Sawtooth Software

RESEARCH PAPER SERIES

A New Approach to Adaptive CBC

Richard M. Johnson & Bryan K. Orme,
Sawtooth Software, Inc.

530 W. Fir St.
Sequim, WA 98382
(360) 681-2300
www.sawtoothsoftware.com

A New Approach to Adaptive CBC

October 2007

Richard M. Johnson
Bryan K. Orme
Sawtooth Software, Inc.

Background

Choice-Based Conjoint (CBC) is the most widely used conjoint technique today. The marketing research community has adopted CBC enthusiastically, for several reasons. Choice tasks seem to mimic what actual buyers do more closely than ranking or rating product concepts as in conventional conjoint analysis. Choice tasks seem easy for respondents, and everyone can make choices. And equally important, multinomial logit analysis provides a well-developed statistical model for estimating respondent partworths from choice data.

However, choice tasks are less informative than tasks involving ranking or rating of product concepts. The respondent must examine the characteristics of several product concepts in a choice set, each described on several attributes, before making a choice. Yet, that choice reveals only which product was preferred, and nothing about strength of preference, or the relative ordering of the non-preferred concepts. Initially, CBC questionnaires of reasonable length offered too little information to support multinomial logit analysis at the individual level. More recently, hierarchical Bayes methods have been developed which do permit individual-level analysis, but interest has remained in ways to design choice tasks so as to provide more information.

Huber and Zwerina (1996) showed that choice tasks are more efficient (statistically) if the alternatives within a task are more nearly equal in utility, giving rise to the term "utility balance." Such choice tasks cannot be designed without knowledge of the respondent's utilities, which is not usually available until after the interview. This "chicken-and-egg" problem has led to several attempts at "adaptive" CBC questionnaires, where inferences from early choice tasks are used in an attempt to create greater utility balance in later choice sets. The authors have participated in three previous attempts to use adaptive principles to produce more efficient choice designs, but without consistent success, two of which were reported at previous Sawtooth Software conferences (Johnson, Huber and Bacon, 2003; Johnson, Huber, and Orme, 2004) and the third attempt reported at the joint Sawtooth Software/SKIM event in Berlin (Johnson, Orme, Huber, and Pinnell, 2005). Those first attempts relied on the assumption that respondents answered in a compensatory manner, consistent with the logit rule. We suspect that we were not more successful because respondents often use non-compensatory decision rules.

In recent years marketing researchers have become aware of potential problems with CBC questionnaires and the way respondents answer CBC questions.

- The concepts presented to respondents are often not very close to the respondent's ideal. This can create the perception that the interview is not very focused or relevant to the respondent.

- Respondents (especially in internet panels) do choice tasks very quickly. According to Sawtooth Software's experience with many CBC datasets, once respondents warm up to the CBC tasks, they typically spend about 12 to 15 seconds per choice task (Johnson and Orme, 1996). For the CBC study presented in this paper, respondents spent about 18 seconds per task (on average across *all* tasks) even when considering 4 alternatives, each specified on 9 attributes. It's hard to imagine how they could evaluate four alternatives each specified on nine attributes in as short a time as 18 seconds (or fewer once warmed up). It seems overwhelmingly likely that respondents accomplish this by simplifying their procedures for making choices, possibly in a way that is not typical of how they would behave if buying a real product.

- To estimate partworths at the individual level, it is necessary for each individual to answer several choice tasks. But when a dozen or more similar choice tasks are presented to the respondent, the experience is often seen to be repetitive and boring, and it seems possible that respondents are less engaged in the process than the researcher might wish.

- If the respondent is keenly intent on a particular level of a critical attribute (a "must have" feature), there is often only one such product available per choice task. Such a respondent is left with selecting this product or "None." And, respondents tend to avoid the "None" constant, perhaps due to "helping behavior." Thus, for respondents intent on just a few key levels, standard minimal overlap choice tasks don't encourage them to reveal their preferences much more deeply than the few "must have" features.

Gilbride *et al.* (2004) and Hauser *et al.* (2006) used sophisticated algorithms to examine patterns of respondent answers, attempting to discover simple rules that can account for respondent choices. Both groups of authors found that respondent choices could be fit by non-compensatory models in which only a few attribute levels are taken into account.

We have done something much simpler, which also suggests that CBC respondents may make choices using simple screening rules:

1. For each respondent, compute a Kendall's Tau coefficient for each attribute level, to measure the relationship between presence of that attribute level and choice of an alternative.

2. Assume that the attribute level with highest Tau is one on which the respondent has screened concepts, and that it accounts for his/her answers for those choice tasks. Remove those choice sets from further consideration.

3. Repeat the process until no choice sets are left. Count the number of attribute levels that are required to account in this way for all of that respondent's choices.

We find that when choice sets are composed so as to have minimal overlap, most respondents make choices consistent with the hypothesis that they pay attention to only a few attribute levels, even when many more are included in product concepts. In a recent study with 9 attributes, 85 percent of respondents' choices could be explained entirely by assuming each respondent paid attention to the presence or absence of at most four attribute levels.

We also examined another CBC data set described in more detail below. This data set had 18 choice tasks, but respondents were given the option of choosing "None." Respondents' answers other than "None" were as follows:

11% answered all 18 tasks by choosing 1 attribute level consistently.
34% answered by choosing at most 2 attribute levels.
80% answered by choosing at most 3 attribute levels.

Such results might lead us to conclude that CBC respondents behave in a way quite different from what we had expected, and contribute less information than we had hoped. And to make matters worse, respondents who apply consistent screening rules involving few attribute levels could easily apply those same rules to holdout choice sets. Thus, success at predicting holdout choices does not imply that respondents are providing informative and thoughtful answers to our questionnaires.

However, the meaning of these results may not be so clear as it appears. A respondent may apply a compensatory model, and yet produce results compatible with a simpler non-compensatory model. To establish this, we used a compensatory model to generate artificial responses to an 18-task CBC questionnaire and then analyzed those responses using the non-compensatory approach. We found that all answers could be accounted for by the hypothesis that the artificial respondent had paid attention to only two of 37 possible attribute levels. Thus, even if a respondent's answers can be explained by a simple non-compensatory model, we cannot be sure that his/her choice process was actually that simple.

Nonetheless, we find these results unsettling. Most CBC respondents answer more quickly than would seem possible if they were giving thoughtful responses with a compensatory model. Most of their answers can be accounted for by very simple screening rules involving few attribute levels. Combine those facts with the realization by anyone who has answered a CBC questionnaire that the experience seems repetitive and boring, and one is led to conclude there is a need for a different way of asking choice questions, with the aim of obtaining better data.

We believe CBC is an effective method that has been of genuine value to marketing researchers, but that it can be improved. And we believe the greatest need at this point is not for better models, but rather for better data.

A New Approach to Data Collection

Like our previous papers on Adaptive CBC, the title for this paper contains the word "Adaptive." However, this time our aim is not to design choice tasks with more statistical efficiency, but rather to acquire better data. We recognize that the respondent may employ screening rules, and we seek to recognize those rules, providing choices among products that pass such screening criteria. In this way we hope to help respondents make choices more thoughtfully, and in a way more like what they would in an actual purchase situation. Our objectives are as follows:

- Provide a more stimulating experience that will encourage more engagement in the interview than conventional CBC questionnaires.

- Mimic actual shopping experiences, which may involve non-compensatory as well as compensatory behavior.

- Screen a wide variety of product concepts, but focus on a subset of most interest to the respondent.

- Provide more information with which to estimate individual partworths than is obtainable from conventional CBC analysis.

The interview has several sections, with each section quite different from the previous (the interview is posted online at www.sawtoothsoftware.com/test/byo/byologn.htm). Throughout the interview we attempt to keep the respondent interested and engaged. The instructions appear on the screen in text, but as though they were spoken by a friendly and attractive female interviewer. Her pictures (images purchased from www.clipart.com) appear frequently at various places in the interview, from different perspectives and in different poses. She explains to the respondent that this is a simulation of a buying experience, and she gives a rationale for each interview section. For example, here is an introductory screen:

Now I'm going to present laptops to you as if you were visiting a store and I were the salesperson assisting you. We have many available configurations to choose from, and it's my job to help you find the right laptop computer.

To help you find the one that best suits you, I'd first like to ask you about the laptop you'd be most likely to purchase.

Next

BYO Section:

In the first section of the interview the respondent answers a "Build Your Own" (BYO) questionnaire to introduce the attributes and levels, as well as to let the respondent indicate the preferred level for each attribute, taking into account any corresponding feature-dependent prices[1]. A typical screen for this section of the interview is shown below:

Feature	Select Feature	Cost for Feature	Help on this item
Screen Size and Weight	15 inch screen, 6 pounds $750	$ 750	
Brand	Dell +$0	$ 0	
Processor Speed (Intel Pentium)	Intel Core 2 Duo T7400 (2.16GHz) + $300	$ 300	?
Operating System	Vista Home Premium +$50	$ 50	?
Memory	1 GB +$100	$ 100	?
Hard Drive	100GB +$50	$ 50	?
Video Card	Select Feature 80GB +$0 100GB +$50 120GB +$100 160GB +$150	$ 0	?
Battery		$ 0	?
Office Software	Select Feature	$ 0	?
	Total:	$ 1250	

[1] We should note that our approach should also be able to accommodate projects for which some attributes do not involve price changes from the base product, or for projects that do not include price at all.

Past research has shown that respondents enjoy BYO questionnaires and answer them rapidly, and that the resulting choices have lower error levels than repetitive choices from CBC questionnaires (Johnson, Orme, and Pinnell, 2006).

Based on answers to the BYO questionnaire, we create a pool of product concepts that includes every attribute level, but for which attribute levels are relatively concentrated around the respondent's preferred attribute levels. Each concept in the pool is generated by altering 2, 3, or 4 attributes from the BYO-specified concept. These concepts are constructed so as to represent a nearly orthogonal design. For this study, we experimented with pools of 40 and 50 concepts.

Screening Section:

In the second section of the interview the respondent answers "screening" questions, where product concepts are shown a few at a time (we have used 5 at a time). Prices are determined by summing the costs of the features involved in the concept (per the BYO exercise) plus or minus 7% or 20%, and rounded to the nearest $50. In the Screening Section, the respondent is not asked to make final choices, but rather just to indicate whether he/she would consider each one "a possibility." We suggest that he/she narrow down the range of possibilities by retaining about half of them, but the number retained is left to the respondent. A typical screen from this section of the interview is shown below:

Here are a few laptops you might like. Do any of these look like they are possibilities? It's helpful if you can keep about half of them for further consideration. But, it's up to you.

Size	14 inch screen, 5 lbs.	17 inch screen, 8 lbs.	15 inch screen, 6 lbs.	15 inch screen, 6 lbs.	15 inch screen, 6 lbs.
Brand	HP	Dell	Acer	Dell	Dell
Processor	Intel Core 2 Duo T7200 (2.00GHz)	Intel Core 2 Duo T7400 (2.16GHz)	Intel Core 2 Duo T7400 (2.16GHz)	Intel Core 2 Duo T7600 (2.33GHz)	Intel Core 2 Duo T7400 (2.16GHz)
Operating System	Vista Home Premium	Vista Ultimate	Vista Home Premium	Vista Home Basic	Vista Home Premium
Memory	1 GB	1 GB	2 GB	1 GB	1 GB
Hard Drive	100 GB	120 GB	160 GB	100 GB	100 GB
Video Card	128 MB Video card, adequate for most use	128 MB Video card, adequate for most use	128 MB Video card, adequate for most use	256 MB Video card for high-speed gaming	128 MB Video card, adequate for most use
Battery	4 hours	3 hours	3 hours	3 hours	6 hours
Productivity Software	Microsoft Office Basic (Word, Excel, Outlook)	Microsoft Office Basic (Word, Excel, Outlook)	Microsoft Office Basic (Word, Excel, Outlook)	Microsoft Office Basic (Word, Excel, Outlook)	Microsoft Works
Price	$1,150	$2,200	$1,800	$1,650	$1,200
	○ A possibility ○ Won't work for me	○ A possibility ○ Won't work for me	○ A possibility ○ Won't work for me	○ A possibility ○ Won't work for me	○ A possibility ○ Won't work for me

Must Haves: After each group of concepts has been presented, we scan previous answers to see if there is any evidence that the respondent is using non-compensatory screening rules. For example, we might notice that he/she has expressed interest in only one level of some attribute, in which case we ask whether that level is an absolute requirement (a "Must Have"). Here is a typical screen for this question:

I don't want to jump to conclusions, but I've noticed that you've selected laptops with certain characteristics, shown below. If any of these describe what you **absolutely need**, it would be helpful to know.

If you'd like, please check the **one most important rule**, and I'll only show you laptops with that feature.

- ○ At least: Intel Core 2 Duo T7200 (2.00GHz)
- ○ At least: Vista Home Premium
- ◉ At least: 1 GB memory
- ○ At least: 100 GB hard drive
- ○ At least: Microsoft Office Small Business (Basic + PowerPoint, Publisher)

- ○ None of the rules above are absolute requirements for my laptop.

Next

Past research with ACA has suggested that respondents are quick to mark many levels as unacceptable that are probably just undesirable. We considered that the same tendency might apply to "must have" rules. To avoid this possibility, we offer only cutoff rules consistent with the respondent's previous choices and we allow the respondent to select only one cutoff rule on this screen. After each new screen of five products has been evaluated, the respondent has another opportunity to add a subsequent cutoff rule.

Unacceptables: If the respondent has systematically avoided an attribute level, we ask whether that level would be completely unacceptable ("Unacceptables)". If the respondent identifies any "must have" or "must avoid" levels, then all further concepts shown will satisfy those requirements. The respondent has several opportunities to express such decision rules, with the result that the number of concepts actually presented to him/her is usually reduced. For this study, respondents needed to evaluate an average of 32 of the 40 product concepts (an average of 8 were automatically screened out due to confirmed decision rules).

Choice Tasks Section:

In the third section of the interview the respondent is shown a series of choice tasks presenting the surviving product concepts (those marked as "possibilities") in groups of three, as in the screen below. In this questionnaire we asked for best and worst in each task, but it would also be possible to ask just for first choices.

Among these three, which is the best option? Which is the worst?
(I've grayed out any features that are the same, so you can just focus on the differences.)

(3 of 7)

Size	15 inch screen, 6 lbs.	15 inch screen, 6 lbs.	17 inch screen, 8 lbs.
Brand	Dell	Dell	Dell
Processor	Intel Core 2 Duo T7600 (2.33GHz)	Intel Core 2 Duo T7400 (2.16GHz)	Intel Core 2 Duo T7200 (2.00GHz)
Operating System	Vista Home Premium	Vista Home Premium	Vista Ultimate
Memory	4 GB	2 GB	1 GB
Hard Drive	100 GB	120 GB	160 GB
Video Card	128 MB Video card, adequate for most use	128 MB Video card, adequate for most use	128 MB Video card, adequate for most use
Battery	3 hours	3 hours	3 hours
Productivity Software	Microsoft Office Small Business (Basic + PowerPoint, Publisher)	Microsoft Office Small Business (Basic + PowerPoint, Publisher)	Microsoft Office Small Business (Basic + PowerPoint, Publisher)
Price	$1,700	$1,650	$1,450
Best	○	○	○
Worst	○	○	○

Next

At this point, respondents are evaluating concepts that are close to their BYO-specified product, that they consider "possibilities," and that strictly conform to any cutoff (must have/unacceptable) rules. To facilitate information processing, we gray out any attributes that are tied across the concepts, leaving respondents to focus on the remaining differences. Any tied attributes are typically the most key factors (based on already established cutoff rules), and thus the respondent is encouraged to further discriminate among the products on the features of secondary importance.

The winning concepts from each triple then compete in subsequent rounds of the tournament until the preferred concept is identified.

Calibration Section (Optional):
The fourth section of the interview may be used to estimate a "None" parameter for the respondent. The section is introduced with this screen:

We are just about done! Now, I'd like to ask you a *different* question: **Would you actually purchase a laptop like those I've been showing you?** I'll ask you about just **5** laptops.

First, I'll show the laptop you originally configured. Next, I'll ask about other laptops you said were possibilities.

Last, I'll ask if you'd actually purchase the laptop you selected as best among those I showed you. I'm thinking you might like this last laptop best.

Next

The respondent is re-shown the concept identified in the BYO section, the concept winning the Choice Tasks tournament, and three others chosen from among those he/she has identified as worthy of consideration. We ask for each of those concepts how likely he/she would be to buy it if it were available in the market, using a standard five-point Likert scale, with a screen similar to the one below:

How likely would you be to purchase this laptop?

*** This is the original laptop you configured ***

Size	15 inch screen, 6 lbs.
Brand	Dell
Processor	Intel Core 2 Duo T7400 (2.16GHz)
Operating System	Vista Home Premium
Memory	1 GB
Hard Drive	100 GB
Video Card	128 MB Video card, adequate for most use
Battery	3 hours
Productivity Software	Microsoft Office Small Business (Basic + PowerPoint, Publisher)
Price	$1,550

Definitely Would Not	Probably Would Not	Might or Might Not	Probably Would	Definitely Would
○	○	○	○	○

Next

This section of the interview is used only for estimation of a partworth threshold for "None." Partworths from other sections of the interview are used to estimate the respondent's utility for each concept, and then a regression equation is used to produce an estimate of the utility corresponding to a scale position chosen by the researcher, such as, for example, somewhere between "Might or Might Not" and "Probably Would." Within the market simulator, if the utility of a product concept exceeds the None utility threshold, it is chosen. (None of the simulations presented in this paper used the "None" utility threshold.)

The interview as a whole attempts to mimic the actual in-store buying experience that might be provided by an exceptionally patient and interested salesperson. For example, after the BYO section she explains that this exact product is not available but many similar ones are, which she will bring out in groups of five, to see whether each is worthy of further interest. The Choice Tasks section is presented as an attempt to isolate the specific product which will best meet the respondent's requirements.

If the respondent has answered conscientiously, he/she will find that the final product identified by the salesperson as best is actually more preferred than the original BYO product. This occurs because the overall prices of the products generated in the product pool are varied as much as +/- 20% from the fixed BYO prices. Therefore, at least one of those (in our case, 40) product concepts will feature better features than the BYO product at the same price, the same features at a lower price, or a combination of these benefits. This makes it seem that the salesperson in our ACBC interview has actually done a good job finding a product that exceeds the quality of the BYO product and fits the needs of the respondent.

Method of Analysis

The data from the first three sections of the questionnaire can be analyzed with a multinomial logit model. Although the respondent was not actually completing conventional choice tasks in the first two sections of the interview, we can structure the data in synthetic choice tasks, as follows.

- The **BYO** section can be considered to produce one choice task per (non-price) attribute. Each task contains information for only a single attribute, and each alternative consists of a single level and an accompanying price.

- The **Screening** section can be considered to produce as many choice tasks as product alternatives that are screened. For each alternative, we compose a choice task that pairs the product alternative versus a constant alternative representing a threshold of acceptability.

- Suppose that **c** concepts are taken into the **Choice Tasks** section. Then the choice data can be arranged in 2 * **c** choice tasks, with half containing three alternatives and half containing two alternatives. If we had asked only for first choices, the number of choice tasks from this section would be **c**.

All of the above real (or synthetic) choice tasks can be combined[2] in one multinomial logit analysis[3]. The amount of information obtained is greater than from a typical CBC interview and may be enough information to permit estimation of individual partworths without having to "borrow" information from other respondents using HB analysis.

Sawtooth Software's current HB algorithm assumes that respondent error is constant across the different kinds of synthetic choice tasks. There is empirical evidence that error levels are higher when more complex judgments are required. (For example, Johnson, Orme and Pinnell (2006) found that BYO data contained less error than CBC data.) Our analysis presented here assumes constant error levels in all questionnaire sections, but Thomas Otter has modeled these same data using modifications of the HB algorithm to permit varying error levels (Otter 2007). His findings confirm that the way we have used HB to estimate partworth utilities works quite well, but also suggest that perhaps even better results could be achieved with more appropriate models.

An Experiment

Early in 2007 we performed an experiment[4] to compare this new type of adaptive CBC questionnaire (ACBC) with conventional CBC. The subject was laptop computers, described by 10 attributes with a total of 37 levels. The attributes and their levels are shown in Table 1.

[2] To investigate the relative contribution of the three main ACBC sections, we omitted each section (retaining the other two sections) and measured the decrease in predictive ability (vis-à-vis holdouts) of the model relative to retaining all information. We found that the relative worth of the sections was, in rank order, 1) BYO, 2) Screening, and 3) Choice Tasks. (Note, in Appendix C, the screening section had the most worth in predicting holdouts in a second ACBC study.) Our procedure helped us get a rough assessment of the impact of the various sections, but we recognize Allenby et al.'s finding that deleting a previous section biases results based on later sections (Allenby et al. 2007).

[3] Our approach to estimation does not treat "unacceptable" levels as "absolutely unacceptable under all conditions." Each respondent's data are consistent with never choosing a product concept that includes the unacceptable level. However, HB shrinks individual estimates toward population parameters, so the unacceptable utility value, while strongly negative, is not scaled so negatively that it becomes an absolute barrier to purchase irrespective of all other potential feature improvements.

[4] A few months later, we had the opportunity to field a second test of ACBC, this time as part of a real study for a client. The results of that test are reported in Appendix C.

Table 1
Attributes and Levels for Laptop Questionnaire

Screen Size/Weight:
14 inch screen, 5 pounds
15 inch screen, 6 pounds
17 inch screen, 8 pounds

Brand:
Acer
Dell
Toshiba
HP

Processor:
Intel Core 2 Duo T5600 (1.86GHz)
Intel Core 2 Duo T7200 (2.00GHz)
Intel Core 2 Duo T7400 (2.16GHz)
Intel Core 2 Duo T7600 (2.33GHz)

Operating System:
Vista Home Basic
Vista Home Premium
Vista Ultimate

Memory:
512 MB
1 GB
2 GB
4 GB

Hard Drive:
80 GB
100 GB
120 GB
160 GB

Video Card:
Integrated video, shares computer memory
128MB Video card, adequate for most use
256MB Video card for high-speed gaming

Battery:
3 hour
4 hour
6 hour

Productivity Software:
Microsoft Works
Microsoft Office Basic (Word, Excel, Outlook)
Microsoft Office Small Business (Basic + PowerPoint, Publisher)
Microsoft Office Professional (Small Bus + Access database)

Price:
$1,000
$1,300
$1,700
$2,200
$2,800

Data were obtained from the Opinion Outpost Internet Panel. Respondents first answered a brief screener to ensure that they had at least moderate familiarity with the product category. Approximately 600 respondents were then divided randomly into two groups, with half participating in an ACBC interview, and half participating in a conventional CBC interview. Respondents in each group first received three holdout choice tasks, each consisting of four alternatives. A fourth holdout task was constructed for each respondent by combining the concepts preferred in the first three tasks.

Following the holdout tasks the ACBC respondents answered the questionnaire described above and the CBC respondents answered a conventional CBC questionnaire with 18 choice tasks, each with four alternatives plus a "None" alternative (see Appendix A for layout of CBC task).

Finally, all respondents received an identical set of questions in which they rated their interview experience on several qualitative aspects.

Experimental Results

We observed that a few respondents in each group had unusually short interview times, and a few others in each group had very long times. We thought the fastest respondents had probably not taken the task seriously, and that the slowest ones might have been distracted and hence not given us their full effort. To minimize the possibility of including respondents of either type, we deleted the fastest 5% and the slowest 10% of each group. The partworth utility estimates and implied importances for the remainder of the respondents, (277 for CBC and 282 for ACBC) appear somewhat similar, as shown in Appendix B. However, one notes greater curvature (disutility for worst levels) for ACBC data, and more reliance on Brand to make product choices among the CBC respondents.

We recorded the interview time and also asked respondents some qualitative questions regarding their experience with the surveys. Here are the qualitative results:

Table 2
Qualitative Results Comparing ACBC with CBC

Median time to complete the CBC or ACBC sections
(excluding the screener questions and post qualitative questions):

ACBC	11.6 minutes
CBC	5.4 minutes

How would you compare your overall experience with this survey compared to other internet surveys you have completed?

	ACBC	CBC	
This survey was far better (5):	**24%**	15%	
This survey was better (4):	**47%**	44%	
This survey was about the same (3):	26%	**35%**	
This survey was worse (2):	2%	**4%**	
This survey was FAR worse (1):	0%	**2%**	
Means:	3.93	3.66	(t = 4.1)

How much do you agree with the following statements about this survey?
(5=Strongly Agree, 1=Strongly Disagree; Top Box % shown beneath means.)

	ACBC	CBC	
Q1. The laptop configurations I was asked to evaluate seemed realistic.	**4.4** **54%**	4.1 37%	(t=4.2)
Q2. This survey was at times monotonous and boring.	**2.6** **4%**	2.8 6%	(t=2.3)
Q3. I'd be very interested in taking another survey just like this in the future.	4.3 52%	4.3 54%	(t=0.4)
Q4. The survey format made it easy for me to give realistic answers that reflect exactly what I'd do if buying a real laptop.	**4.3** **48%**	4.1 37%	(t=2.7)
Q5. The way the laptops were presented made me want to slow down and make careful choices.	**4.1** **38%**	3.9 27%	(t=2.9)

Qualitative Results:
The average ACBC interview took about twice as long as the average CBC interview. That may appear to be a disadvantage at first, but it seems less so when one realizes that

CBC respondents spent an average of only 18 seconds per choice set, seemingly inadequate time to provide truly thoughtful answers.

ACBC had significantly more favorable answers than CBC on five of the six questions, despite its greater interview time. This suggests that we may have achieved our goal of providing a more stimulating experience to encourage more engagement in the interview.

Hit Rates:

Hit rates for the holdout tasks revealed interesting differences between groups. Recall that there were three holdout tasks each having four alternatives, and a final holdout task that was custom-made for each respondent, containing the winners from his/her first three holdout tasks. Prior to collecting the data, we hypothesized that ACBC would have an advantage over conventional CBC in predicting the outcome for the final holdout task (that presented the three winning concepts from the previous holdout tasks). We did not know what to expect for first three static holdout concepts.

Table 3
Holdout Hit Rates

	ACBC	CBC	
First three holdouts	55.7%	57.0%	
Fourth holdout	60.8%	50.0%	(t = 2.54)

For the first three holdout tasks there is no significant difference, although for these samples of respondents CBC has a slight advantage. However, for the fourth holdout there is a large and significant difference in favor of ACBC.

Earlier, we presented evidence that CBC respondents may be using simplified strategies for responding to choice tasks, in which they may pay attention to only a few attribute levels. The first three holdout tasks, as well as the calibration tasks in the CBC questionnaire, were constructed with "minimal overlap," with the alternatives in each choice set being as different from one another as possible. For example, with four brands and four alternatives in a choice set, there was always one alternative with each brand. Thus a respondent who happened to answer by always choosing a particular brand could answer every choice task consistently, and could also answer the holdout questions using the same strategy. Respondents behaving in this way could be expected to do well on the first three holdout tasks.

However, the fourth holdout task did not have this characteristic, since it was assembled from those alternatives previously preferred by each respondent. For example, if a respondent had consistently chosen a particular brand, then all three alternatives in the final holdout task would have featured that brand. If a respondent had answered the calibrating questions simply by choosing a preferred brand, his answers would contain no information with which to predict his choice in the fourth holdout. (Although, with HB estimation of partworths, the borrowing of information from other respondents would have provided some relevant information.) Thus, the fact that ACBC had a significantly

better hit rate than CBC on the fourth holdout tends to confirm that some CBC respondents may have resorted to simplification of their decision processes, and that ACBC captures a greater depth of attribute processing that is more predictive of challenging choice scenarios where concepts are closer in utility and perhaps tied on key aspects.

Share Predictions:

The fourth holdout choice set was custom-made for each respondent, so it was not useful for share predictions, which require that the same choice sets be shown to many respondents. We thought it desirable to have more than three holdout choice sets for share predictions, and also thought it would be interesting to see how well our two treatment groups could predict holdout shares generated by an entirely different sample of respondents.

Accordingly, we used another group of 955 panelists who completed the same screener to assure familiarity with the product category, and who then answered 12 choice tasks (standard CBC format with 4 concepts per task, without a "None") that were identical for all respondents. These were generated to have a modest degree of level overlap. We arbitrarily deleted the fastest 28 and the slowest 27 respondents, leaving a total of 900. These were divided into three groups of 300 on the basis of their times taken to answer the holdout questionnaire. Table 4 gives Mean Absolute Errors of share predictions for the CBC and ACBC respondents, when used to predict shares for all 900 holdout respondents, as well as each third of them based on holdout interview time. (For each prediction, we tuned the scale factor to minimize the MAE.)

Table 4
Mean Absolute Errors for Prediction of Holdout Shares

	Total Holdout Sample (n=900)	Fastest 1/3 (n=300)	Middle 1/3 (n=300)	Slowest 1/3 (n=300)
ACBC	4.52	5.24	4.72	4.42
CBC	4.49	4.88	4.95	4.98

We are not aware of good statistical tests for comparing differences in MAE for choice shares across 12 choice tasks, but the differences in Table 4 tell a consistent story. ACBC essentially matches the overall prediction accuracy of CBC, but excels in predicting the shares generated by the slower two groups of holdout respondents. In contrast, CBC has smaller prediction errors when predicting holdout shares generated by respondents who answered the holdout tasks most quickly.

These results also seem consistent with the hypothesis that some CBC respondents may use simple decision rules, such as choosing products that have a small number of critical attribute levels. It seems reasonable that holdout respondents who take longer with their choices may be using more elaborate and potentially different decision rules. To investigate this possibility, we used the Swait/Louviere test to assess whether the slow and fast responders to the 12 holdout questions differed significantly with respect to main effect parameters after controlling for scale (the design of the 12 holdout tasks supported main effects estimation). The test for difference in parameters was strongly significant, with $p<0.001$. We also found that the slower group had a scale factor 40% larger than the faster respondent group, implying less error in their responses. Table 4 provides evidence favorable for ACBC. Despite the strong methods bias which should favor CBC in predicting CBC holdouts, ACBC can match CBC's prediction accuracy overall. More importantly, ACBC produces better predictions of the shares generated by more thoughtful holdout respondents.

Further Evidence of Simplification

At the 2006 Sawtooth Software Conference, Hoogerbrugge and van der Wagt (H&W) presented an interesting paper titled "How Many Choice Tasks Should We Ask" (2006). They re-analyzed a large number of CBC data sets with HB, in which they first estimated respondent partworths using only the first choice task, then again using the first two choice tasks, etc. For each re-analysis they used the estimated partworths to predict a holdout choice task, and they measured success with hit rates. They found that hit rates increased as the number of calibration choice tasks increased until about ten choice tasks, but from then on the hit rates were essentially flat. They concluded that there was often little reason to administer more than ten choice tasks to a CBC respondent. These results were surprising to many researchers, because general statistical experience has led us to expect that prediction will be better with more information. If respondents were using compensatory models to make their choices, one would expect that more information would indeed permit better predictions.

Our CBC respondents had a total of 18 calibration choice tasks plus four holdout tasks. We have duplicated the H&W analysis with our data, and we reach similar conclusions. Our hit rates increase gradually when using from 2 to 12 calibration tasks, after which there is no further systematic improvement.

If respondents are simplifying their decision processes by paying attention to only a few attribute levels, that pattern can be detected after relatively few choice tasks. Therefore, it may be that H&W have provided additional evidence that respondents are in fact simplifying their decision processes.

Benefits from Increased Information

We have pointed out that the ACBC interview should provide more information than a conventional CBC interview. This raises the question of whether ACBC data may be especially useful when the researcher is faced with small samples, or even for individual-level estimation.

To examine the effect of small samples, we drew 10 random samples of 25 respondents of each type and re-estimated individual partworths in each sample using HB. We reasoned that if ACBC provided more information, the estimates of population parameters should be more precise, leading in turn to better estimation of individual partworths. We used the partworths estimated from each sample to predict choice shares for the holdout respondents on the 12 holdout choice sets.

ACBC had an advantage with a mean absolute error of 6.29 share points compared to 6.91 for CBC. This difference of 0.62 share points may be compared to a corresponding difference of 0.03 share points in favor of CBC for estimates obtained when all respondents are used to estimate population parameters (see Table 4). Thus it appears that ACBC has an advantage over CBC when sample sizes are small.

It should be noted that ACBC's superior performance occurs despite the disadvantage that the holdout responses being predicted are CBC responses. The presence of any "methods bias" would be a disadvantage for ACBC.

We have also used a simple monotone regression algorithm (Johnson, 1975) to estimate partworths. This approach makes no assumptions about error distributions. It simply seeks a set of partworths that satisfy the inequality constraints implied by the data. Any levels that were marked as unacceptable for the respondent were given an arbitrary low partworth. Each respondent's partworths are estimated using only information from his own responses, so it provides strictly "individual-level" estimation. Table 5 provides hit rates for partworths estimated by monotone regression, compared to previously shown hit rates for HB estimates.

Table 5
ACBC Hit Rates, Including Monotone Regression Estimation

	HB Estimation		Monotone Regression
	CBC	ACBC	ACBC
First three holdouts	57.0%	55.7%	52.2%
Fourth holdout	50.0%	60.8%	57.0%

The first fact evident from Table 5 is that hit rates for monotone regression are inferior to those for HB. When even small samples are available, HB appears to be the preferred estimation method. The second fact regards the fourth holdout choice set, which was deliberately constructed so as to be difficult to predict from choice data in which respondents had paid attention to few attribute levels. Both methods for estimating partworths from ACBC seem to perform better than conventional CBC under HB estimation.

To examine ACBC's success at holdout share predictions when monotone regression is used for estimation, in Table 6 we repeat the overall results from Table 4, but with an additional column for monotone regression predictions.

Table 6
Mean Absolute Errors for Prediction of Holdout Shares, Including Monotone Regression

	HB Estimation		Monotone Regression
	CBC	ACBC	ACBC
Mean Absolute Error	4.49	4.52	4.54

ACBC's share predictions from monotone regression are essentially as good as those from HB estimation. This is somewhat unexpected. We'd generally recommend that HB be used for estimation whenever possible (especially given an improved HB estimation approach detailed by Thomas Otter in his paper presented at this same conference), but that when strictly individual estimation is required, monotone regression can still provide useable results.

Because ACBC contains relatively more information than conventional CBC, it may provide additional benefits for segmentation research, whether via demographic variables or latent classes. There is less shrinkage to population parameters when using HB with ACBC data and correspondingly larger scale, which is beneficial when characterizing distinct preferences of segments. Additionally, monotone regression can produce partworths that are truly individual, uninfluenced by group averages.

Summary and Conclusions

Results from this study help in understanding the previously puzzling results of our earlier ACBC attempts.

- Our previous attempts assumed that respondents answered in a compensatory way consistent with the logit model, and created choice tasks so as to increase statistical efficiency as defined by that model (by increasing utility balance). But for respondents who behave non-compensatorily, utility balance is irrelevant, since such a model says that the respondent looks for presence or absence of specific attribute levels irrespective of other aspects of the products. In our second ACBC paper, where our attempt at ACBC seemed to have failed, we found that we *had* increased statistical efficiency, but without improved predictions. That could be expected for respondents behaving non-compensatorily.

- In our first three attempts, we measured success with holdout tasks that had minimal overlap. A non-compensatory respondent can easily make consistent holdout choices among alternatives with minimal overlap, where consistency may be even easier to achieve than for a compensatory respondent. Thus, prediction of holdout choices among alternatives that have minimal overlap is not necessarily a good test of success under the logit rule.

For this current research, we were motivated to investigate new ways of collecting choice data consistent with the idea that many CBC respondents simplify their decision processes, paying attention to only a few critical attributes. This is a convenient way of dealing with a task perceived as being confusing, repetitive and boring. We believed it might be possible to structure a more interesting and engaging interview which let respondents identify any "must have" or "must avoid" attribute levels, and which encouraged more thoughtful evaluation of products compatible with those requirements.

While some researchers have tried to accomplish a more thorough evaluation of attributes through partial-profile models (both ACA and partial-profile CBC), we have accomplished this while maintaining the more realistic full-profile context.

We believe that the Adaptive CBC (ACBC) method for collecting data provides several improvements over conventional CBC.

- Although the interview takes longer (11.6 minutes rather than 5.4 in our experiment), respondents appear to have found the ACBC interview more interesting and engaging than CBC, and a more faithful simulation of the buying experience.

- ACBC produces better predictions for a choice set that was custom-designed for each respondent from concepts preferred in previous choice sets. ACBC was superior to CBC when predicting choice shares from the group of holdout

respondents who had taken longer to answer, and had therefore presumably been more thoughtful. In both of these cases methods bias significantly favored CBC, since the holdout tasks were CBC tasks.

- ACBC's superiority over CBC is also particularly evident when used with small samples of respondents. ACBC also permits estimation of truly individual-level partworths without the need to borrow information from other respondents (although they are probably not as successful as HB estimates).

Most choice researchers admit that task simplification at the individual level must exist, but many have believed that the aggregate effect of hundreds of respondents (each employing different simplification strategies) should counteract this problem and fairly accurately reflect the more careful processing of information of real-world decisions. Our results suggest that respondents who take more time to complete CBC questionnaires provide different aggregate shares, and that a data collection technique that encourages a greater depth of processing may produce more accurate share predictions. Of course, we cannot be certain that ACBC performs better at predicting real world choices than standard CBC until a more complete validation experiment involving actual purchases is available.

There are some types of CBC studies that wouldn't seem a good fit for the ACBC approach we've described here. Brand-Package-Price studies, for which conventional CBC has been very popular and quite successful, would not seem to us to benefit from this adaptive approach. However, for studies involving about five attributes or more, the adaptive procedure may offer compelling benefits.

Despite our success with this comparative study, there are ways that our approach to ACBC may be improved. For example:

- In a pilot test of the interview we created a pool of 50 concepts to be considered by the respondent in the Screening section, but in the experiment reported here we used 40. Further work is required to learn the optimal number, and how it may be related to the numbers of attributes and levels.

- Further work can be done to estimate the optimal amount to vary the attributes from the BYO concept when constructing the pool of products that each respondent evaluates.

- We showed those concepts to respondents in groups of five, which was an arbitrary decision based on screen size, legibility, and clutter. We don't know if this layout was optimal.

- In the Choice section we asked for identification of both "best" and "worst" alternatives. The information about worst alternatives was of almost no value in improving estimation, but the fact that we asked that question may have improved the quality of respondents' choices of "best."

Another significant potential source of improvement is in estimation rather than data collection. There is good reason to believe that respondents are more careful and provide better answers to the BYO section of the questionnaire than to the more repetitive and complex considerations of products profiled on many attributes simultaneously. Yet the HB algorithm we used for estimation assumes constant error levels in all parts of the questionnaire. At this same conference, Thomas Otter has presented a compelling way to deal with this problem, and his work shows that the predictive ability of ACBC can be further improved by using a specialized HB methodology that uses a better way to combine information from the three ACBC sections (Otter 2007).

References

Allenby, Greg, Thomas Otter, and Qing Liu (2007), "Endogeneity Bias: Fact or Fiction?" *Sawtooth Software Conference Proceedings*.

Gilbride, Timothy and Greg M. Allenby (2004), "A Choice Model with Conjunctive, Disjunctive, and Compensatory Screening Rules," *Marketing Science*, 23, 3, (Summer) 391-406.

Hauser, John R., Ely Dahan, Michael Yee, and James Orlin (2006) "'Must Have' Aspects vs. Tradeoff Aspects in Models of Customer Decisions," *Sawtooth Software Conference Proceedings*, 169-181.

Hoogerbrugge, Marco and Kees van der Wagt (2006) "How Many Choice Tasks Should We Ask?," *Sawtooth Software Conference Proceedings*, 97-109.

Huber, Joel and Klaus Zwerina (1996), "The Importance of Utility Balance in Efficient Choice Designs," *Journal of Marketing Research*, 33 (August) 307-317.

Johnson, Richard M. (1975) "A Simple Method for Pairwise Monotone Regression," *Psychometrika*, 40, 163-168.

Johnson, Richard M. and Bryan Orme (1996), "How Many Questions Should You Ask in Choice-Based Conjoint Studies?" downloaded from www.sawtoothsoftware.com/techpap.shtml.

Johnson, Richard M., Bryan Orme, and Jon Pinnell (2006). "Simulating Market Preference with 'Build Your Own' Data," *Sawtooth Software Conference Proceedings*, 239-253.

Johnson, Richard M., Bryan Orme, Joel Huber, and Jon Pinnell (2005). "Testing Adaptive Choice-Based Conjoint Designs," *Design and Innovations Conference (Berlin)*, available at www.sawtoothsoftware.com/techpap.shtml.

Johnson, Richard M, Joel Huber, and Bryan Orme (2004), "A Second Test of Adaptive Choice-Based Conjoint Analysis (The Surprising Robustness of Standard CBC Designs)" *Sawtooth Software Conference Proceedings*, 217-234.

Johnson, Richard M., Joel Huber, and Lynd Bacon (2003), "Adaptive Choice-Based Conjoint," *Sawtooth Software Conference Proceedings*, 333-343.

Otter, Thomas (2007), "Hierarchical Bayesian Analysis for Multi-Format Adaptive CBC," *Sawtooth Software Conference Proceedings*.

Appendix A

CBC Task Layout

If these were your only options, which laptop would you choose?
Choose by clicking one of the buttons below:

Screen Size/Weight:	14 inch screen, 5 pounds	17 inch screen, 8 pounds	15 inch screen, 6 pounds	17 inch screen, 8 pounds	
Brand:	Toshiba	Acer	HP	Dell	
Processor:	Intel Core 2 Duo T7600 (2.33GHz)	Intel Core 2 Duo T7200 (2.00GHz)	Intel Core 2 Duo T7400 (2.16GHz)	Intel Core 2 Duo T5600 (1.86GHz)	
Operating System:	Vista Home Basic	Vista Ultimate	Vista Home Premium	Vista Ultimate	
Memory:	2 GB	4 GB	512 MB	1 GB	
Hard Drive:	100 GB	120 GB	160 GB	80 GB	
Video Card:	128MB Video card, adequate for most use	Integrated video, shares computer memory	256MB Video card for high-speed gaming	Integrated video, shares computer memory	
Battery:	6 hour	3 hour	4 hour	3 hour	
Productivity Software:	Microsoft Works	Microsoft Office Basic (Word, Excel, Outlook)	Microsoft Office Professional (Small Bus + Access database)	Microsoft Office Small Business (Basic + PowerPoint, Publisher)	NONE: I wouldn't choose any of these.
Price:	$1,700	$1,000	$2,800	$1,300	
	○	○	○	○	○

Appendix B

Average Partworths (Normalized)

	ACBC n=282	CBC n=277
14 Inch, 5 pounds	-24.38	-16.47
15 Inch, 6 pounds	8.12	-0.50
17 Inch, 8 pounds	16.26	16.97
Acer	-25.81	-35.69
Dell	24.59	24.55
Toshiba	-3.44	-5.71
HP	4.65	16.85
1.86GHz Processor	-35.66	-11.93
2.00GHz Processor	2.36	-2.08
2.16GHz Processor	13.26	0.20
2.33GHz Processor	20.04	13.81
Vista Basic	-9.67	-4.30
Vista Premium	6.05	-2.14
Vista Ultimate	3.62	6.44
512MB RAM	-90.69	-89.30
1GB RAM	-11.21	-9.26
2GB RAM	40.94	35.69
4GB RAM	60.96	62.87
80GB Hard Drive	-48.07	-28.40
100GB Hard Drive	-3.18	0.39
120GB Hard Drive	18.22	4.20
160GB Hard Drive	33.02	23.81
Integrated Video	-37.88	-23.58
128MB Card	10.86	0.46
256MB Card	27.02	23.12
3 hour battery	-26.78	-19.32
4 hour battery	7.91	-2.02
6 hour battery	18.87	21.35
MS Works	-39.80	-34.31
MS Office Basic	12.58	7.72
MS Office Small Bus	19.44	13.33
MS Office Professional	7.78	13.25
Price	-101.29	-97.39

Average Importances

	ACBC n=282	CBC n=277
Size/Weight	8.25	8.25
Brand	8.10	13.90
Processor	7.67	5.74
OS	4.84	4.46
RAM	16.68	17.14
Hard Drive	9.87	7.47
Video Card	8.96	7.57
Battery	6.49	5.92
Software	9.79	8.72
Price	19.35	20.83

Appendix C

A Second Experiment to Test ACBC

A few months after completing the first test of ACBC as reported in this paper, we used ACBC in a real client study of a mechanical product for recreational equipment. We are grateful to Joe Curry of Sawtooth Technologies for sponsoring this project. Because this was an actual client project, we were not able to design the experiment as rigorously as our first test of ACBC (e.g. the ACBC respondents were collected a few weeks after the CBC respondents, with minor deviations in the method of recruitment). For this second test the questionnaire did not include graphics representing an interviewer, which we believe would have made a positive contribution. Also, we weren't able to collect a separate sample of holdout respondents. Despite these differences, the findings are quite similar to those of the first test.

This second study involved the following characteristics for the ACBC interview:

- 8 attributes (29 total attribute levels) plus price
- 36 products used in the Screening Section, shown in triples
- The Choice Tasks section asked just first choice from triples
- No graphic representing an interviewer was shown

Approximately 500 respondents completed a standard CBC survey and 400 completed the ACBC survey. The CBC survey involved 14 choice tasks shown in pairs. Four CBC-looking holdout tasks were included for all respondents, with the final holdout composed of the three concepts chosen in the earlier three fixed holdout tasks.

Qualitative Findings:
The ACBC respondents spent about triple the time doing the conjoint section of their questionnaire as respondents who completed the more abbreviated standard CBC interview (about 15 minutes compared to 5 minutes). Approximately 6% of the ACBC respondents dropped out of their survey during the conjoint questions compared to 1% of CBC respondents. The ACBC respondents reported that their survey was more monotonous than CBC respondents did (this is opposite what we found with the first laptop study), but both groups reported equal interest in taking another survey like theirs in the future. In addition, the ACBC respondents reported that the products they were shown were more realistic (confirming findings of the laptop study).

Quantitative Findings:
The aggregate utilities were correlated 0.91 between ACBC and CBC respondents. Attribute importances were very similar, with one attribute (warranty) appearing significantly more important for CBC respondents. We did not note any enhanced "curvature" (loss avoidance) for the worst levels from ACBC utilities compared to the CBC utilities (in contrast to what we observed with the laptop study). The hit rates for the fixed CBC-like holdout tasks favored CBC slightly but not significantly. Hit rates for the customized holdout CBC choice task differed more strongly, and in favor of ACBC:

62.2 vs. 59.5 (difference not significant). Again, methods bias was strongly in favor of CBC in terms of ability to predict this holdout, as the choice task was a CBC task.

In contrast to the first ACBC study where we found that the BYO section was the most valuable of the three sections (BYO, Screening Section, Choice Tasks), this time it was least valuable. The rank-order of contribution toward predicting holdouts was 1) Screening Section, 2) Choice Tasks, 3) BYO. We think we have an explanation for this discrepancy. The BYO section focuses on the tradeoff between each feature and price. In this second study, price overall was not very important relative to the other attributes. Therefore, BYO's focused effort on estimating price sensitivity didn't pay off as well here (in terms of predicting holdout choices) as with the laptop study where price carried much more importance.

Discussion:
It is impressive that ACBC again beats CBC in predicting the customized (and difficult) CBC-looking task, despite the methods bias in favor of CBC and the fact that ACBC utilities are not equivalent to those of CBC. Of course, the best test of validity would involve actual purchases, for which we do not have data. Respondents described the ACBC interview as more monotonous compared to CBC (opposite our findings from the laptop study), and we wonder whether not showing a graphic of an engaging facilitator made a difference, or if it is principally explained by the greater relative difference in task length for ACBC relative to CBC in this study.