Reprinted from *2009 Sawtooth Software Conference Proceedings*.
Delray Beach, FL, March 2009.

CBC vs. ACBC: Comparing Results with Real Product Selection

Christopher N. Chapman
Microsoft Corporation

James L. Alford
Volt Information Sciences

Chad Johnson, Ron Weidemann
Answers Research

Michal Lahav
Sakson & Taylor Consulting

Abstract

We examined consumer preference for a computer accessory product line with an online survey. Every respondent completed both a choice-based conjoint (CBC) procedure and an adaptive choice-based conjoint (ACBC) procedure using the same attributes. Our goal was to predict market share and to answer both methodological and managerial questions. In evaluating within-subjects results, CBC and ACBC gave generally similar estimates, with ACBC estimating greater price sensitivity and giving a smaller standard deviation of respondent utilities. When estimated head-to-head preference between two products was compared to actual market data, ACBC agreed closely with actual market data. In a 4-product portfolio, ACBC predictions of preference share were in somewhat closer agreement to observed market data than those using CBC data; the ACBC error proportion was 15-25% lower than the error rate with CBC data.

Background and Managerial Question

Our product team was preparing to launch a PC accessory consumer electronics (CE) product (hereafter *Product A*) when a competitor released an identically priced product (*Product B*) that had a higher specification in one feature area that is highly salient to consumers. We wished to determine the competitive threat from Product B and decide whether to immediately launch a development effort to update Product A so that it would match the specification of Product B.

Working with the product management team, we refined the competitive question to one that was amenable to research. Balancing the development cost against market size, we determined a threshold for action of 25% head-to-head preference share. That is, if Product A achieved at least 25% consumer preference vs. Product B, we would keep Product A unchanged. However, if its preference share was less than 25% vs. Product B, we would undertake to update the product.

Because of the managerial importance of this question, we wished to assess the preference with multiple methods in order to have maximal confidence in the result. To assess preference, we chose to use four methods: traditional Choice-Based Conjoint analysis (*CBC*; Sawtooth Software, 2008); the newly developed Adaptive Choice-Based Conjoint analysis (*ACBC*; Johnson & Orme, 2007); a single monadic item presented as a CBC holdout task; and a final offer in which respondents were allowed to receive either Product A or B at no cost.

Method

We designed a single online survey using Sawtooth Software's SSI Web, comprising both CBC and ACBC exercises. The choice exercises had eight attributes: brand, price, and six product features, ranging from 2-5 levels per attribute. All attributes and levels were identical on the CBC and ACBC portions, except that Price showed minor variance between the two due to different methods for constructing conditional pricing in Sawtooth Software CBC vs. ACBC questionnaires. The difference was typically no more than $5-10 for comparable products, well within the range of conditional price randomization.

The survey was designed for all respondents to answer both CBC and ACBC exercises in randomly counterbalanced order. The CBC section included a fixed task that directly assessed the preference between Product A, Product B, and a third product designed to have minimal appeal. The CBC exercise comprised 12 random tasks plus 2 fixed tasks. The ACBC exercise comprised a BYO configurator task, 9 screening tasks of 3 concepts each, a choice tournament with 3 concepts in each group, and a final four-item calibration section (unused in the analyses presented here).

At the end of the survey, 25% of respondents were randomly selected and offered a free product, where they could choose to receive either Product A or Product B at no cost.

We also wished to examine the extent to which traditional randomized task/attribute format choice tasks could predict real market behavior (where products may be described differently than as a list of features). Thus, the final "real" product selection task presented a richer, more complete description of the available products than the CBC and ACBC exercises, comprising a superset of the choice exercises' feature descriptions. By doing this, we could contrast the rich selection task with both CBC and ACBC estimates and with a traditional holdout task presented in CBC format, potentially yielding a different estimate of overall preference that would be informative for the managerial question with regards to the sensitivity of our findings.

Respondent utility scores were computed for both CBC and ACBC sections using Sawtooth Software Hierarchical Bayes (HB) estimation with 40000 total iterations. CBC and ACBC results were compared on the basis of overall utility estimation pattern, sample-level correlation between equivalent attribute/level utilities, within-subject correlation across utilities, holdout task prediction, and prediction of actual product selection. We used respondents' individually estimated mean HB scores (i.e., *beta* scores), converted to zero-centered difference scores to reduce between-respondent scale effects (Sawtooth Software, 1999).

After six months, actual market data was available for comparison of real channel sales performance as compared to the predicted preference in the survey. The product team agreed at the inception of the product that we were interested in consumer *preference* with all other factors being equal – as opposed to *actual* market share in which other factors are not equal. Still, it was of interest to determine the degree to which the methods reflected actual market performance.

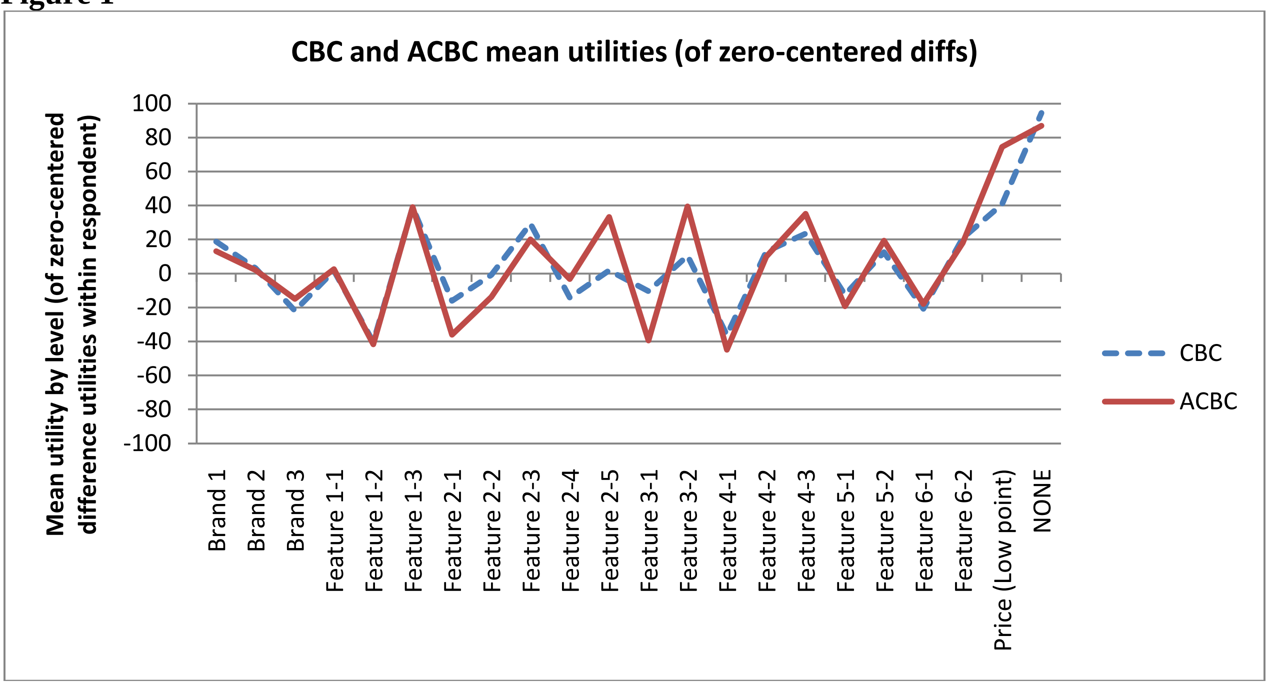
Results

N=400 respondents (PC-using adults in the US, enriched for broadband users) completed the survey. Median time to completion for CBC (when it was taken first, N=201) was 244 seconds, while ACBC (taken first, N=199) had median 442 seconds.

In terms of subjective respondent experience, on the scales proposed by Johnson & Orme (2007), respondents showed no significant differences between CBC and ACBC tasks on self-report ratings of product realism, survey interest, realistic answers, or attention (t-test, all p>.05). There was a modest effect for "This survey is at times monotonous and boring," where respondents who took CBC first reported being more bored after the first choice section than were those who started with ACBC (means 3.04 and 2.75 on 5-point scale; t=2.33, df=397, p<.05). In other words, ACBC was perceived as somewhat less boring than CBC despite that it took 80% longer on average (including the calibration items, which were not used in HB estimation).

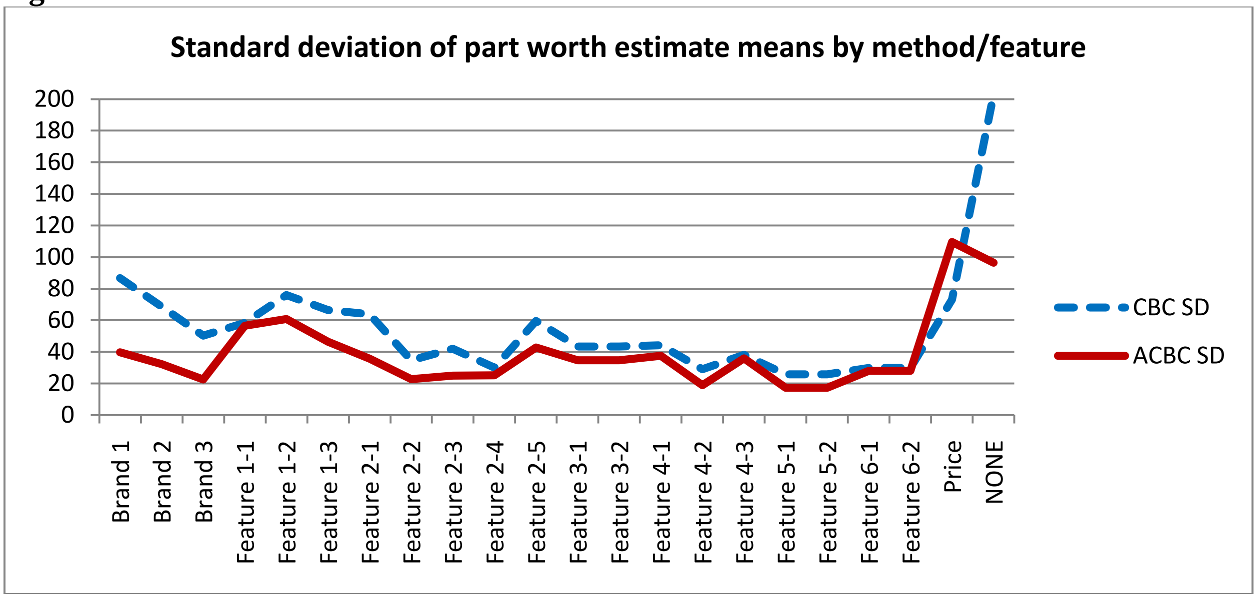
ACBC and CBC yielded moderately different utility estimates. Figure 1 shows the mean utility by feature and level, for zero-centered difference (normalized) utility scores. Apart from Brand, ACBC yielded utilities with an equivalent or larger range than CBC between best and worst mean levels within an attribute. This is consistent with the ACBC method's hope to obtain better estimates of feature value beyond a few important attributes (Johnson & Orme, 2007). In particular, ACBC yielded a higher utility weight for Price (at the low point), i.e., ACBC estimated a higher average respondent price sensitivity than did CBC.

Figure 1



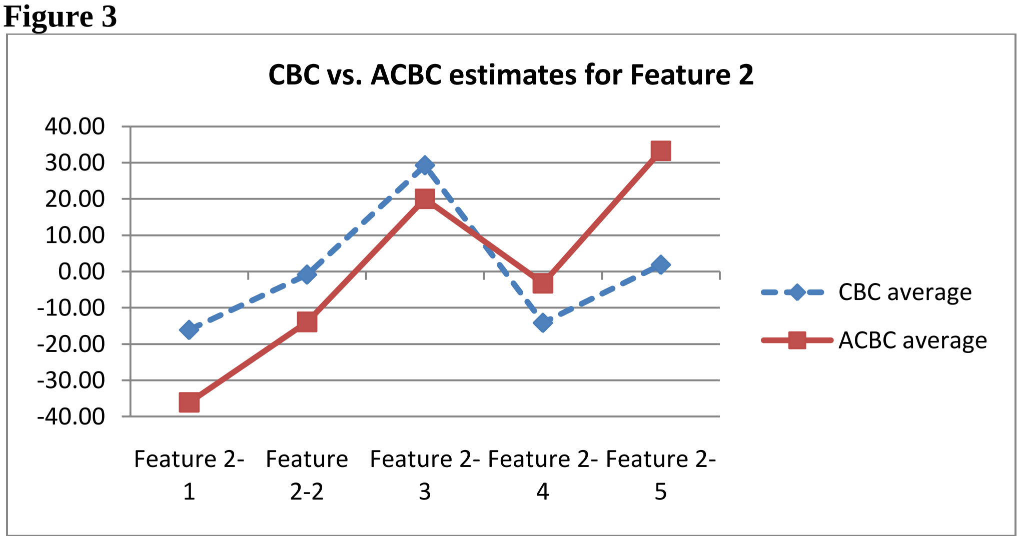
Although the mean utility levels were somewhat more extreme, ACBC generally yielded lower standard deviations within feature level than CBC. The ratio of ACBC:CBC standard deviations ranged from 0.46 to 1.24 with an average ratio of 0.79 ACBC:CBC standard deviation on raw scores; and 0.45 to 1.50 with mean 0.69 for zero-centered differences (normalized) utilities. This suggests that to achieve comparable standard errors of group-level utility means with a similar product and feature set, ACBC would need about 38% fewer respondents than CBC.

Figure 2



One product attribute, Feature 2, had five levels, of which two sets of two levels were naturally ordered (e.g., similarly to memory size or price, where one level should always be preferred to another). Within Feature 2, we expected to see Level 4 > Level 2, and Level 5 > Level 3. As shown in Figure 3, without constraints on HB estimation, we observed reversal on both sets with CBC/HB estimation, but ordering as expected with ACBC/HB estimation.

Figure 3



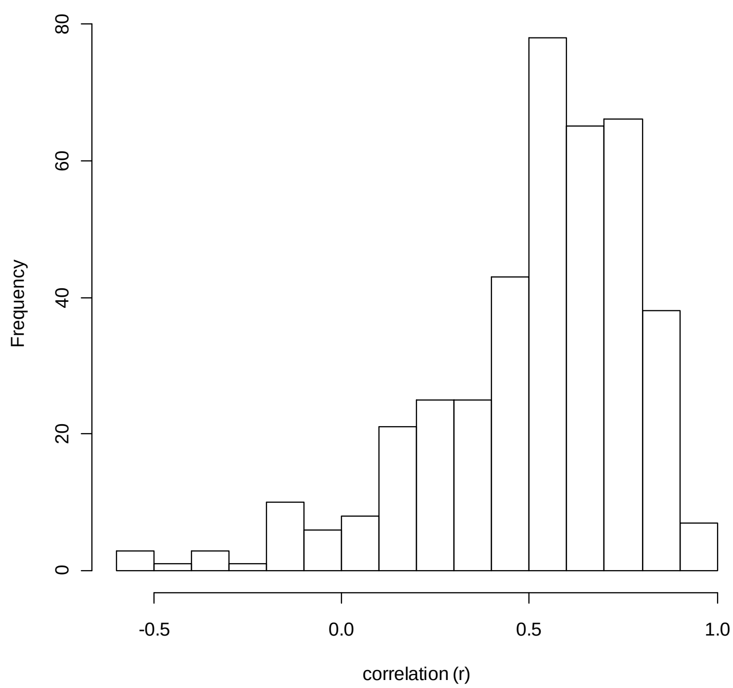
Across respondents, correlations between zero-centered difference utilities estimated by CBC and ACBC on 18 identical attribute levels (e.g., a specific brand; we omitted one level from two-level attributes, as they are identical except for direction) ranged from *r*=0.12 (for one level of a

5-valued attribute; p<.05, df=398) to *r*=0.62 (for Price, estimated at one point; p<.01, df=398), with a median correlation of *r*=0.42. Correlation after transforming to multinormality with the Box-Cox procedure (Box & Cox, 1964) yielded nearly identical correlations. These correlations demonstrate that CBC and ACBC utilities for a given feature level are positively associated but are not equivalent.

More telling are within-subject correlations between utility scores estimated from CBC and ACBC tasks. We took K-1 levels per attribute with 2-4 levels, and K-2 levels for the attribute with 5 levels (except for Price, where we used only one point because of its linear estimation) and then correlated the utility weights for CBC and ACBC within each respondent. The median within-subject correlation was *r*=0.57, with 95% between *r*=(-0.15,0.88), as shown in Figure 4.

In short, the differences in utility correlations both within- and across-subjects, the mean-level utility estimates, and differences in variance all establish that CBC and ACBC yielded somewhat different utility estimates for respondents.

Figure 4
Distribution of within-subject correlation (r) between CBC & ACBC estimates



This pattern demonstrates that the utility patterns between CBC and ACBC were modestly different, but did this mean that CBC and ACBC gave different preference predictions? Which was more correct?

We examined this question with three procedures: (1) predicting a holdout CBC task; (2) predicting the selection between two free products offered to respondents; and (3) predicting head-to-head market share of the same two products in the actual marketplace.

Predicting CBC holdout. Using strict first choice preference to predict a CBC holdout task with three options, CBC utilities achieved 70.5% accuracy within-subject (N=400), with a base rate likelihood of 49.4%, yielding an agreement coefficient (Cohen's *kappa*; Cohen, 1960) k=0.42 or moderate agreement. ACBC utilities achieved 65.5% agreement predicting the CBC holdout (N=400), with base rate 54.9%, giving k=0.24 or fair agreement. Although both methods performed better than chance (which would have k≈0), neither was highly accurate at within-subject prediction (which would occur with k>0.7). Not surprisingly, CBC did slightly better as the holdout task was in CBC-style format and was part of the CBC trial block.

Predicting final product selection. N=100 respondents were offered a dichotomous free choice between Products A and B, presented at the end of survey with a rich product description (similar to a web-based shopping format). 53/100 respondents chose to receive Product A, while 47/100 chose Product B.

Preference differentiation was stronger for CBC utilities than for ACBC utilities. Of N=162 respondents estimated by CBC to prefer Product 1 in head-to-head preference, the median preference estimate (i.e., logit rule summed utility value) was 0.927, with sd=0.158; of N=130 estimated by ACBC to prefer Product 1, the median preference was 0.802, sd=0.157.

Using strict first-choice preference, CBC utilities accurately predicted 56/100 actual choices of the free product, while ACBC correctly predicted 53/100 actual choices. With a standardized base rate (combined marginal probability) of 49/100, this yields *kappa* agreement coefficients of k=0.14 and k=0.08, respectively, or "slight" agreement between within-subject prediction and actual behavior. Thus, in this study, neither CBC nor ACBC utilities were very good predictors of *within-subject* choice on the free product offer, and their performance was lower for this task than for the assessed holdout task described above. The reasons for these low agreement rates are unknown, and may include the effects of differing task presentation, respondent inconsistency, "found money" effects from a free product offer (although the prices were equivalent), or a novelty effect (e.g., if a respondent happened to own one of the products already).

More interesting is the group-level difference in preference share between CBC, ACBC, and real choice. CBC utilities estimate a head-to-head preference for the product of interest (using strict first-choice preference) as 43.5%±5.0%, while ACBC utilities estimate 33.0%±4.8% preference. Actual selection by N=100 respondents was 53/100 choosing the target product on the free selection offer, with a binomial confidence interval of ±10% (i.e., confidence range of 43%-63% preference). Thus, the CBC confidence range (but not center) overlapped the confidence range for observed product selection on the free product offer, while the ACBC range did not.

It is important to emphasize that the free selection task differed from CBC/ACBC tasks in two ways: (1) it was an actual, not a hypothetical choice; (2) it came at the end of the survey and presented more features, with more descriptive text, in richer context, emulating an Internet-style product display. We have found previously (Chapman et al, unpublished) that CBC performance at predicting actual behavior was higher when the CBC survey format closely matched the actual product selection format (cf. also Martin & Rayner, 2008).

Predicting market response. As noted above, in a head-to-head comparison, CBC predicted 43.5%±5.0% preference share for Product A, while ACBC predicted 33.0%±4.8% preference share. Actual market data for Product A showed a head-to-head *market share* of 34.6% (this was calculated as *unit sales of A* / (*unit sales of A + unit sales of B*), with an unknown confidence interval). The ACBC prediction of preference share was not statistically significantly different from the observed actual market share – an impressive result for the product team.

Tuning of response. Although head-to-head preference prediction was the key research question and ACBC answered the question quite well, a more interesting question involves prediction among more than 2 products. Utility estimates from conjoint analysis surveys often need to be "tuned" to adjust the utilities by a constant factor that allows better prediction in comparison to actual market results (Orme & Heft, 1999).

We evaluated the tuning coefficients vs. actual market data using the Sawtooth Software SMRT market simulator with randomized first choice selection, including in the simulation all four products in the studied category that have greater than 5% unit share and are in the price range we studied. We selected the best-selling product as the baseline and adjusted the utility exponent until that product's share most closely matched actual sales data. The best fit for CBC data was obtained with a utility multiplier (i.e., exponent) of 0.28, while ACBC best fit had an exponent of 0.14.

As shown in Table 1, mean absolute error (MAE) and root mean squared error (RMSE) were reduced by 20-50% by tuning the utility exponent. After tuning, ACBC had errors that were 15-25% smaller than those of CBC.

Table 1
Actual and Predicted Shares, Before and After Utility Exponent Tuning

Product	**Actual market share**	**CBC Exp 1**	**TUNED CBC Exp 0.28**	**ACBC Exp 1**	**TUNED ACBC Exp 0.14**
Product A	0.141	0.285	0.282	0.164	0.252
Product B	0.266	0.122	0.172	0.169	0.181
Other 1 *[tuning baseline]*	**0.401**	**0.511**	**0.403**	**0.636**	**0.403**
Other 2	0.192	0.082	0.143	0.031	0.164
MAE (3 products)	--	0.127	0.095	0.129	**0.082**
RMSE (3 products)	--	0.128	0.102	0.151	**0.075**

Conclusion

For our CE product, ACBC and CBC yielded generally comparable utility estimates. Combining the two methods gave us high confidence in the ability to answer the managerial question of interest with reduced dependence on a single method. The ACBC procedure gave slightly smaller standard deviation of utility *beta* estimates across respondents, indicating that it may produce stable results with smaller sample sizes. ACBC also estimated higher price sensitivity of respondents, and yielded order effects without constraints that were more closely aligned to expectation than CBC. Neither method was highly effective at prediction of within-subject preference, but this was of marginal interest for our application, as we wished to predict group-level preference.

In head-to-head product preference estimation, ACBC was a substantially closer match to observed market data, with no statistically significant difference between predicted and observed head-to-head preference between the two products of interest. When modeling a lineup of three products, with exponent tuning to match a holdout product share, ACBC had an average error of 7.5-8.2% vs. market data, compared to 9.5-10.2% for CBC.

In short, the performance of ACBC for our CE product was similar to CBC and somewhat better in alignment with market data. We believe future research would be useful to determine whether this pattern of results (better prediction; higher price sensitivity; lower standard deviation) continues with other product categories. We hope other researchers will systematically include the kinds of external market validity and between-methods measures that we investigated.

References

Box, G.E.P., and Cox, D.R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society*, Series B 26: 211–246. http://www.jstor.org/stable/2984418.

Chapman, C.N., Alford, J., Lahav, M., Johnson, C., and Weidemann, R. (unpublished). Conjoint Analysis Prediction and Actual Behavior: Three Studies.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, vol.20, no.1, pp. 37–46.

Johnson, R.M., and Orme, B. K. (2007). A New Approach to Adaptive CBC. Sawtooth Software, Sequim, WA.

Martin, B, and Rayner, B. (2008). An Empirical Test of Pricing Techniques. Paper presented at Advanced Research Techniques Forum (A/R/T Forum) 2008, Asheville, NC.

Orme, B.K., and Heft, M.A. (1999), Predicting Actual Sales with CBC: How Capturing Heterogeneity Improves Results. Available online at http://www.sawtoothsoftware.com/download/techpap/predict.pdf; last accessed April 13, 2009.

Sawtooth Software (1999). Scaling Conjoint Part Worths: Points vs. Zero-Centered Diffs. Available online at http://www.sawtoothsoftware.com/education/ss/ss10.shtml; last accessed April 13, 2009.

Sawtooth Software (2008). CBC 6.0 Technical Paper. Available online at http://www.sawtoothsoftware.com/download/techpap/cbctech.pdf; last accessed April 13, 2009.

Acknowledgements

The authors thank Bryan Orme of Sawtooth Software for extensive feedback and discussion of both the research plan and data analysis, as well as discussion at the Sawtooth Software Conference (SSC) 2009, and thank Rich Johnson of Sawtooth Software for review and comments on initial results. We also thank Dr. Edwin Love of Western Washington University, with whom we had fruitful discussions of the research plan and data analysis; colleagues at Microsoft Hardware who posed the initial research questions and supported our sharing methods and results with the research community; and many attendees of SSC 2009 for their insightful questions and discussion.