach respondents directly, then Silver might be appealing because a few people strongly prefer it. Ultimately this decision should be made on the basis of market simulation (more on that below), yet understanding the preference structure more fully may help an analyst understand the market and generate hypotheses that otherwise might be overlooked.

Figure 3: Distribution of Individual-Level Utilities from HB Estimation



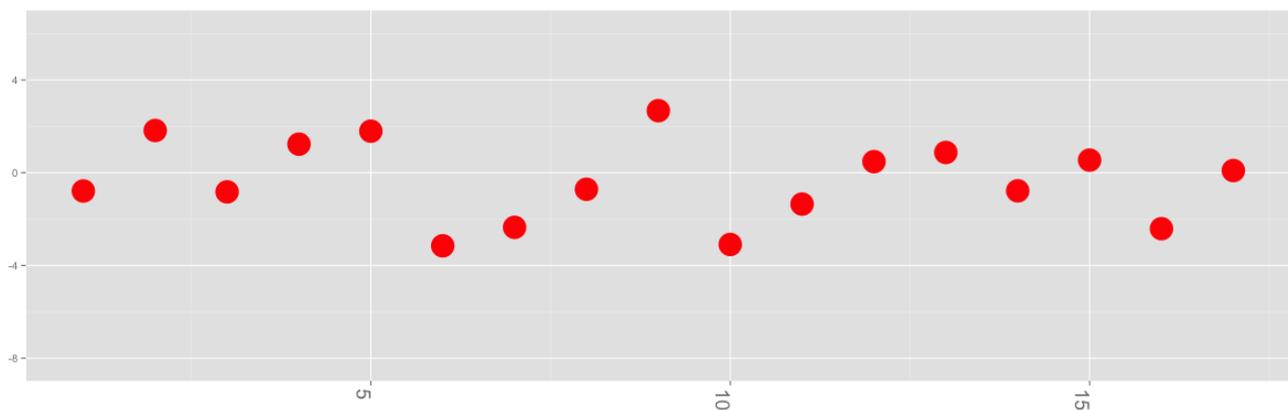
The client takeaway is this: *CA (using HB) gives us a lot more information than just average utility.* We should use that information to have a much better understanding of the distribution of preference.

Mistake #5: There Is a True Score

The issue about average utility (problem #4 above) also arises at the individual level. Consider Figure 4 which presents the mean betas for one respondent. This respondent has low utilities for features 6 and 10 (on the X axis) and high utilities for features 2, 5, and 9.

It is appealing to think that we have a psychic X-ray of this respondent, that there is some “true score” underlying these preferences, as a social scientist might say. There are several problems with this view. One is that behavior is contextually dependent, so any respondent might very well behave differently at another time or in another context (such as a store instead of a survey). Yet even within the context of a CA study, there is another issue: we know much more about the respondent than the average utility!

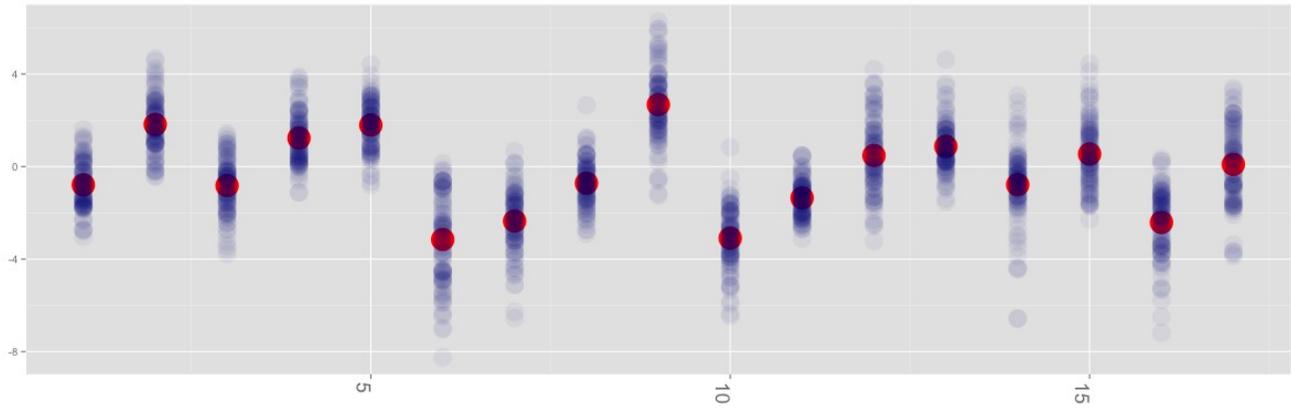
Figure 4: Average Utility by Feature, for One Respondent



Now compare Figure 5 with Figure 4. Figure 5 shows – for the same respondent – the within-

respondent distribution of utility estimates across 100 draws of HB estimates (using Monte Carlo Markov chain, or MCMC estimation). We see significant heterogeneity. An 80% or 95% credible interval on the estimates would find few “significant” differences for this respondent. This is a more robust picture of the respondent, and inclines us away from thinking of him or her as a “type.”

Figure 5: Distribution of HB Beta Estimates by Feature, for the Same Respondent



What I tell clients is this: *understand respondents in terms of tendency rather than type*. Customers behave differently in different contexts and there is uncertainty in CA assessment. The significance of that fact depends on our decisions, business goals, and ability to reach customers.

Mistake #6: CA Tells Us the Best Product to Make (*rather easily*)

Some clients and analysts realize that CA can be used not only to assess preference share and price sensitivity but also to inform a product portfolio. In other words, to answer “What should we make?”

An almost certainly wrong answer would be to make the product with highest utility, because it is unlikely that the most desirable features would be paired with the best brand and lowest price. A more sophisticated answer searches for preference tradeoff vs. cost in the context of a competitive set. However, this method capitalizes on error and precise specification of the competitive sets; it does not examine the sensitivity and generality of the result.

Better results may come by searching for a large set of near-optimum products and examine their commonalities (Chapman and Alford, 2010; cf. Belloni, et al, 2008). Another approach, depending on the business question, would be to examine likely competitive response to a decision using a strategic modeling approach (Chapman and Love, 2012). An analyst could combine the approaches: investigate a set of many potential near-optimal products, choose a set of products that is feasible, and then investigate how competition might respond to that line.

Doing this is a complex process: it requires extraordinarily high confidence in one's data, and then one must address crucial model assumptions and adapt (or develop) custom code in R or some other language to estimate the models (Chapman and Alford, 2010; Chapman and Love, 2012). The results can be extremely informative – for instance, a product identified in Chapman and Alford (2010) was identified by the model fully 17 months in advance of its introduction to the market by a competitor – but arriving at such an outcome is a complex undertaking built on impeccable data (and perhaps luck).

In short, when clients wish to find the “best product,” I explain: *CA informs us about our line, but*

Mistake #8: Make CA Fit What You Want to Know

To address tough business questions, it's a good idea to collect customer data with a method like CA. Unfortunately, this may yield surveys that are more meaningful to the client than the respondent.

I find this often occurs with complex technical features (that customers may not understand) and messaging statements (that may not influence CA survey behavior). Figure 7 presents a fictional CBC task about wine preferences. It was inspired by a poorly designed survey I once took about home improvement products; I selected wine as the example because it makes the issue particularly obvious.

Figure 7: A CBC about Wine

Imagine you are selecting a bottle of wine for a special celebration dinner at home.

If the following wines were your only available choices, which would you purchase?

Blend	75% Cabernet Sauvignon 20% Merlot 4% Cabernet Franc 1% Malbec	75% Cabernet Sauvignon 15% Merlot 10% Cabernet Franc
Winery type	Custom crush	Negotiant
Bottle size	700ml	750ml
Cork type	Grade 2	Double disk (1+1)
Fining agent	(None, unfinned)	Potassium caseinate
Bottling line type	Mobile	On premises
Origin of bottle glass	Mexico	China
	○	○

Our fictional marketing manager is hoping to answer questions like these: should we fine our wines (cause them to precipitate sediment before bottling)? Can we consider cheaper bottle sources? Should we invest in an in-house bottling line (instead of truck that moves between facilities)? Can we increase the Cabernet Franc in our blend (for various possible reasons)? And so forth.

Those are all important questions but posing their technical features to customers results in a survey that only a winemaker could answer! A better survey would map the business consideration to features that a consumer can address, such as taste, appearance, aging potential, cost, and critics' scores. (I leave the question of how to design that survey about wine as an exercise for the reader.)

This example is extreme, yet how often do we commit similar mistakes in areas where we are too close to the business? How often do we test something “just to see if it has an effect?” How often do we describe something the way that R&D wants? Or include a message that has little if any real information? And then, when we see a null effect, are we sure that it is because customers don't care, or could it be because the task was bad? (A similar question may be asked in case of significant effects.) And, perhaps most dangerously, how often do we field a CA without doing do a small-sample pretest?

The implication is obvious: *design CA tasks to match what respondents can answer reliably and validly.* And before fielding, pretest the attributes, levels, and tasks to make sure!

(non!-) Mistake #9: It's Better than Using Our Instincts

Clients, stakeholders, managers, and sometime even analysts are known to say, “Those results are interesting but I just don't believe them!” Then an opinion is substituted for the data.

Of course CA is not perfect – all of the above points demonstrate ways in which it may go wrong, and there are many more – but I would wager this: a well-designed, well-fielded CA is almost always better than expert opinion. Opinions of those close to a product are often dramatically incorrect (cf. Gourville, 2004). Unless you have better and more reliable data that contradicts a CA, go with the CA.

If we consider this question in terms of expected payoff, I propose that the situation resembles Figure 8. If we use data, our estimates are likely to be closer to the truth than if we don't. Sometimes they will be wrong, but will not be as wrong on average as opinion would be.

Figure 8: Expected Payoffs with and without Data

	Decision correct	Decision incorrect	<i>Net expectation:</i>
Use data	High precision (high gain)	Low inaccuracy (modest loss)	Positive
Use instinct	Low precision (modest gain)	High inaccuracy (large loss)	Negative

When we get a decision right with data, the relative payoff is much larger. Opinion is sometimes right, but likely to be imprecise; when it is wrong, expert opinion may be disastrously wrong. On the other hand, I have yet to observe a case where consumer data has been terribly misleading; the worst case I've seen is when it signals a need to learn more. When opinion and data disagree, explore more. Do a different study, with a different method and different sampling.

What I tell clients: *it's very risky to bet against what your customers are telling you!* An occasional success – or an excessively successful single opiner – does not disprove the value of data.

Mistake #10 and Counting

Keith Chrzan (2013) commented on this paper after presentation at the Sawtooth Software Conference and noted that attribute importance is another area where there is widespread confusion. Clients often want to know “Which attributes are most important?” but CA can only answer this with regard to the relative utilities of the attributes and features tested. Including (or omitting) a very popular or unpopular level on one attribute will alter the “importance” of every other attribute!

Conclusion

Conjoint analysis is a powerful tool but its power and success also create conditions where client expectations may be too high. We seen that some of the simplest ways to view CA results such as average utilities may be misleading and that despite client enthusiasm they may distract from

answering more precise business questions. The best way to meet high expectations is to meet them! This may require all of us to be more careful in our communications, analyses, and presentations.

The issues here are not principally technical in nature; rather they are about how conjoint analysis is positioned and how expectations are set and upheld through effective study design, analysis, and interpretation. I hope the paper inspires you – and even better, inspires and informs clients.

Acknowledgements

I'd like to thank Bryan Orme, who provided careful, thoughtful, and very helpful feedback at several points to improve both this paper and the conference presentation. If this paper is useful to the reader, that is in large part due to Bryan's suggestions (and if it's *not* useful, that's due to the author!) Keith Chrzan also provided thoughtful observation and reflections during the conference. Finally, I'd like to thank all my colleagues over the years, and who are reflected in the reference list. They spurred the reflections more than anything I did.

References

- Belloni, A., Freund, R.M, Selove, M., and Simester, D. (2008). Optimal product line design: efficient methods and comparisons. *Management Science* 54: 9, September 2008, pp. 1544-1552.
- Chapman, C.N., Alford, J.L., and Ellis, S. (2013). Rcbc: marketing research tools for choice-based conjoint analysis, version 0.201. [R code]
- Chapman, C.N., and Love, E. (2012). Game theory and conjoint analysis: using choice data for strategic decisions. *Proceedings of the 2012 Sawtooth Software Conference*, Orlando, FL, March 2012.
- Chapman, C.N., and Alford, J.L. (2010). Product portfolio evaluation using choice modeling and genetic algorithms. *Proceedings of the 2010 Sawtooth Software Conference*, Newport Beach, CA, October 2010.
- Chapman, C.N., Alford, J.L., Johnson, C., Lahav, M., and Weidemann, R. (2009). Comparing results of CBC and ACBC with real product selection. *Proceedings of the 2009 Sawtooth Software Conference*, Del Ray Beach, FL, March 2009.
- Chapman, C.N., Alford, J.L., and Love, E. (2009). Exploring the reliability and validity of conjoint analysis studies. Presented at Advanced Research Techniques Forum (A/R/T Forum), Whistler, BC, June 2009.
- Chrzan, K. (2013). Remarks on “9 things clients wrong about conjoint analysis.” Discussion at the 2013 Sawtooth Software Conference, Dana Point, CA, October 2013.
- Ding, M. (2007). An incentive-aligned mechanism for conjoint analysis. *Journal of Marketing Research*, 2007, pp. 214-223.
- Gourville, J. (2004). Why customers don't buy: the psychology of new product adoption. Case study series, paper 9-504-056. Harvard Business School, Boston, MA.