Sawtooth Software

RESEARCH PAPER SERIES

The "Importance" Question in ACA: Can It Be Omitted?

W. Christopher King, Aaron Hill and Bryan Orme,
Sawtooth Software, Inc.

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

The "Importance" Question in ACA: Can It Be Omitted?

W. Christopher King,
Aaron Hill,
and Bryan Orme,
Sawtooth Software, Inc.

Background

Adaptive Conjoint Analysis (ACA) was developed by Sawtooth Software's founder, Rich Johnson, based on his early work in tradeoff matrices dating back to the late 1960s. During the 1970s, Rich and his colleagues refined the technique by applying early versions of ACA to various real-world business problems. In the late 1970s, the first versions of ACA were proprietary and developed on the Apple II platform. In 1983, Rich founded Sawtooth Software and released ACA as a commercial software product in 1985. ACA soon became the most widely-used conjoint analysis program and technique in the world (Wittink and Cattin, 1989). It held that position until about the year 2000, when choice-based conjoint (CBC) techniques became more widely used (Sawtooth Software, 2003a). Even still, ACA is a widely used method for studying more attributes than is prudent for full-profile approaches. It is especially valuable for product design research and segmentation studies. In this paper, we assume the reader is already familiar with the basic mechanics of ACA. For more information, please see the "ACA Technical Paper" available for download at www.sawtoothsoftware.com in the Technical Papers Library.

ACA has benefited from scrutiny over the years, particularly from leading academics. In a paper published in 1991 (Green, Krieger, and Agarwal, 1991), the authors criticized ACA's self-explicated "importance" question. They argued that the attribute importance ratings employed in ACA's "priors" section were "…too coarse; only four response values [were] permitted." Later that year, William McLaughlan conducted a split-sample research study that compared a 9-pt and a 4-pt importance question within ACA, finding virtually no difference in the quality of the final part worth utilities (McLaughlan, 1991). Green, Krieger, and Agarwal also questioned the scale compatibility of self-explicated importance ratings and the subsequent conjoint (pairwise) judgments. They argued that it was not proper to combine both types of responses (as dependent variables) within the same regression tableau. Based on that suggestion, Sawtooth Software modified the way ACA utilities were estimated under OLS to address these potential scale incompatibilities. The results were quite similar to the earlier estimates, though time has shown that neither utility estimation procedure clearly dominates. As an unexpected benefit, it was later discovered that the newer OLS method seemed to slightly reduce the number of level effects problem (Orme, 1998). Since the late 1990s, we at Sawtooth Software have recommended hierarchical Bayes (HB) estimation as the gold standard for ACA, though the modified OLS procedure, nearly identical to that developed by Johnson in the early 1990s, remains the default option in the base system.

One prevalent criticism of ACA has been that the self-explicated importance judgments probably reduce the discrimination among attributes, leading to "flatter" importances in the final utility estimates. Furthermore, many researchers have reported that respondents can have difficulty understanding the importance questions, and have a tendency to over-use the upper portion of the rating scale. Given the availability of HB and its proven superiority for estimating quality part worth utilities using less information from each individual (due to its powerful information-borrowing mechanism), we have wondered whether we might achieve equally good (or even better) results by skipping the importance questions altogether. This research seeks to answer this question. As will be shown, it appears for most research applications that the importance question *can* be skipped (if applying HB estimation), resulting in shorter questionnaires and less confusion and burden for respondents.

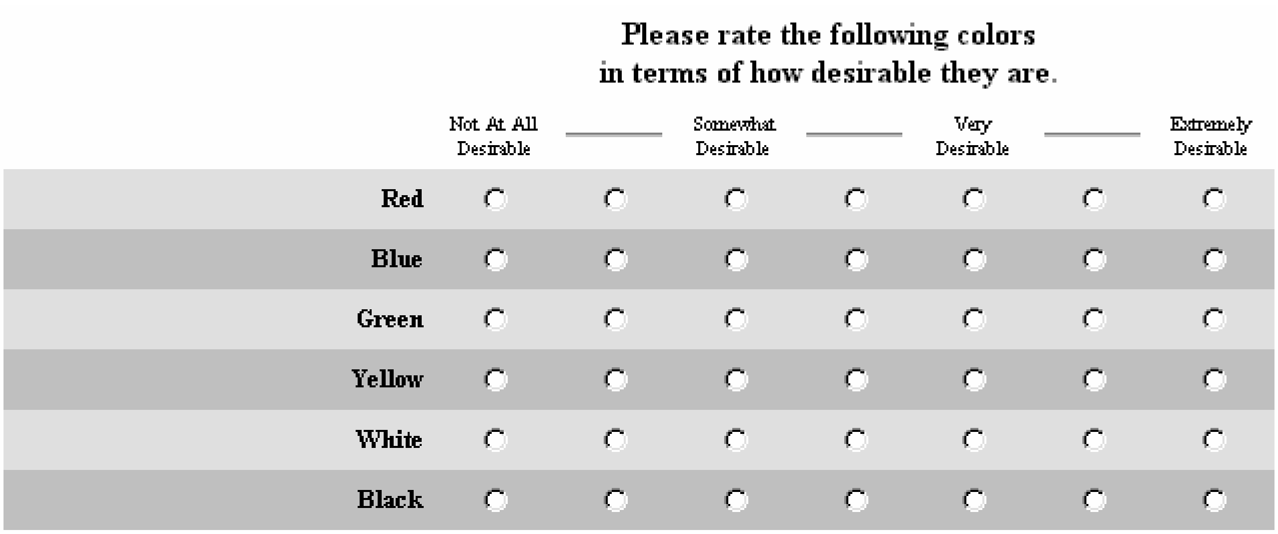
Parts of the ACA Interview

ACA is a multi-stage hybrid conjoint instrument that features the following sections:

1. Unacceptables (optional, and rarely used)
2. Rating/Ranking of levels <u>within</u> attributes
3. Importance ratings of attributes
4. Pairs (the "conjoint" section)
5. Calibration Concepts (optional section)

The first three sections represent a simple self-explicated scaling exercise and are often called the "Priors." The first section allows respondents to indicate which levels of an attribute are "unacceptable," even if these levels were included in a product that was excellent in all other ways. This section is not available in ACA/Web and is rarely used, so we skipped it for the purposes of this study. In the second section, the respondent is asked to rate/rank the levels of any attributes for which we cannot know ahead of time the rational order of preference (such as brand, colors, etc.). Figure 1 is an example of a ratings style question.

Figure 1

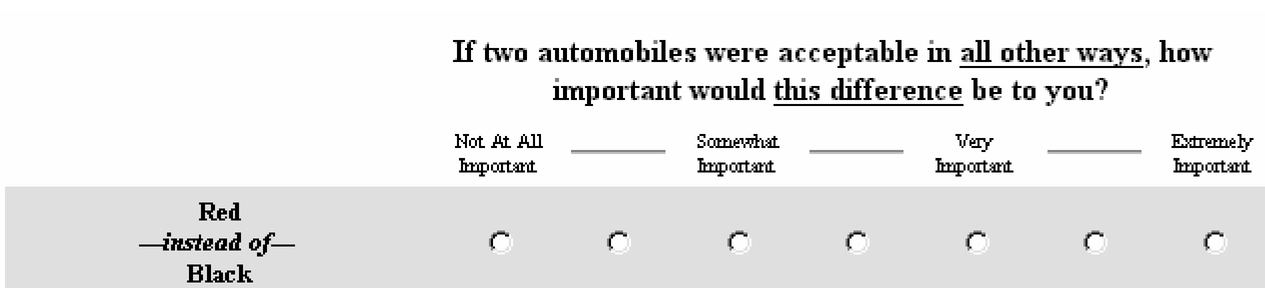


Please rate the following colors in terms of how desirable they are.

	Not At All Desirable		Somewhat Desirable		Very Desirable		Extremely Desirable
Red	○	○	○	○	○	○	○
Blue	○	○	○	○	○	○	○
Green	○	○	○	○	○	○	○
Yellow	○	○	○	○	○	○	○
White	○	○	○	○	○	○	○
Black	○	○	○	○	○	○	○

Next follow the Importance questions, which are the focus of this research. After the first ranking/ratings section (or based on other pre-specified attributes where we know ahead of time the rational order of preference), we display the best and worst levels for each attribute. We ask respondents to indicate how important it is for them to receive the best level instead of the worst level (there are other ways to phrase this question). Figure 2 shows an example, assuming the respondent has rated Red as the most preferred level and Black as the least preferred:

Figure 2



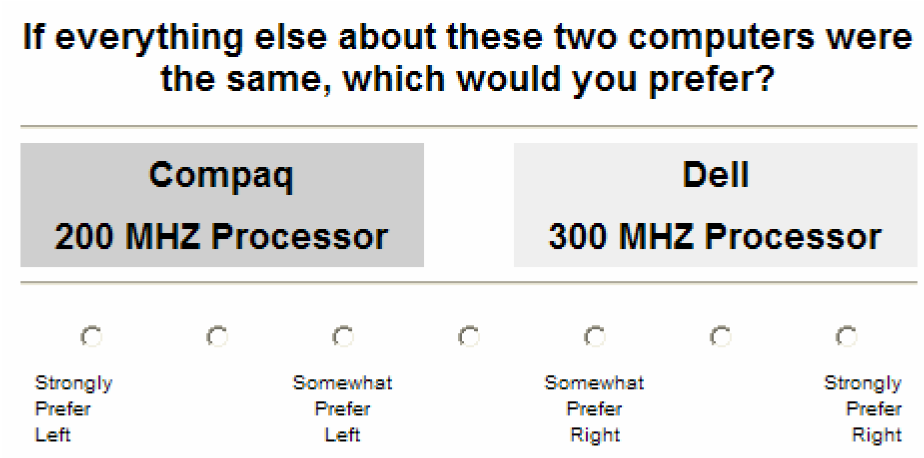
If two automobiles were acceptable in all other ways, how important would this difference be to you?

	Not At All Important		Somewhat Important		Very Important		Extremely Important
Red *—instead of—* **Black**	○	○	○	○	○	○	○

The importance rating is used to prioritize the order that attributes are investigated in the conjoint "Pairs" section. The metric information from this rating scale is used to develop the rough "prior" utilities used for the utility balancing in ACA's Pairs design algorithm, and also in the final utility estimates. The importance question is currently a *required element* in the ACA survey, though we investigate an experimental version of ACA that drops this section.

To this point in the ACA interview, no "conjoint" questions have been asked. Figure 3 shows an example conjoint pair, which is the workhorse of the ACA interview and that provides the critical tradeoff information needed for refining the part worth utilities. After each pair is asked, the information augments that previously collected (updates the prior utilities), and is used to select the next question asked of the respondent (respondents typically answer 12 to 30 Pairs).

Figure 3



After the Pairs section, an optional Calibration Concepts section may be included to gauge each respondent's purchase likelihood for different product concepts. Purchase Likelihood judgments were not important to this research, so we skipped this final section.

Concerns Regarding the Importance Question

Again, researchers have expressed a number of valid concerns regarding the Importance question, including:

- For some attributes, it is difficult to understand what is being asked—especially a binary attribute which either has or doesn't have a feature.
- Researchers sometimes have difficulty with formatting and wording.
- Lack of context can become an issue, particularly with attributes exploring sensitive topics (i.e. salary).
- Respondents do not answer in a truly ratio-scaled manner (a rating of "4" does not mean that the attribute is twice as important as one rated a "2"), yet the standard OLS utility estimation routine in ACA assumes we have captured ratio data.
- Respondents don't discriminate enough among the attributes using the provided rating scale. The responses typically load on the upper part of the scale.
- In studies with a large number of attributes, the importance questions can become tedious. (Perhaps a respondent's time would be better spent answering more Pairs questions. The advent of HB potentially removes the need for individual self-explicated importance information.)

- Importance scores may artificially constrain and usually do flatten the differences among attributes.
- It is difficult for respondents to establish the proper scaling (framing) for the first few attributes (though this problem might be reduced by first reviewing the list of attributes prior to starting the ACA process).

In their as-of-yet unpublished work "Adventures in Conjoint Analysis" (Krieger, Green, Wind, 2004), the authors state:

> "Considerable research, both conceptual and empirical, suggests that a respondent's self explicated attribute level desirability ratings are much more reliable than self explicated attribute importances.
>
> "A recent examination of reliability in a conjoint setting indicated that the median test/retest correlation for attribute-level desirabilities was 0.925, while that for importance ratings was only 0.574."

Alternatives to the Importance Question

- Initially, set all importance scores equal (for design purposes), but let the final part worths be estimated using level ratings and conjoint Pairs data.
- Use starting importance scores (for design purposes) based on the average of previous respondents' final derived importance values.
- Bayesian updating routine.
- The researcher might establish importance scores (for design purposes) based on previous research, focus groups, sales data, or client whim.
- Derive importances using the observed spread among self explicated ratings within each attribute.
- Use a different conjoint method (partial-profile choice?)

Research Study

Because of the confidentiality of the information, we cannot disclose the sponsor of this research or details regarding the product or features. We can say that the purpose of the research was to compare accessibility and content issues related to an existing electronic information service over the internet. We studied 20 attributes, where roughly half were binary (had two levels), and the others ranged in complexity to a maximum of five levels per attribute. None of the attributes had *a priori* level order, so respondents had to provide level ratings for all 20 attributes in the study. Graphics were used to reflect the attribute levels to make the content (which at times was quite complex) more accessible to the respondents.

We randomly split respondents among three design cells. The study was conducted over the internet, and respondents were recruited in a variety of ways, including an opt-in list and banner ads, according to the client's specifications. A total of 1,419 respondents were used in the final analysis, after data cleaning. Respondents were discarded if they completed the survey too quickly (fastest 5%), if their Pairs ratings showed little discrimination (>=85% identical), or if they did not complete the holdout tasks. Less than 15% of the data was discarded.

Respondents in each cell received different ACA treatments. The three design cells were as follows:

> **Cell 1 (n=463):** Standard ACA procedure, as currently available in ACA/Web v5. The questionnaire included 20 level ratings grid questions (one for each attribute), 20 Importance questions, and 14 Pairs questions (6 at 2 attributes and 8 at 3 attributes).
>
> **Cell 2 (n=508):** Modified ACA where all importance ratings were set to be equal (for Pairs design purposes). **The Importance section was skipped**, and final part worths were developed using HB estimation (fitting the Pairs, and constrained by within-attribute level ratings). Updating of part worths during the Pairs section used the standard ACA procedure. Attribute importance was therefore recalculated after each Pairs question, so the initial importances had less influence as the survey progressed. The questionnaire included 20 level ratings grid questions (one for each attribute), no Importance questions, and 20 Pairs questions (8 at 2 attributes and 12 at 3 attributes).
>
> **Cell 3 (n=448):** Modified ACA where the importance ratings were projected for each respondent based on the derived importances from respondents that had already completed the survey. All other aspects were identical to Cell 2.

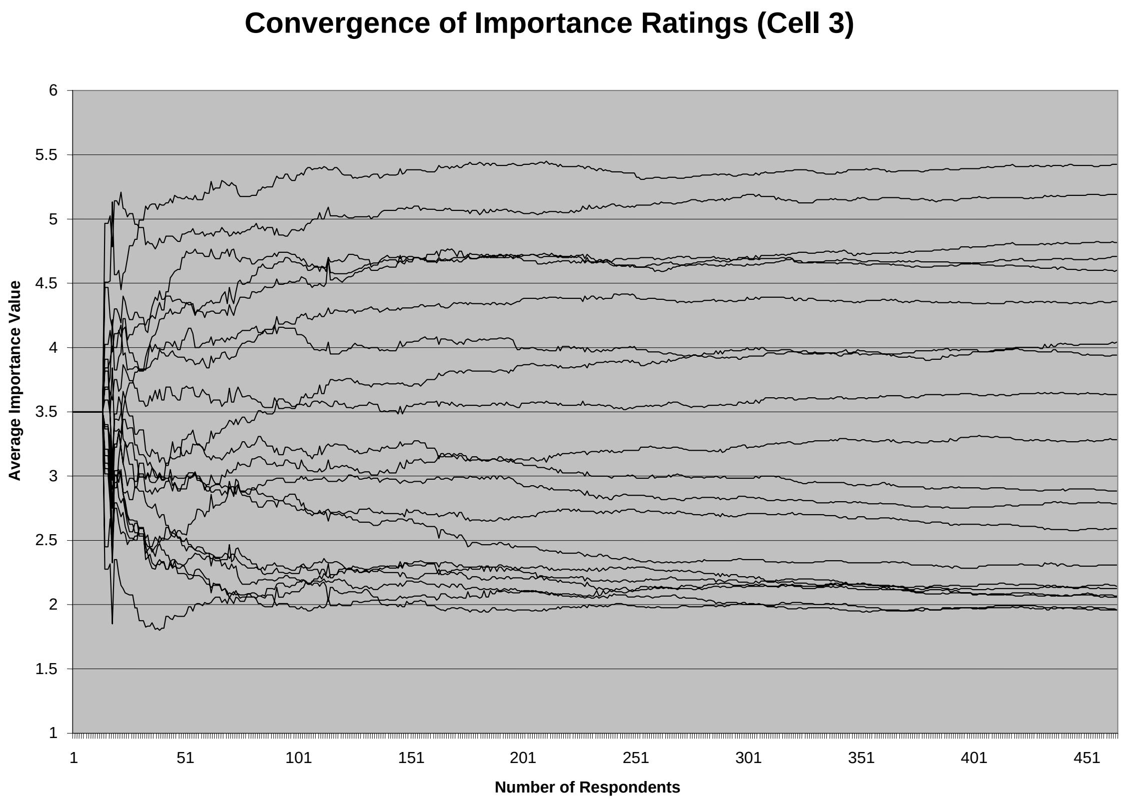
In addition, we included partial-profile holdout choice questions in the survey. After completing the ACA sections, each respondent answered six of a total of eighteen possible holdout tasks (selected randomly). Each task included 4 of the 20 attributes, with 3 alternatives per task. Across the tasks, all 20 attributes and all levels were represented. To conserve time, no Calibration Concepts were asked (Hill, Baker & Pilon, 2001, found that calibrating utilities did not help share predictions).

Prior to fielding the study, we hypothesized that the Importance questions might not be very useful, and might actually misinform utility estimation. We hypothesized that the time savings from skipping the Importances could be used more profitably in asking additional conjoint Pairs questions. We expected that the importances based on prior respondent averages would provide reasonable starting points for design purposes in the Pairs. We also hypothesized that skipping the Importances would lead to higher completion rates.

Convergence of Attribute Importances for Cell 3

As previously described, Cell 3 of our design used the utility information from previous respondents as initial importance scores for design purposes for the conjoint Pairs section. Because there was no information available when the first respondent entered the survey, we initially set the importances equal. After 12 respondents, we began to use the simple average of all previous respondents to initialize the design procedure for the Pairs algorithm (for the next respondent). We were curious to know how quickly the averages would stabilize for the respondent population. Figure 4 displays the average importance ratings for all 20 attributes by number of respondents who had taken the survey.

Figure 4



As can be seen, the average importance scores stabilized quite rapidly. After 50 respondents had completed the survey, only minor switches in the rank order of attribute importances can be seen. We should note that we used the prior group importances in a deterministic manner for arranging the attributes in rank-order (for purposes of the round-robin sequence used in designing Pairs with two attributes at a time). In hindsight, we recognize that we may have done slightly better by using the prior importances in a probabilistic way for establishing the initial rank-order of attributes for these first Pairs questions. This would have lead to greater variation in the design for the first Pairs (shown at two attributes at a time) across respondents.

Time to Complete and Completion Rates

Because we were interviewing using computers, we were able to capture the time (in seconds) to complete each section of the survey. As shown in Figure 5, it took 2.6 minutes for respondents in Cell 1 to complete the 20 Importance questions. Cells 2 and 3 omit the Importance questions and substitute six additional pairs. We can further see in

Figure 5 that the overall interview time for the ACA section was *lower* for Cells 2 and 3 relative to Cell 1 (though the Pairs section took slightly longer to complete). We could have substituted even more conjoint Pairs questions to match the interview time of Cell 1.

Figure 5

(in Minutes)	**Median Overall Survey Completion Time**	**Median time per ACA section** Ratings	Importance	Pairs	Total ACA
Cell 1 (14 pairs)	23.0	6.5	2.6	2.9	11.9
Cell 2 (20 pairs)	21.4	6.5	0.0	4.0	10.5
Cell 3 (20 pairs)	20.9	6.6	0.0	4.2	10.8

More than half of the people that started the survey ended up quitting at some point (there was no monetary incentive to complete the survey). This figure is disappointing. The percent of respondents that completed the survey once they started were 39.6%, 42.9% and 42.6% for Cells 1, 2, & 3 respectively. The shorter interviews (Cells 2 and 3) indeed had directionally higher completion rates, but the margin of difference is relatively small.

Part Worth Utility Analysis

We used the following methods to estimate the part worth utilities.

> **ACA/OLS:** Standard OLS estimation, as offered in ACA (only applicable for Cell 1 respondents).
>
> **ACA/HB with Importance Scores:** ACA/HB procedure, fitting the metric information of the conjoint pairs, constrained by both the importance scores and the within-level ratings (only applicable for Cell 1 respondents).
>
> **ACA/HB without Importance Scores:** ACA/HB procedure, fitting the metric information of the conjoint pairs, constrained only by the within-level ratings*.

(Some ACA/HB users may wonder how we employed ordinal constraints using only the within-attribute ratings. To do this, we modified the .ACD file, and artificially set the importance scores to be equal for all 20 attributes. This "trick" has worked well for us in a few data sets, where the self-explicated importances seemed suspect.)

Findings—Hit Rates

As previously explained, respondents completed six (of eighteen) choice tasks including three alternatives each (shown in partial profile). We asked respondents to indicate the favored and least favored alternative within each task (a best/worst approach).

Using the part worth utilities, we predicted which product alternatives we would have expected respondents to choose and reject within the holdout choice tasks. We scored each product alternative (by adding the part worth utilities corresponding to the levels included in that alternative) and projected that the respondent would choose the alternative with the highest utility and reject the alternative with the lowest utility. We compared predicted choices to actual choices, and scored the "hits" and "misses." The hit rates, by design cell and different utility estimation methods, are given in Figure 6.

Figure 6

Hit Rates	**Cell 1**	**Cell 2**	**Cell 3**
ACA/OLS	67.1 (0.63)	N/A	N/A
ACA/HB with Importance scores	67.9 (0.64)	N/A	N/A
ACA/HB without Importance scores	66.7 (0.61)	66.2 (0.61)	65.6 (0.66)

(Standard errors are shown in parentheses. A difference of 1.7% in hit rate is required between groups for significance at 95% confidence.)

The best hit rate occurs for traditional ACA (Cell 1) with ACA/HB estimation, using the importance scores and HB estimation. However, this is only marginally better than the design cells that omitted the Importance question (and substituted six more Pairs questions). Three key points occur to us: 1) Omitting the importances and substituting six more pairs (an overall time savings) results in nearly the same hit rates, 2) Previous research suggests that Importance ratings reduce measurement error within each individual at the expense of bias, which often benefits hit rates, 3) Hit rates are not as important to managers as accuracy of share predictions, which we'll report in the next section.

Findings—Share Predictions

As mentioned previously, share prediction accuracy is of most importance to managers. To compute share prediction accuracy, we used the Randomized First Choice simulation model as offered in Sawtooth Software's market simulator (though the standard logit simulation approach would produce very similar findings). We used part worth utilities for the respondents in each cell to predict the overall shares of first choices for all respondents and all eighteen holdout choice tasks. This, importantly, is a more stringent test of internal predictive validity, as it holds out both choice tasks *and* respondents. We tuned the shares (using the Exponent, or scale factor) to achieve highest predictive accuracy within each cell. This ensures that observed differences in predictive accuracy are due to substantive differences in the utilities rather than arbitrary differences in scale.

Figure 7 reports the predictive accuracy in terms of Mean Absolute Error (MAE). MAE characterizes by how much (on average) the predicted shares differ from the actual choice shares. Lower MAE values indicate more accurate predictions.

Figure 7

Share Predictions (MAE)	**Cell 1**	**Cell 2**	**Cell 3**
ACA/OLS	7.1%	N/A	N/A
ACA/HB with Importance scores	6.8%	N/A	N/A
ACA/HB without Importance scores	6.7%	6.4%	5.9%

Cells 2 and 3 (which omitted the Importance questions and substituted six additional Pairs) show the best predictive accuracy. The additional Pairs questions seem to benefit this analysis, and the lack of individualized importance information for designing the Pairs hasn't seemed to hurt. (We cannot, unfortunately, test for statistically significant differences). We expect that the margin of victory for omitting the Pairs would have been even greater if additional pairs had been asked with the time savings from skipping the Importance questions.

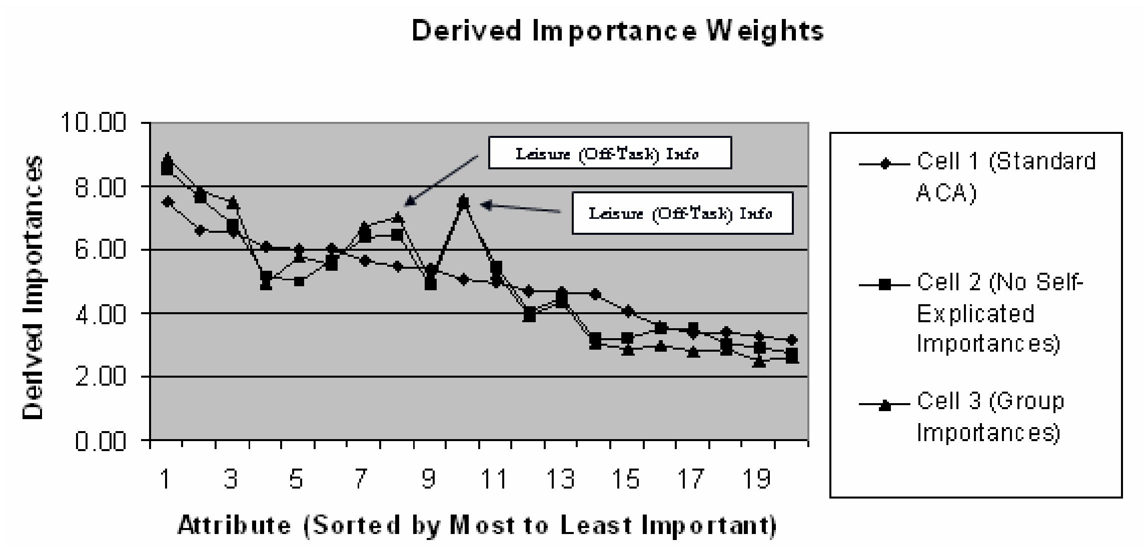
It is important to note that the use of the Importance ratings in Cell 1 seem to provide no benefit for share prediction accuracy. Throwing out the Importance ratings for Cell 1 results in directionally lower errors in share prediction (6.7%) than when using the Importance information during HB estimation (6.8%).

As has been seen in other studies (Sawtooth Software, 2003b), HB estimation provides more accurate predictions than the OLS procedure.

Findings—Importance Scores

We may wonder why the modified versions of ACA that skipped the Importance questions resulted in more accurate share predictions. Were the part worth utilities more precise due to the additional pairs asked? Was the information provided by Importance ratings different from that implied by the tradeoffs in the conjoint pairs (and holdout choice tasks)? We compared the *derived* Importance scores (by comparing the best to worst utilities for each attribute, and percentaging) for the three design Cells in Figure 8. In each case, we used HB estimation. We sorted the attributes in order of derived Cell 1 importances.

Figure 8



Of note:

- The importances for Cell 1 (standard ACA) are "flatter" than for Cells 2 and 3.
- The derived importances are nearly identical for Cells 2 and 3 and are quite different in some instances from Cell 1.

This suggests that the information provided by the self-explicated importances not only flattens the differences in derived importance among the attributes (relative to the conjoint Pairs information), but also leads to different conclusions regarding the order of importance of attributes. It turns out that the two attributes that show the highest negative deviations for Cell 1 importances versus Cells 2 and 3 involved leisure (off-task) information. It is plausible that it was not socially desirable to state (in a self-explicated importance question) that the off-task leisure information was as important as on-task information, and that the truer importance was revealed in the tradeoff tasks. Anecdotal evidence supports the derived importances suggested by the modified ACA procedures. The client commented that previous research at the firm suggested that the off-task leisure information was quite important.

Conclusions

This research suggests that for most situations ACA researchers may be better off by skipping the Importance questions and adding more Pairs questions. For this study, using the Importance questions during estimation only provided modest benefit in terms of hit rates. And, the Importance questions may actually hurt share predictions.

In summary, skipping 20 importance questions, but adding six additional pairs:

- provides nearly the same overall hit rate accuracy,
- seems to improve the share prediction accuracy,
- yields “steeper” and <u>different</u> derived importances,
- reduces the average length of the survey by over one minute.

Using aggregate information as proxy for prior importances yields directionally lower hit rates but directionally better share predictions. This experiment has not conclusively shown that this more elaborate procedure adds any value, though it is theoretically pleasing. Perhaps applying the prior group information in a non-deterministic way in the first stage of the Pairs design might tip the scales in favor of this approach rather than naïve prior importances (Cell 2).

Using within-attribute ratings information seems enough to initially provide informative (non-dominated) tradeoffs in ACA.

Future Directions:

Based on this research, it would be valuable to:

- Study other methods of incorporating prior importance scores (particularly, using prior group importances in a non-deterministic manner),
- Validate the findings with additional studies,
- Integrate the ability to drop the Importance question within ACA software (it is currently impossible to do without customization).

References

Hill, Aaron, Gary Baker and Tom Pilon (2001), "A Methodological Study to Compare ACA Web and ACA Windows Interviewing," *Sawtooth Software Conference Proceedings*: Sequim, WA.

Krieger, Abba, Paul Green, and Yoram Wind (2004), "Adventures in Conjoint Analysis: A Practitioner's Guide to Tradeoff Modeling and Applications," Unpublished Manuscript.

Green, Paul, Abba M. Krieger, and Manoj K. Agarwal (1991), "Adaptive Conjoint Analysis: Some Caveats and Suggestions", *Journal of Marketing Research*, (May), 215-22.

McLaughlan, William (1991), "Scaling Prior Utilities in Sawtooth Software's Adaptive Conjoint Analysis," *Sawtooth Software Conference Proceedings*: Ketchum, ID.

Orme, Bryan (1998), "Reducing the Number-of-Attribute-Levels Effect in ACA with Optimal Weighting," Sawtooth Software Technical Paper, available at www.sawtoothsoftware.com.

Sawtooth Software (2003a), "More Evidence CBC Is Most Popular Conjoint Method," *Sawtooth Solutions*, Summer 2003.

Sawtooth Software (2003b), "ACA/Hierarchical Bayes v2.0 Technical Paper," Technical Paper available at www.sawtoothsoftware.com.

Wittink, D.R. and P. Cattin (1989), "Commercial Use of Conjoint Analysis: An Update," *Journal of Marketing*, 53 (July) 91-6.