Sawtooth Software

TECHNICAL PAPER SERIES

ACBC Technical Paper

1457 E 840 N
Orem, UT 84097
(801) 477-4700
www.sawtoothsoftware.com

The Adaptive Choice-Based Conjoint (ACBC) Technical Paper

Orem, Utah USA
February, 2014

Background

Choice-Based Conjoint (CBC) has been the most widely used conjoint technique among our user base since about the year 2000. 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. For this reason, CBC studies have typically required larger sample sizes than ratings-based conjoint methods. In CBC studies, 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. Later, hierarchical Bayes methods were developed to permit individual-level analysis, but interest has remained in ways to design choice tasks so as to provide more information.

Problems with Traditional CBC

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. It's hard to imagine how they could evaluate so much information on the screen in such short order. 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 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.

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.

There has been a lot of effort dedicated to designing efficient CBC experiments featuring orthogonality and high D-efficiency. These efforts have assumed that respondents answer using an additive process consistent with the logit rule. We have become increasingly convinced that *most* respondents to complex conjoint studies employ non-compensatory heuristics at odds with the logit rule, and that efforts to improve design efficiency assuming compensatory processing may be misdirected. ACBC's design strategy is effective for respondents that employ various degrees of non-compensatory and compensatory processes. In terms of the traditional design criteria, ACBC's designs would be judged inferior. But this is an inappropriate standard given that so many respondents apply cutoff rules, such as screening based on must-have or unacceptable levels.

The Quest for 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 greater need at this point is not for better models, but rather for better data. Adaptive CBC is a new, promising method that responds well to the problems above. It uses the trusted full-profile method of presenting product concepts. Its surveys are more engaging for respondents. Rigorous comparison studies to traditional CBC suggest that our adaptive form leads to data that are more accurate and predictive of choice behavior. It captures more information at the individual level than traditional CBC surveys and may be used even with small (for example, business-to-business) samples. As of 2014, more than 1500 ACBC studies have been conducted by Sawtooth Software users (based on our annual tracking survey of our customers).

Adaptive CBC is best applied for conjoint-type problems in which there are about five attributes or more. Both standard and advanced alternative-specific designs are supported. Studies involving few attributes (such as the brand+package+price studies commonly done in FMCG research) wouldn't seem to benefit from ACBC. The store-shelf approach in our standard CBC software would seem more appropriate for such studies.

Sections and Flow for Adaptive CBC Surveys

The objectives of the Adaptive CBC interview are as follows:

- Provide a 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.

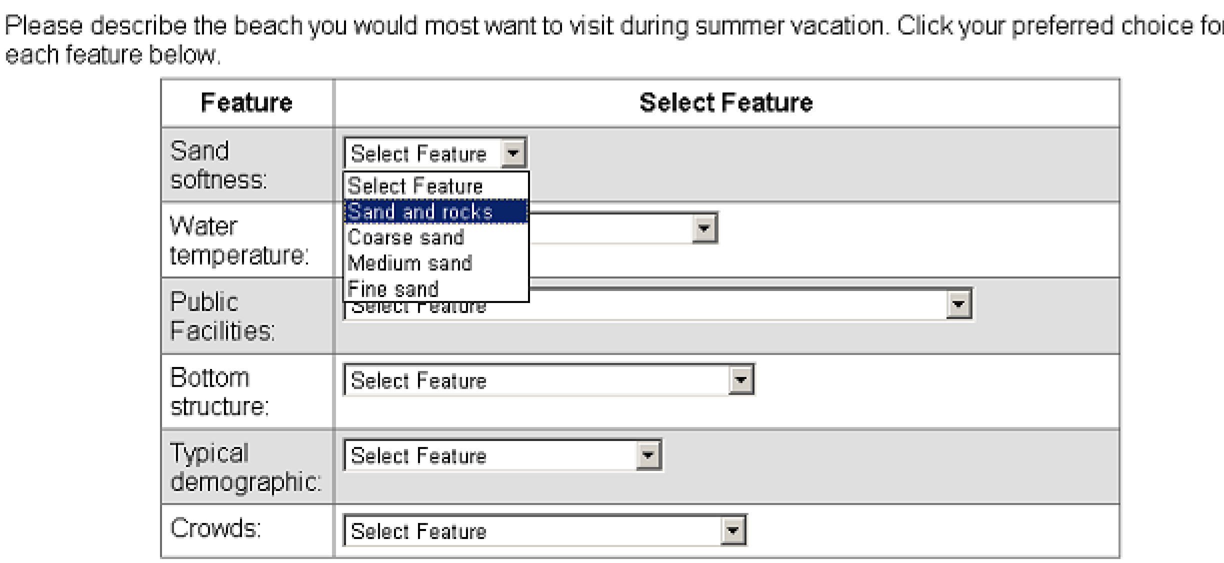
Typically, an ACBC interview includes the following three core sections:

- BYO (Configurator)
- Screening Section
- Choice Tasks

But, ACBC offers the flexibility to skip sections, for advanced researchers who decide (for certain research situations) that some sections do not fit the buyer's (chooser's) decision making process.

BYO (Configurator) Section:

In the first section of the interview the respondent answers a "Build Your Own" (BYO) question 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. A typical screen for this section of the interview is shown below:


Please describe the beach you would most want to visit during summer vacation. Click your preferred choice for each feature below.

Feature	Select Feature
Sand softness:	Select Feature (Select Feature / Sand and rocks / Coarse sand / Medium sand / Fine sand)
Water temperature:	
Public Facilities:	Select Feature
Bottom structure:	Select Feature
Typical demographic:	Select Feature
Crowds:	Select Feature

An alternate display incorporates radio buttons rather than combo boxes:

Please describe the beach you would most want to visit during summer vacation. Click your preferred choice for each feature below.

Feature	Select Feature
Sand softness:	○ Sand and rocks ○ Coarse sand ○ Medium sand ○ Fine sand
Water temperature:	○ 60°F/15°C average (wetsuit required) ○ 70°F/21°C average ○ 80°F/27°C average ○ 90°F/32°C average (bathwater)
Public Facilities:	○ No public restrooms, showers or changing facilities ○ Public restrooms available, but no showers or changing facilities ○ Public restrooms, showers & changing facilities available
Bottom structure:	○ Sandy bottom underwater past shoreline ○ Rocky bottom underwater past shoreline ○ Coral bottom underwater past shoreline ○ Mud bottom underwater past shoreline
Typical demographic:	○ Popular with young partygoers ○ Popular with young families ○ Popular with mature adults
Crowds:	○ Uncrowded - ample open space ○ Somewhat crowded - some open space ○ Crowded - very little open space

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.

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.

Some attributes with obvious *a priori* preference order would seem to be out of place within a BYO question. It may not make sense in some situations to assign price premiums to the preferred levels, and it would be "obvious" to ask respondent to indicate their preferred level. In those cases, you can drop such attributes from the BYO question type (but these attributes appear within the remaining sections of the ACBC survey).

Advanced users sometimes entirely skip the BYO section (by indicating that all attributes should not be included in the BYO section). This has an effect on the later product concepts shown to respondents. If no BYO section is shown, then we equally sample across all levels within the experimental design, much like a standard CBC experiment (rather than oversampling the BYO-selected levels).

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 recommend showing 3 to 5 at a time per screen, for about 7 total screens of concepts). In the Screening Section, the respondent is not asked to make final choices, but rather just indicates whether he/she would consider each one "a possibility" or "not a possibility." A typical screen from this section of the interview is shown below:

Here are a few beaches you might like. Do any of these look like possibilities? For each, indicate whether it is a possibility or not. (1 of 6)

Sand softness:	Fine sand	Coarse sand	Fine sand	Medium sand
Water temperature:	90°F/32°C average (bathwater)	80°F/27°C average	80°F/27°C average	80°F/27°C average
Water safety:	Frequent rip currents, occasional dangerous animals	Occasional rip currents, occasional dangerous animals	No rip currents, few dangerous animals	Occasional rip currents, occasional dangerous animals
Beach cleanliness:	Trash, glass and manmade rubble rarely found	Trash, glass and manmade rubble occasionally found	Trash, glass and manmade rubble commonly found	Trash, glass and manmade rubble occasionally found
Public Facilities:	Public restrooms, showers & changing facilities available	Public restrooms, showers & changing facilities available	Public restrooms available, but no showers or changing facilities	Public restrooms, showers & changing facilities available
Bottom structure:	Coral bottom underwater past shoreline	Coral bottom underwater past shoreline	Sandy bottom underwater past shoreline	Mud bottom underwater past shoreline
Typical demographic:	Popular with young families	Popular with young families	Popular with young families	Popular with mature adults
Crowds:	Crowded - very little open space	Somewhat crowded - some open space	Uncrowded - ample open space	Somewhat crowded - some open space
	○ 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

The Screening Section is also used to estimate the "None" parameter threshold.

Advanced users may choose to skip the Screening Section and in that case all generated product concepts are taken forward into the Choice Tasks Section, as if the respondent had indicated that each one was "a possibility." If you skip the Screening Section, a "None" parameter threshold is not available, unless you include the final Calibration Section and also perform the additional step of calibrating the estimated utilities.

Unacceptables:

After a few screens of concepts have been evaluated, 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 systematically avoided an attribute level (or range of levels for ordered attributes). In that case, we ask whether that level would be completely unacceptable (“Unacceptables)”. Here is a typical screen for this question:

Would any beach having the features below be **totally unacceptable**? If so, mark the **one feature** that is most unacceptable to you, so I can focus better on beaches that meet your needs.

- ○ Bottom structure: - Mud bottom underwater past shoreline
- ○ Bottom structure: - Rocky bottom underwater past shoreline
- ○ Public Facilities: - No public restrooms, showers or changing facilities
- ○ Sand softness: - Medium sand
- ○ Crowds: - Crowded - very little open space
- ○ Typical demographic: - Popular with young partygoers
- ○ Water temperature: - 60°F/15°C average (wetsuit required)
- ○ Water temperature: - 90°F/32°C average (bathwater)

- ○ None of these is entirely unacceptable

If you include price in Unacceptables questions, ACBC scans the previous answers to determine the highest price ever selected as “a possibility” in the Screener or BYO questions. That highest price is offered as an unacceptable threshold.

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 with ACBC. 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 screen of typically three to five products has been screened (as a “possibility” or not), another “unacceptable” screen is shown and the respondent has another opportunity to add a subsequent cutoff rule. If the respondent identifies any “unacceptable” levels, then all further concepts shown will satisfy those requirements.

Must Haves:

We also scan all previous answers to determine if the respondent has expressed interest in only one level of some attribute. In that case, we ask whether that level is an absolute requirement (a "Must Have").

For example:

I don't want to jump to conclusions, but I've noticed you've chosen beaches with certain characteristics shown below. Is any of these an **absolute must** for the beach you would like to visit? If so, mark the **one most important feature**, so I can focus better on beaches that meet your needs.

- ○ Typical demographic: - Popular with young families
- ○ Water safety: - At most: Occasional rip currents, occasional dangerous animals
- ○ Water temperature: - At least: 80°F/27°C average
- ○ Water temperature: - At most: 80°F/27°C average

- ○ None of these is an absolute must

Based on early evidence (Orme 2009), it seems better to place Must Have questions directly following Unacceptable questions, and on a separate page. We suggest you wait to ask Must-Have questions until at least two Unacceptable questions have been asked.

Choice Tasks Section:

Once the respondent has completed the planned number of screens of Screening questions (typically 7 to 9 screens, where each screen includes 4 to 5 concepts), we transition to the Choice Tasks Section (tournament). 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:

Among these three, which beach would you most want to visit for summer vacation? (I've grayed out any identical features, so you can just focus on the differences.)

(1 of 4)

Sand softness:	Fine sand	Medium sand	Fine sand
Water temperature:	80°F/27°C average	80°F/27°C average	80°F/27°C average
Water safety:	Occasional rip currents, occasional dangerous animals	Occasional rip currents, occasional dangerous animals	Occasional rip currents, occasional dangerous animals
Beach cleanliness:	Trash, glass and manmade rubble occasionally found	Trash, glass and manmade rubble occasionally found	Trash, glass and manmade rubble occasionally found
Public Facilities:	Public restrooms, showers & changing facilities available	Public restrooms available, but no showers or changing facilities	Public restrooms, showers & changing facilities available
Bottom structure:	Coral bottom underwater past shoreline	Coral bottom underwater past shoreline	Coral bottom underwater past shoreline
Typical demographic:	Popular with young partygoers	Popular with young families	Popular with mature adults
Crowds:	Somewhat crowded - some open space	Crowded - very little open space	Uncrowded - ample open space
	○	○	○

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. If displaying concepts in triples, it takes $t/2$ choice tasks to identify the overall winner, where t is the number of concepts marked as "possibilities" from the previous section (in the case that t is odd and $t/2$ is not an integer, one rounds down to determine the required number of triples).

Although it may seem to some that the goal of the tournament section is to identify an overall winning concept, the actual goal is to engage respondents in a CBC-looking exercise that leads to good tradeoff data for estimating partworth utilities.

Calibration Section (Optional):

The fourth section of the interview is optional, and may be used to estimate a different "None" parameter from that provided by the Screening Section.

The respondent is re-shown the concept identified in the BYO section, the concept winning the Choice Tasks tournament, and (typically) four others chosen from among both previously accepted and rejected concepts. 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 visit this beach for summer vacation?

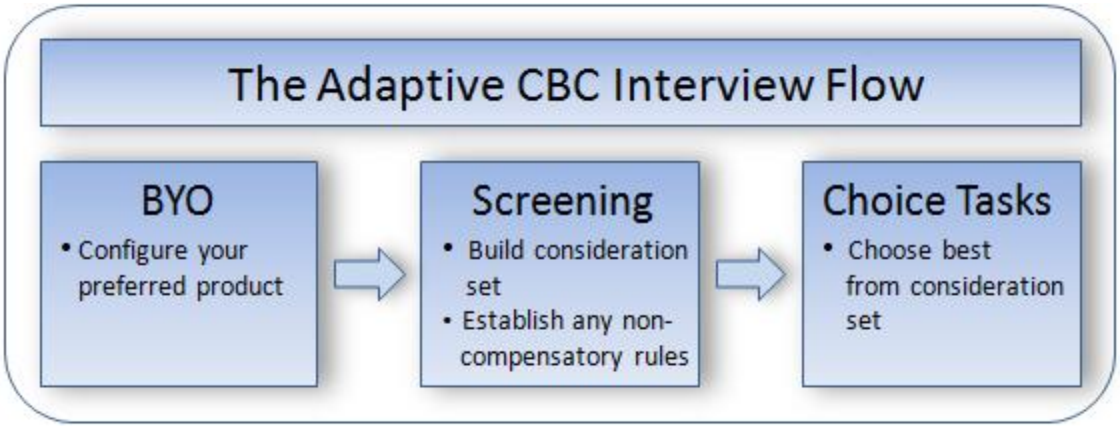
(3 of 4)

Sand softness:	Fine sand
Water temperature:	80°F/27°C average
Water safety:	Occasional rip currents, occasional dangerous animals
Beach cleanliness:	Trash, glass and manmade rubble occasionally found
Public Facilities:	Public restrooms, showers & changing facilities available
Bottom structure:	Coral bottom underwater past shoreline
Typical demographic:	Popular with young partygoers
Crowds:	Somewhat crowded - some open space
	○ Definitely Would ○ Probably Would ○ Might or Might Not ○ Probably Would Not ○ Definitely Would Not

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.

Summary:

The standard Adaptive Choice survey process can be summarized in the following graphic:



Respondents first configure their preferred product via a **BYO** question. Based on that preferred product, we create a set of similar products for the respondent to evaluate in the **Screening** section. Respondents indicate which of these similar products they would consider, and reveal non-compensatory cutoff rules. Finally, respondents make a final product selection among products screened into their consideration set. This is done via a **Choice Tasks** section, using a tournament format.

Researchers may also customize the interview flow, choosing to skip one or more of these sections to meet the needs of the project.

Attributes and Levels for ACBC

ACBC supports both standard attribute lists (where each attribute applies generically to all product concepts) and alternative-specific attribute lists (where some attributes only apply to, say, certain brands). For the most part, designing attributes and levels for ACBC follows the same guidelines as for other conjoint methods. But, there are some key differences related to the treatment of price and the possibility of using constructed lists to customize the attribute list each respondent sees.

Most projects will probably present each respondent the full list of attributes and levels in the study. However, ACBC gives you the option to drop attributes and/or levels from consideration (you need to do this prior to the respondent starting the ACBC questions). This is facilitated by SSI Web's constructed list building logic. As you enter your attributes and levels for your ACBC study, you'll be specifying them within Predefined Lists. The attribute list is a Predefined List with as many elements as attributes in your study. Each element in the attribute list is associated with a Predefined List that enumerates the levels.

Generally, we recommend that no more than about 12 attributes[1] with no more than about seven levels each be taken forward to the ACBC questionnaire. If your study involves more attributes and levels than this, you may want to use preliminary questions within SSI Web (CiW) to eliminate any attributes that are completely unimportant and any levels that are completely irrelevant.

[1] The extremes that have been attempted and published include as few as 4 attributes (Orme 2009) to as many as 14 attributes in each respondent's ACBC exercise (Binner *et al.* 2009).

Price is often included in choice studies, and initial evidence suggests that ACBC should be a strong approach for studying the impact of price on choice. Although it is possible to include a Price attribute as an attribute with discrete levels (such as $10, $15, $20), we generally recommend using ACBC's Summed Price approach and to treat price as a continuous variable.

Summed Prices

In traditional conjoint studies, the researcher applies price as another attribute (factor) in the study design, and specifies typically 3 to 5 levels of price (e.g. $100, $120, $140, $160). (One certainly could take this approach with ACBC.) The problem with such an approach is that $100 is sometimes shown with a collection of high-end features and $160 is sometimes shown with low-end features. With our standard CBC software, conditional pricing allowed the researcher to specify that certain attributes should by conditional on price, and always include a premium or discount. With ACBC, we have taken the idea of conditional pricing further, by allowing the researcher to specify incremental prices for up to all attributes in the study. When displaying the total product price, we sum the prices associated with the levels across all attributes in the product concept, and then we vary that summed price by a randomly drawn price variation (such as anywhere from -30% to +30%), as specified by the researcher. The researcher can also indicate that prices (after being disturbed randomly) should be rounded to, say, the nearest $100, or $5, $1, or $0.10 (for whatever currency symbol is appropriate).

Using the "summed" pricing approach leads to product concepts that show realistic prices (and therefore reflect higher utility balance across product concepts relative to traditional CBC). Products with high-end features will generally carry higher prices, and products with low-end features will generally carry lower prices. Under summed pricing, thousands of potential unique prices will have been shown to respondents, and the utility function is estimated by fitting a linear (or non-linear function). Under summed pricing, we may estimate price as a continuous function. When treating price in this manner, we are able to partial-out the effect of price vis-a-vis the effects of other attributes' levels. Therefore, one can interpret the utilities for the other levels independent of the price increments that were associated with them (which one cannot do when using conditional price in our CBC software).

Design Generation Strategy for ACBC

ACBC creates an array of product concepts for the respondent to evaluate within the Screener and Choice Tasks sections of the questionnaire. These concepts are chosen as "near-neighbors" to the product concept the respondent chooses in the BYO task, but still include the full range of levels taken into each respondent's ACBC survey. Because the BYO-specified product concept differs across respondents and the number of attributes and levels taken into each respondent's ACBC exercise can be dynamic (though in most projects the list will be static), it isn't possible to create an experimental design prior to fielding the study (as is done in non-adaptive CBC studies). Customized designs must be generated on-the-fly for each respondent. Because dozens or even hundreds of respondents might simultaneously be completing surveys over the web, we are concerned about the amount of computing resources demanded of the server to generate designs. Therefore, we've developed a relatively quick algorithm. The algorithm cannot be said to produce optimal designs, but its designs are near-orthogonal, and have proven to work exceptionally well in many methodological studies to date comparing ACBC to standard CBC.

Inputs

The respondent provides the following input to the customized design:

- $\mathbf{C_0}$, a vector with as many elements as attributes included in this respondent's BYO question, describing which levels were included in the BYO concept.

The analyst provides some inputs that control the design:

- **T**, the number of total product concepts to generate
- $\mathbf{A_{min}}$, the minimum number of attributes to vary from the BYO concept
- $\mathbf{A_{max}}$, the maximum number of attributes to vary from the BYO concept (restricted to be no more than 1/2 the number of attributes in the BYO question +1, not including Summed Price),
- If a "summed price" attribute is used, a range is provided specifying how much price should vary (randomly) from the summed components' price (e.g. 30% below to 20% above summed price).

The Design Algorithm

Near-orthogonal designs are generated using a controlled, randomized process. The steps involved in selecting each of **T** concepts in the design are as follows:

1. Randomly select an integer ($\mathbf{A_i}$) from $\mathbf{A_{min}}$ to $\mathbf{A_{max}}$ that specifies how many attributes within $\mathbf{C_0}$ will be modified to create new (near-neighbor) concept $\mathbf{C_i}$.
2. Randomly select $\mathbf{A_i}$ elements within $\mathbf{C_0}$ to modify.
3. Randomly select new (non-BYO selected) levels for the attributes chosen in step 2 (all other attributes remain at the BYO-selected levels).
4. Check to ensure that the concept chosen doesn't violate any prohibited pairs and is not a duplicate to another concept previously selected for this respondent. If prohibited or duplicate, discard the concept and return to step 1.
5. For non-BYO selected levels, examine whether relabeling levels to another non-BYO selected level within the same attribute improves the relative D-efficiency of the design for this respondent. Examine whether swapping non-BYO selected levels between two concepts improves the relative D-efficiency. Any relabeling or swapping that increases the efficiency while not making the target level count balance worse is accepted.

Issues Related to Level Balance and Efficiency

"Counts" arrays (at the individual level) are maintained to ensure that the designs have much greater balance than would be achieved if the above strategy involved purely random selections. A counts array keeps track of how many times each element has been selected or modified. For example, we maintain a counts array for each attribute that records how many times each level within that attribute (other than the BYO-selected level) has been included in the design. When a relative deficit occurs in the target frequency of selection, we increase the likelihood that the element with the deficit will be selected in the next concept. This allows for "controlled" randomized designs, and leads to a relatively high degree of level balance (but not *perfect* level balance). Our approach leads to a high degree of balance for: a) how many times $\mathbf{A_{min}}$ to $\mathbf{A_{max}}$ attributes were varied when generating **T** concepts, b) how many times each attribute was varied, and c) how many times each level (other than the BYO-specified level) was included across the **T** concepts.

From the standpoint of achieving maximum statistical efficiency, the optimal approach would vary all attributes independently, as in the traditional orthogonal array. By choosing only a narrow subset of the attributes to vary (from the BYO-specified concept) when creating each new concept, lower statistical efficiency results. However, ACBC's approach has three important benefits to counteract the loss in statistical efficiency: 1) Respondents can answer with less noise when fewer attributes are varying within each concept, 2) The concepts seem more relevant and plausible to the respondent, since they are near-neighbors to the BYO-specified product, and 3) The design concentrates on learning about preferences for the levels directly surrounding the respondent's highest levels of preference. In our previous methodological tests with about eight or nine total attributes, we have found that using 2 and 4 for $\mathbf{A_{min}}$ and $\mathbf{A_{max}}$, respectively, works quite well. We tried varying from 2 to 5 attributes when generating each concept in one of our experiments, and didn't see an increase in performance. However, the optimal amount of variation to use when generating concepts depends on the number of attributes in the study. Recent research published at the 2013 Sawtooth Software Conference (2013 Goodwin, 2013 Fotenos *et al.*) reported that the quality of the results are quite robust to how much variation from the BYO-selected concept is employed in generating new concepts for the respondent to evaluate.

Because the designs are adaptive, one cannot know for certain the design efficiency for any given respondent prior to fielding the study. However, ACBC software includes a Test Design module for simulating robotic respondents (who answer randomly). The D-efficiency for each robotic respondent’s design is reported as well as the standard errors for the parameter estimates across the sample (using aggregate logit).

Generating Replacement Cards

A key aim of ACBC questionnaires is to identify any levels that are "must haves" or "unacceptables." Rather than ask the respondent upfront regarding cutoff rules, we formulate hypotheses regarding what levels might be must-haves or unacceptables based on the respondent's observed choices within the Screener section. For example, if we notice that the respondent only selects as "a possibility" concepts featuring Brand A, we might suspect Brand A is a "must-have" level. After we have observed a pattern of such choices, we present a list of "must-have" or "unacceptable" rules that we suspect the respondent might be employing. If the respondent confirms a non-compensatory rule on the list, we then mark as "not a possibility" any concepts not yet evaluated that would fail to satisfy that rule. This leads to more efficient questionnaires and the opportunity to create "replacement cards" that probe deeper within the respondent's relevant preference space. For example, if after a respondent has evaluated the first 10 concepts (marking each as a "possibility" or "not a possibility"), the respondent verifies that Brand A is a "must-have" rule, this might eliminate from consideration 8 of the upcoming concepts (all featuring a brand other than Brand A) that we had planned to ask the respondent to evaluate. We generate 8 replacement concepts (all featuring Brand A) according to the design algorithm specified above, the exception being that the brand attribute will be removed from consideration as a potential selection in step 2 (above).

Analysis of ACBC Data

Counting different outcomes in ACBC data can provide useful insights. ACBC software lets you count:

- **BYO**: How often levels were included
- **Unacceptables**: How often levels were unacceptable
- **Must-Haves**: How often levels were must-haves
- **Screeners**: How many products were screened into the consideration set
- **Choice Tournament**: How often levels were included in the "winning" concept

The information in the core three sections of the ACBC questionnaires can be coded as a sequence of choice tasks and may be estimated using maximum likelihood estimation under the MNL model. By default, non-price attributes are coded as partworth functions using effects coding (a type of dummy coding procedure).

BYO

The choice of level for each attribute may be coded as a choice task, where respondents chose 1 of K levels (traded off with price if level prices are being used within the Summed price feature). Thus, if 10 total attributes (other than price) are included in the study, this section will contribute 10 tasks.

Screener Section

Each respondent has marked T concepts as "possibility" or "not a possibility." We treat these each as binary choices, where the respondent is assumed to compare the utility of the concept to the utility of a constant threshold prior to making a choice. The constant threshold is included in the design matrix as an additional dummy coded column.

Choice Tasks Tournament

These are coded as per standard CBC practice.

Modeling the Price Function

With ACBC, we have taken the approach of treating the price attribute as a "summed," continuous variable. With summed price, the survey author assigns component-based (level-based) prices (and optionally, a base price for the product). When we display the total price of the product (in the Screener, Choice Tasks, and Calibration sections), we add the prices for its levels (plus any optional base price), and then we vary the price randomly by a user-specified amount of random variation (e.g. -30% to +30%). This has the benefit of displaying overall prices that are more in line with the general quality of the products (for example, it avoids showing a poor product at relatively high prices). But, the drawback is that if price is not varied randomly (and independently) enough, strong enough correlations between price and product features can result and will lead to relatively imprecise estimates of both. In general, we recommend that overall price be varied by at least -30% to +30% to reduce such multicollinearity and ensure reasonably precise estimates. More guidance related to this can be found in the white paper: "Three Ways to Treat Overall Price in Conjoint Analysis" available in our Technical Papers library at www.sawtoothsoftware.com.

Because there are myriad combinations of attribute levels that can make up a product (and each level can have assigned component prices) and prices are varied randomly within a specified range of variation, thousands of unique "displayed" prices can be found across a sample of respondents. To model the effect of price on product choice, we have chosen three effective coding methods: *Linear*, *Log-Linear*, and *Piecewise*.

Linear Function: The total price shown to respondents is coded as a single column in the design matrix. This assumes that the impact of price on utility is a linear effect. (However, note that when projecting respondents' utilities to choices in the simulation module, one typically exponentiates the total utility for products, so the final impact of price on choice probabilities is modestly non-linear.) The benefit of linear coding is that only one parameter is estimated, conserving degrees of freedom. But, if the actual impact of price on utility is non-linear, the linear assumption for this model is simplistic and suboptimal.

Log-Linear Function: The total price shown to respondents is transformed by the natural log and coded as a single column in the design matrix. This assumes that the impact of price on utility is log-linear, or somewhat curved.

Piecewise Function: This is a more advanced and flexible function—and the approach we recommend for most datasets. The user provides specific breakpoints, and a different slope (linear function) is estimated to characterize the relationship of price to utility between each breakpoint.

Partworth Estimation

Partworths may be estimated using ACBC's built-in HB tool, though the data may be exported to a .CHO file for estimation using our Latent Class system. ACBC also offers a method of purely individual-level analysis of partworths called Monotone Regression (Johnson 1975).

By default, interaction effects are not estimated in ACBC's HB routines. But, ACBC designs (across people) do a fine job of supporting interaction effects and we encourage you to include selected first-order interactions (interactions between two attributes). An interaction search tool is available within ACBC software for investigating (using aggregate logit) the value of all potential first-order interaction effects. The search tool prioritizes these interaction effects for you from most to least additional fit (over the main effects model) in its report.

One of the benefits of traditional CBC modeling is the ability to include a "None" option. However, interpretation of the strength of the None parameter as well as appropriate application within market simulations have been difficult issues in CBC research. Adaptive CBC can estimate a None threshold using two approaches:

- None parameter estimated from Screening Tasks
- None parameter inferred from purchase intent scales via Calibration Concepts

One consideration when estimating partworth utilities from ACBC is the large difference in scale (response error) for the three main sections of the questionnaire. If using generic HB estimation, a key assumption is that the three types of tasks can be combined within the same dataset, even though we have observed that each section has different scale factor. Even though the BYO section has larger scale than the other two sections, we ignore this fact when using generic HB to estimate the parameters across the tasks.

We have wondered about the practical effect of the differences in scale for the three sections when using a generic HB model. We have been fortunate to benefit from the expertise of Dr. Thomas Otter, a recognized expert regarding HB application to choice experiments. Otter has built more sophisticated HB models that separately model scale for the three sections in ACBC questionnaires. He has found that the generic HB model performs very nearly as well as the more sophisticated models. You can turn on "Otter's Method" for accounting for differences in scale during HB estimation through the HB interface provided in ACBC. In our experience, it takes about 300 or more respondents to develop stabile estimates of differential scale in ACBC via Otter's method.

ACBC: How Well Does It Work?

To this point, Sawtooth Software has conducted four fairly ambitious methodological studies comparing ACBC to standard CBC[2]. Because we have documented the results of those experiments in three white papers published on our website, we refer the reader to those articles and just cover the highlights here.

To read more about the details of our studies, we recommend you download and read the following white papers, available from our Technical Papers Library at www.sawtoothsoftware.com:

- "A New Approach to Adaptive CBC"
- "Testing Adaptive CBC: Shorter Questionnaires and BYO vs. 'Most Likelies'"
- "Fine-Tuning Adaptive CBC Questionnaires"

Judging Success

The critical design issue when comparing one conjoint method to another is the criteria for success. Ideally, the researcher would have access to sales data (or subsequent actual choices of respondents) and could compare predicted choices with actual choices.

In the absence of data on real-world choices, many researchers turn to holdout choice tasks included within stated preference surveys. Traditionally, these have looked exactly like standard CBC questions: typically 3 to 5 product concepts. One of the key things we and other researchers have learned about standard CBC tasks is that respondents often key on just a few levels to make decisions. They often establish a few cutoff rules for excluding or including concepts within their consideration sets. And, if one uses standard CBC tasks that employ "minimal overlap" (where each level is typically available no more than once per task), often only one product concept can satisfy respondents. Choices to such tasks often reflect simplified (superficial) behavior, and other choice tasks designed in the same way not surprisingly are quite successful in predicting answers to those holdouts. We have found that ACBC does about as well as CBC (sometimes better, sometimes worse) in predicting these kinds of holdout CBC tasks. And, that parity doesn't concern us, as we'll further explain.

We have wondered whether traditional holdout tasks really do a good job in mimicking actual purchase behavior, or whether they reflect simplified processing heuristics that are a byproduct of respondents (especially internet sample) completing marketing research surveys with less motivation than they would apply to real-world purchases. When people make real decisions, they often narrow down the choices to an acceptable consideration set (with respect to must-have and must-avoid features) and then make a final

---

[2] Others outside Sawtooth Software have also conducted comparison studies between CBC and ACBC, notably Chapman 2009 and Fotenos et al. 2013.

choice within the consideration set. To better mimic that process in the holdouts, our first three methodological studies used a customized type of holdout CBC task that involved comparing winning concepts chosen from previous CBC tasks. For example, the respondent might be shown four standard (fixed) CBC holdout tasks, but the fifth holdout included the four winning concepts from the first four tasks. Such customized tasks lead to questions that can probe beyond just the first few critical levels of inclusion and reflect more challenging tradeoffs. We have found that ACBC consistently predicts these choices more accurately than traditional CBC.

Some researchers (such as Elrod 2001) have stressed that if we validate to holdout tasks within survey questionnaires, not only should the holdout tasks be excluded from the calibration tasks used for estimating the utilities, but a sample of holdout respondents should be held out as well. For example, we should be using calibration tasks from respondent group A to predict holdout tasks from respondent group B (and the actual tasks used in group B should not have been used in the experimental design for group A). When this is done, one does not compute individual-level hit rates (since it is no longer a within-respondent analysis), but one uses market simulations and measures share prediction accuracy. Simulations of market choices for group A are used to predict choices to fixed holdouts for group B.

We were fortunate enough to have a robust enough sample in our first ACBC methodological study to follow Elrod's recommendations. In addition to the 600 respondents in the sample used to estimate partworth utilities, we were able to collect a separate sample of 900 respondents who all received a single-version (fixed) set 12 CBC tasks. We found that ACBC predicted shares of choice for these scenarios just as well as CBC. But, when we split the 900 respondents into three equal-sized groups based on interview time, we found that ACBC performed worse than CBC in predicting the choices for respondents who were the quickest but better than CBC for respondents that took the most time (and presumably were providing more in-depth and careful answers).

We mention these conclusions and findings to provide direction to researchers who may want to test the merits of ACBC relative to more traditional methods like CBC. If standard holdout CBC tasks are used as the criterion for success, there is methods bias that strongly favors CBC in predicting those tasks. If respondents are using the same (probably unrealistic) simplification heuristics to answer calibration CBC tasks, they will employ the same heuristics in answering holdouts. One needs to consider carefully the appropriate criteria for success. You may find that ACBC does not consistently beat CBC in predicting static minimal-overlap holdout choice tasks, for the reasons outlined above. It occurs to us, however, that if static choice tasks include a greater number of concepts than are traditionally used, more overlap will be introduced, increasing the likelihood that respondents will need to make a more careful decision regarding multiple acceptable concepts per task prior to making a decision. In those cases (which are probably more representative of real choice situations), ACBC may very well consistently beat CBC.

We suspect that ACBC will be more predictive of real-world buying behavior than CBC. One study published by Chapman et al. (2009) supports this. But, we need practitioners and firms with access to the appropriate sales data to conduct additional comparisons. This is an exciting new area of research, and we are optimistic that ACBC will indeed prove better than traditional CBC for complex conjoint problems involving about five or more attributes.

Previous Results and Conclusions

To date, we have completed four substantial methodological studies comparing ACBC to CBC. The study names, sample sizes, and numbers of attributes are as follows:

- Laptop PC Study (n~600 calibration respondents; n~900 holdout respondents; 10 attributes)
- Recreational Equipment Study (n~900; 8 attributes)
- Home Purchases study (n~1200; 10 attributes)
- Fast-Food Drive-Through (n~650; 4 attributes)

The experiments all used web-based interviewing. In the case of the Laptop, Home Purchase, and Drive-Through studies, we used Western Wats panel (since purchased by SSI). For the Recreational Equipment study, customer sample was provided by the client. For the Laptop, Home Purchase, and Fast-Food Drive-Through studies, we implemented a carefully controlled random split-sample design, where respondents were randomly assigned to receive ACBC or CBC questionnaires, or assigned to be holdout respondents. With the Recreational Equipment study, because of conditions related to working within the client's timeline, we weren't able to conduct a perfectly controlled random assignment of respondents to different conditions, and the sample pull receiving the ACBC survey had potentially different characteristics from those receiving CBC. This weakens our ability to draw firm conclusions from that study.

Interview Time

The ACBC questionnaires generally took longer than the CBC questionnaires to complete (50% longer in the Housing survey, 100% longer in the Laptop study, and 200% longer in the Recreational Equipment study). We suspect that the Recreational Equipment study was relatively longer because it interviewed respondents from a customer list, who may have been even more enthusiastic and engaged in the process. The Housing survey was relatively shorter, because we investigated whether we could shorten the survey length without sacrificing much in terms of predictive validity (see results below, which show that even the shortened ACBC surveys were successful). In the Drive-Through survey, we purposefully made the CBC questionnaire quite long (24 tasks) so as to better time-equalize the tasks between ACBC and CBC. We actually overshot the mark slightly, and the CBC questionnaire turned out to take 25% *longer* than the parallel ACBC questionnaire.

A key finding from the first three studies is that even though ACBC respondents were asked to do a longer task, they rated the survey as more interesting and engaging than respondents taking CBC surveys rated the CBC surveys. Our thought is that when collecting conjoint data, speed alone shouldn't be the goal, as we cannot ignore quality. And, if we can improve the survey experience for the respondent even though we are asking them to spend more time, then the additional investment to acquire better data is well worth the effort.

Interest/Satisfaction with the Survey

For the Laptop and Home Purchases studies (which provide the strongest results due to the controlled experiments), the ACBC respondents reported higher satisfaction and interest in the survey than CBC respondents. They reported that the survey was more realistic, less monotonous/boring, and that it made them more likely to want to slow down and make careful choices than the CBC respondents reported. All these differences were significant at $p<0.05$. The Drive-Through study did not include these qualitative evaluation questions.

Hit Rates

In the two studies involving random assignment to either ACBC or CBC cells (and with a custom holdout task), the ACBC respondents had higher hit rates than CBC respondents in two cases, with differences significant at $p<0.05$:

Hit Rates, Custom Holdout Task		
	ACBC	CBC
Laptop Study:	61	50
Houses Study:	44	37

For the Recreational Equipment study (which lacked a completely random split-sample control), ACBC had a 2-point edge in hit rate prediction over CBC, but the result was not statistically significant. For the Drive-Through study, we included a standard (non-customized CBC holdout task), and the hit rate was tied between CBC and ACBC. All these figures reflect the generic HB estimation method.

Some researchers have wondered whether the edge in predictive ability demonstrated by ACBC over CBC in predicting custom holdout tasks might be accounted for by the longer questionnaire time for ACBC. Maybe ACBC's success is due to more precise estimates because we've asked more questions of each respondent. While this seems plausible, we doubt that this can entirely explain the results, because it has been demonstrated that hit rates for CBC respondents gain very little beyond about 10 to 15 choice tasks (and the CBC surveys for the Laptop and Houses studies both included 18 tasks). We come to this conclusion based on a meta-study of 35 commercial data sets by Hoogerbrugge and van der Wagt of SKIM Analytical as presented at the 2006 Sawtooth Software Conference. They randomly threw out one task at a time and estimated the loss in hit rate precision depending on questionnaire length. Based on their research, in the range of 10 to 15 choice tasks, the gains in hit rate rapidly decreased and in some cases leveled off through up to 19 tasks. The net pattern and absolute magnitude of the gains (observed across 35 data sets) by extending the questionnaire would not seem to be able to eventually make up the gaps of 11 and 7 absolute hit rate points in favor of ACBC as seen in our two experiments.

Share Predictions

In the Laptop study, we were able to estimate shares for holdout CBC tasks completed by a new sample of 900 respondents. The 900 new respondents completed 12 fixed CBC tasks. The partworths generated from ACBC performed just as well as those from CBC in predicting these holdout shares, when including all respondents. But, we split the 900 respondents into three equal samples, based on time taken in the interview. After segmenting based on time, we found that CBC predicted the fastest respondents better than ACBC, but ACBC predicted the middle and slower respondent groups better than CBC (with the best prediction occurring for the holdouts completed by the slowest respondents). In each case, we tuned the simulations for scale, so the differences cannot be explained by the random error components (which were significantly different for the three segments of respondents). Also, a Swait-Louviere test (which also controls for scale) found differences in the parameters across the groups of holdout respondents. This finding suggests that ACBC can do a better job at predicting choices made by respondents who take the most time and presumably are being more careful and considerate in their choices.

For the Houses Study, we did not collect holdout respondents. However, we were able to use the partworths from the ACBC and CBC respondents to predict choices made by the two groups of respondents combined (a total of 1200 respondents in the validation sample). Thus, ACBC partworths were being used to predict holdout responses made by both ACBC and CBC respondents; and CBC respondents the same. Based on predictions of four fixed holdout tasks (each with four product alternatives), the ACBC predictions were better than those from CBC (the mean absolute errors in prediction were 3.00 and 5.47 share points for ACBC and CBC, respectively).

With the Drive-Through study, we purposefully made the CBC questionnaire quite long (24 tasks) to roughly time-equalize the ACBC and CBC tasks. Only 4 attributes were included in the study. The mean absolute errors in holdout prediction were 2.02 and 2.04 for ACBC and CBC, respectively.

External Validity

External validity is a more stringent (and more valuable) test of predictive validity than internal hit rates. For the first study reported by Johnson and Orme, they interviewed an additional 900 respondents that were used only for predictive validity purposes. These holdout respondents completed 12 standard CBC tasks, and their data were not used to estimate partworth utilities for the market simulation models.

Orme and Johnson found that the ACBC model was able to predict shares of choice for the holdout respondents as well overall as standard CBC. But, when the holdout respondents were divided into three groups based on time spent completing the questionnaire, the authors found that ACBC did better at predicting the answers from the respondents who took the most time, and worse than CBC in predicting the answers for respondents who hurried.

These results seem consistent with the hypothesis that many CBC respondents 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, the authors tested whether slow and fast responders to the 12 holdout questions differed significantly with respect to partworth utilities estimated from those 12 tasks. The test for difference in parameters was strongly significant. This last finding is important, because it suggests that adaptive CBC is able to capture more careful and *different* decision-making than standard CBC.

A validation test was reported at the 2009 Sawtooth Software Conference by Chris Chapman of Microsoft Corporation (Chapman *et al.* 2009). He compared standard CBC to Adaptive CBC predictions for sales of a new PC peripheral device. The predictions for ACBC were slightly better than CBC. At that same conference, Tang *et al.* also presented research comparing CBC to ACBC (Tang *et al.* 2009). Their study focused on preferences and price sensitivity for features of mobile internet devices. They used a willingness-to-pay (WTP) methodology to estimate WTP for features of these devices. They found ACBC results to provide WTP estimates that were more realistic and in line with established industry prices than CBC.

Future Directions for Research

ACBC is a promising new technique, and it opens a variety of new avenues for research. Highest on the list is whether ACBC generally works better than other conjoint/choice methods for predicting actual buyer choices. The methodological research so far suggests this is the case and Chapman's study finds a modest edge for ACBC, but additional work is needed.

With the latest release of ACBC (within SSI Web v8.3), researchers have more flexibility regarding whether to use or skip the BYO, Screening, or Choice Tournament sections. One investigation (Fotenos *et al.* 2013) suggests that ACBC can do just as well without the Screening tasks. Other researchers will certainly test results when skipping both the BYO and Screening tasks (relying only on the choice tournament tasks). The early evidence suggests ACBC will be robust to different questionnaire design decisions, which would give researchers great latitude to fit ACBC to their particular research needs for given research contexts and specific respondent populations.

How well ACBC can work on small screen mobile devices is another current topic. Diener *et al.* fielded a complex conjoint design with nine attributes on mobile devices as well as regular laptops/desktops (Diener *et al.* 2013). They programmed multiple versions of the questionnaire, including CBC, partial-profile CBC, and a shortened version of ACBC which included prior prescreening of attributes to drop those of little importance via constructed lists, a BYO section, an abbreviated Screening section, and skipping the Tournament section. Their abbreviated ACBC survey worked well on the mobile devices, with quality results in terms of part-worth estimation and holdout hit rate accuracy, suggesting that even complex conjoint studies are possible on mobile devices using ACBC. More research is needed, but this is a promising development.

References

Binner, Stefan, Richard Neggers, and Marco Hoogerbrugge (2009), "ACBC: A Case Study," presented at the joint SKIM/Sawtooth Software Training Event, Prague, Czech Rupublic.

Chapman, Chris, James Alford, Chad Johnson, Ron Weidemann, and Michal Lahav (2009), "CBC vs. ACBC: Comparing Results with Real Product Selection," Sawtooth Software Conference Proceedings, Sequim, WA.

Diener, Chris, Rajat Narang, Mohit Shant, Hem Chander, and Mukul Goyal (2013), "Making Conjoint Mobile: Adapting Conjoint to the Mobile Phenomenon," Sawtooth Software Conference Proceedings, Orem, UT.

Elrod, Terry. (2001), "Recommendations for Validation of Choice Models," Sawtooth Software Conference Proceedings, Sequim, WA.

Fotenos, Christopher (2013), Jeroen Hardon, and Marco Hoogerbrugge, "ACBC Revisited," Sawtooth Software Conference Proceedings, Orem, UT.

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

Goodwin, Robert (2013), "Quantitative Marketing Research Solutions in a Traditional Manufacturing Firm: Update and Case Study," Sawtooth Software Conference Proceedings, Orem, UT.

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, M. and K. van der Wagt (2006), "How Many Choice Tasks Should We Ask?" Sawtooth Software Conference Proceedings, Sequim, WA.

Johnson, Rich (1975) "A Simple Method of Pairwise Monotone Regression," Psychometrika, 1975, pp 163-168.

Johnson, Rich and Orme, Bryan (2007), "A New Approach to Adaptive CBC," Sawtooth Software Conference Proceedings, Sequim, WA.

Orme, Bryan (2009), "Fine-Tuning CBC and Adaptive CBC Questionnaires," Technical Paper Available at www.sawtoothsoftware.com.

Orme, Bryan and Rich Johnson (2008), "Testing Adaptive CBC: Shorter Questionnaires and BYO vs. 'Most Likelies'," Technical Paper Available at www.sawtoothsoftware.com.

Otter, Thomas (2007), "HB Analysis for Multi-Format Adaptive CBC," Sawtooth Software Conference Proceedings, Sequim, WA.

Tang, Jane, Andrew Grenville, Vicki Morwitz, Amitav Chakravarti, Gulden Ulkumen (2009), "Influencing Feature Price Tradeoffs Decisions in CBC Experiments," Sawtooth Software Conference Proceedings, Sequim, WA.