A Researcher's Guide to Studying Large Attribute Sets in Choice-Based Conjoint

Megan Peitz
Numerious
Mike Serpetti
Dan Yardley
Gongos

Abstract

There is no arguing that choice-based methods have become dominant in our industry. Yet, there is no clear answer on what a researcher should do as the number of attributes increases (>6). Design techniques including Partial Profile and Adaptive Choice-Based Conjoint offer solutions, but past research has yet to crown a winner. This paper sets out to revisit prior research and to explore and validate both Partial Profile (PP) and Adaptive Choice-Based Conjoint (ACBC) in comparison to Full Profile (FP) Choice-Based Conjoint (CBC) with real respondents on experiments of 10, 15, and 20 attributes to determine which method is best across multiple scenarios.

Introduction

Most of the choice-based research done today is Full Profile (FP), where a level from every attribute is shown in every product profile. However, some argue that there comes a point when a FP choice task is too cumbersome and overwhelming, forcing respondents to use a simplification heuristic that could affect the model's predictability. Since the work of Green and Srinivasan (Green, P. & Srinivasan, 1978), we have been historically taught to use around six attributes (depending on level text, category, and more). Two solutions to this problem include Partial Profile (PP) and Adaptive Choice-Based Conjoint (ACBC). PP is where a level from only a subset of attributes, usually 7 or fewer, is shown in every product profile. The subset of attributes changes across every screen so that respondents evaluate all attributes, but only 7 at a time (Chrzan, K. & Elrod, T., 1995).

ACBC takes respondents through three main phases: 1) BYO (configuration) phase, 2) Consideration phase, and 3) Choice Tournament phase, adapting the design depending upon answers within these phases to account for non-compensatory decision making that can happen (Johnson, R. & Orme, B., 2007). Screen shots of the FP CBC, PP CBC, and ACBC exercises can be found in the appendix.

Study Design

We chose smartphones as the product and created sample cells with 10 attributes, 15 attributes, and 20 attributes to describe the product profiles. The set of attributes and levels tested is in Figure 1.1.

Figure 1.1: Attributes and Levels Tested

10	15	20	ATTRIBUTE	LEVEL 1	LEVEL 2	LEVEL 3	LEVEL 4	LEVEL 5	LEVEL 6
X	X	X	**Brand**	Apple	Samsung	Motorola	Google	LG	Sony
X	X	X	**Price**	$700	$800	$900	$1,000	$1,100	
X	X	X	**Screen Size**	4.6"	5.2"	5.8"	6.4"		
X	X	X	**Storage**	64 GB	128 GB	256 GB	512 GB	1TB	
X	X	X	**Quality of Camera**	8 megapixels	12 megapixels	16 megapixels	24 megapixels		
X	X	X	**Generation**	4G	5G				
X	X	X	**Battery Life**	14 hours	20 hours	26 hours	32 hours	40 hours	
X	X	X	**RAM**	2GB RAM	4GB RAM	6GB RAM	8GB RAM		
X	X	X	**HDR**	No	Yes				
X	X	X	**Waterproof**	No	Yes				
	X	X	**Wireless Charging**	No	Yes				
	X	X	**Headphone Jack**	No	Yes				
	X	X	**Screen Quality**	Standard (800x400)	High Def (1280x720)	Full HD (1920x1080)			
	X	X	**Dual Camera**	No	Yes				
	X	X	**SD Slot**	No	Yes				
		X	**Color**	White	Black	Blue	Gold	Pink	Silver
		X	**Weight**	5 oz	6 oz	7 oz	8 oz		
		X	**Facial Recognition**	Touch ID	Face ID				
		X	**Display**	LCD	OLED				
		X	**Video Recording Quality**	Up to 720p	Up to 1080p	Up to 4k			

Each cell was assigned a specific methodology, shown in Figure 1.2.

Figure 1.2: Methodologies Tested

CHOICE-BASED CONJOINT
FULL PROFILE

All attributes and all levels are shown in every product profile.

CHOICE-BASED CONJOINT
PARTIAL PROFILE

The computer randomly selects a subset of 7 attributes to show in every product profile.

Each screen shows a different subset; Brand & Price were not forced.

ADAPTIVE
CHOICE-BASED CONJOINT
FULL PROFILE

All attributes and all levels are shown in every product profile.

ADAPTIVE
CHOICE-BASED CONJOINT
PARTIAL PROFILE

Each respondent chooses a subset of attributes that are important to them, between 2 and 10. Only those attributes are shown in every product profile.

Each screen shows the same subset; Brand & Price were not forced.

Sample Cells

There were 11 sample cells in total. Three with 10 attributes, four with 15 attributes, and four with 20 attributes. See Figure 1.3 for detail on the sample cells. Each sample cell was weighted equally by gender, age, and ethnicity. Sample was provided by Dynata.

Figure 1.3: Sample Cell Overview

METHOD	PROFILE	N	ATTRIBUTES SHOWN	ATTRIBUTES IN TOTAL	# TASKS
CBC	Full	170	10	10	8
CBC	Partial	183	7	10	10
ACBC	Full	164	10	10	BYO, 10 Screening, 10 Choice
CBC	Full	188	15	15	10
CBC	Partial	201	7	15	12
ACBC	Full	200	15	15	BYO, 10 Screening, 13 Choice
ACBC	Partial	207	10	15	BYO, 8 Screening, 10 Choice
CBC	Full	229	20	20	12
CBC	Partial	279	7	20	14
ACBC	Full	244	20	20	BYO, 12 Screening, 16 Choice
ACBC	Partial	249	10	20	BYO, 8 Screening, 10 Choice

Design Strategy

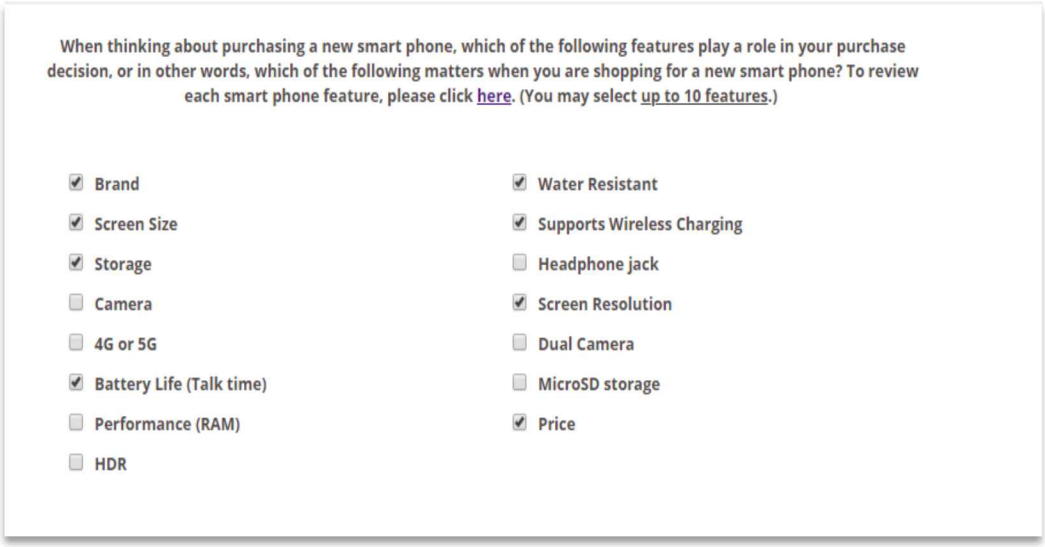
The FP CBC design was generated using Sawtooth Software's Shortcut design algorithm. The PP CBC design was generated using Sawtooth Software's Complete Enumeration design algorithm (as the attributes not shown in the product profile are essentially assuming level overlap). The ACBC designs were generated using Sawtooth Software's default settings.

Partial Profile Attribute Selection

For the PP CBC, the computer selected 7 attributes on each screen based upon the design algorithm. Brand and Price were not forced onto every screen in the PP CBC.

For the PP ACBC, respondents could choose between 2 to 10 attributes. Brand and Price were not forced into the respondent's subset for PP ACBC. Figure 1.4 shows an example of how respondents configured their subset.

Figure 1.4: PP ACBC 15 Attribute Selection Question



15 Attribute PP ACBC Selection

Respondents vary in the number of attributes they find play a role in their decision to purchase a smartphone. 71% include Price and only 60% include Brand. Figure 1.5 shows the distribution of the number of attributes per respondent. Figure 1.6 shows the proportion of respondents that chose that attribute.

Figure 1.5: % of # of Attributes in 15 Attribute PP ACBC per Respondent

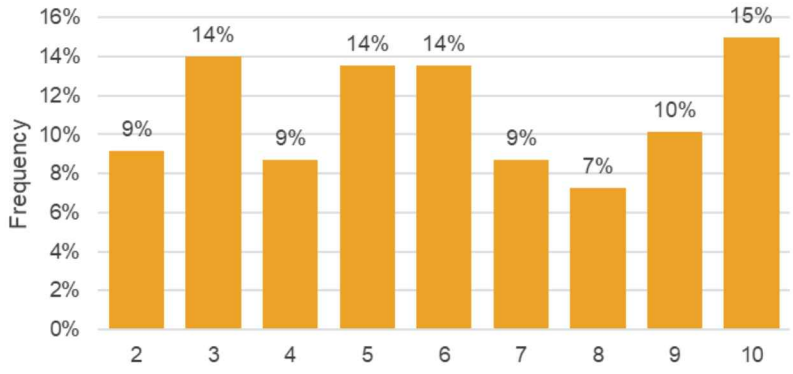
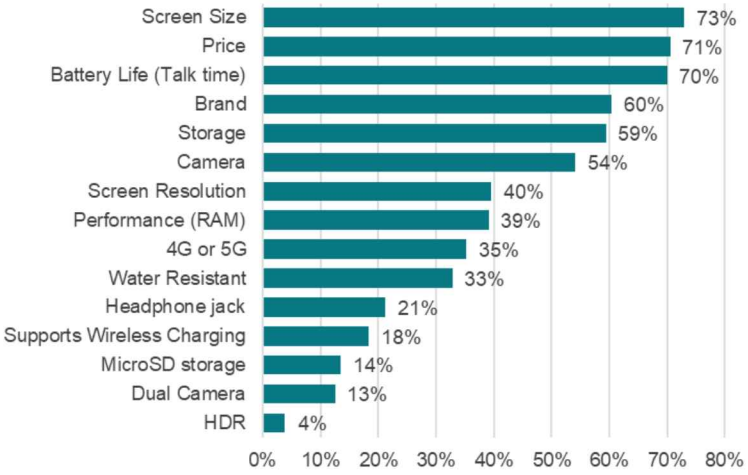


Figure 1.6: % of Respondents Including Each Attribute in Their Subset



20 Attribute PP ACBC Selection

When shown 20 attributes, respondents are more likely to choose the maximum number of attributes allowed (10) in comparison to 15 attributes. 66% include Price and 61% include Brand. Figure 1.7 shows the distribution of the number of attributes per respondent. Figure 1.8 shows the proportion of respondents that chose that attribute.

Figure 1.7: % of # of Attributes in 20 Attribute PP ACBC per Respondent

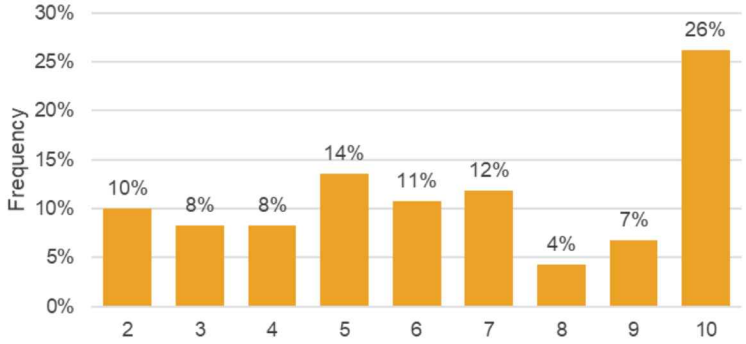
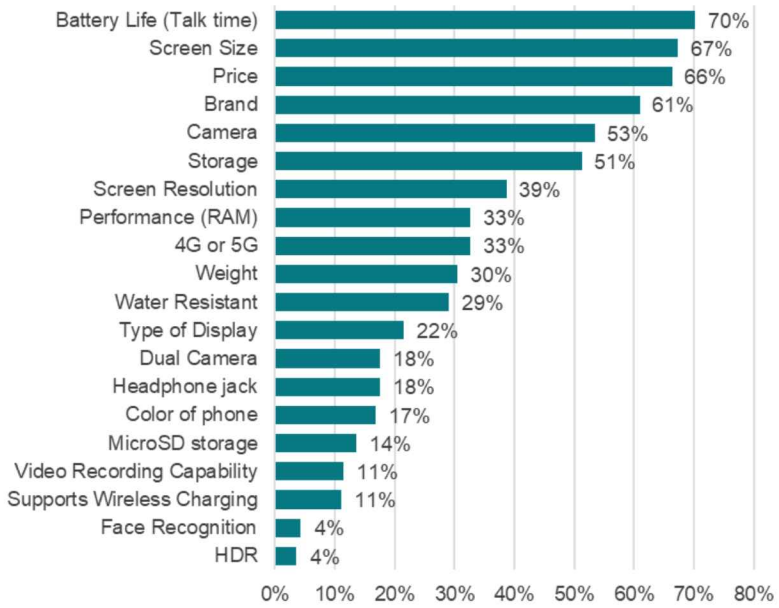


Figure 1.8: % of Respondents Including Each Attribute in Their Subset

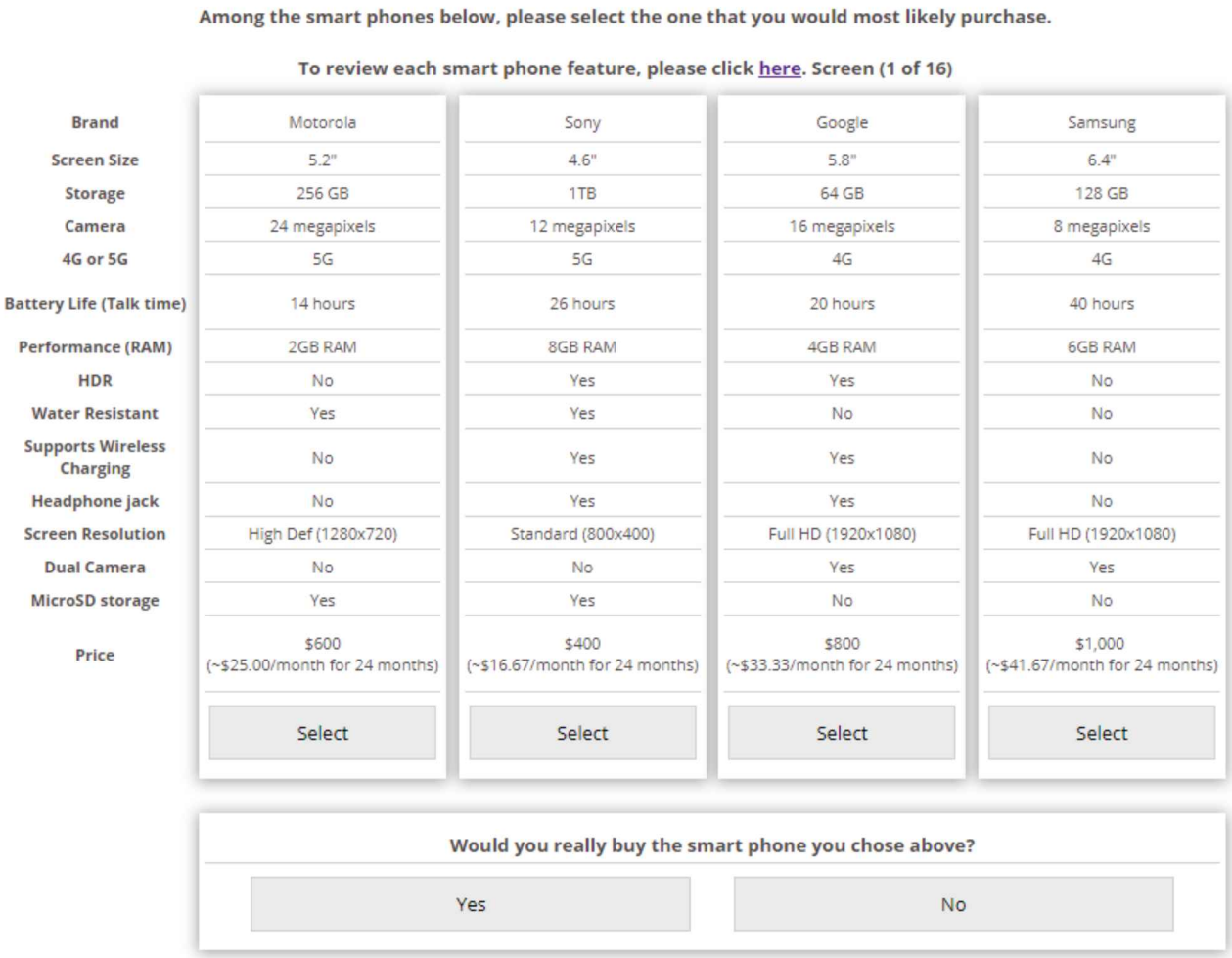


In future projects, analysts should consider forcing brand and price into every subset of the PP CBC and list of attributes in the PP ACBC. Doing this may depict a more accurate representation of the real world, given the influence these attributes have on decision making.

HOLDOUT TASKS

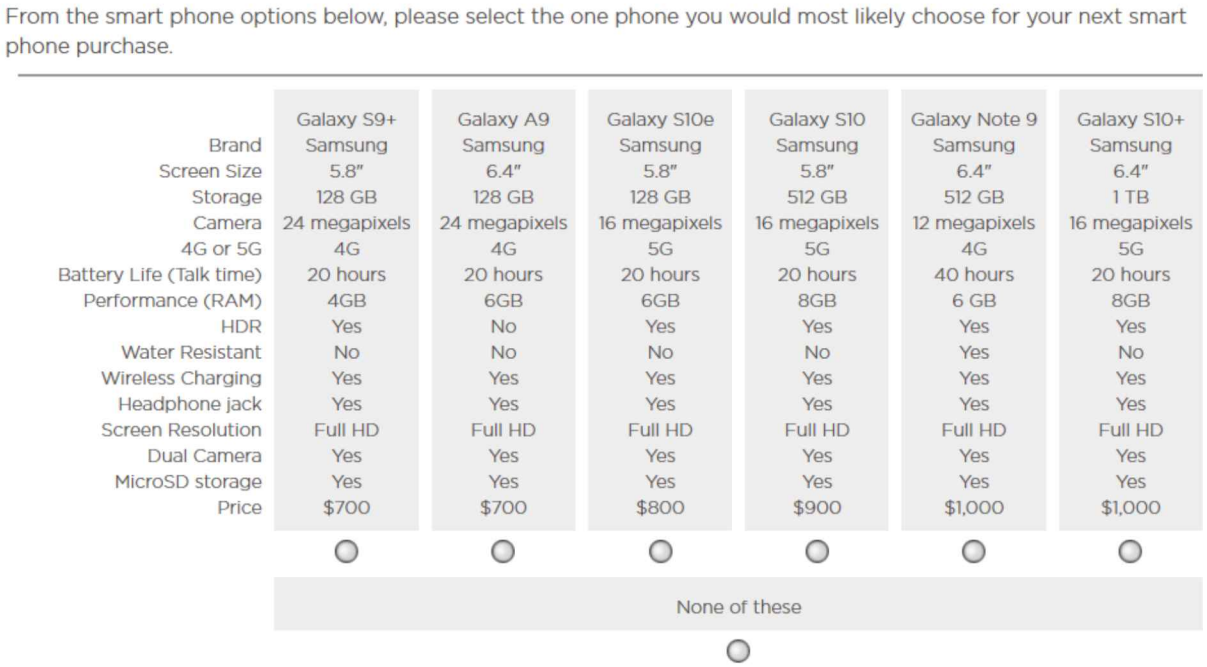
Holdout tasks can be used to compare the predictive validity of one conjoint method to another. In a holdout task, the researcher specifies exactly which combinations of attribute levels to show in a product profile and all respondents see this exact scenario. Chrzan (2015) suggests that at least 5 holdout tasks, if not more, are needed to be confident in these conclusions. This study includes 6 CBC-looking holdout tasks (Figure 2.1).

Figure 2.1: CBC Holdout Example



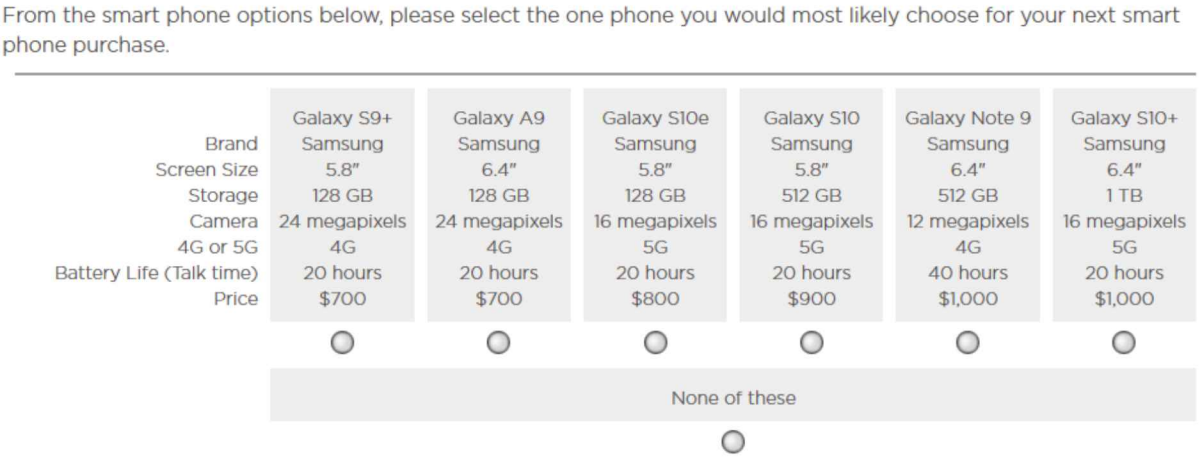
Since the best measure of success is each model's ability to predict real-world market shares, two shelf holdouts were also included to reflect the actual consumer decision-making process. The first shelf is a Full Profile and includes 20 total products throughout the duration of the exercise (Figure 2.2 shows one of these screens that tests 6 Samsung Galaxy products).

Figure 2.2: FP Shelf-Holdout Example



Research has shown that FP is better than PP at predicting hit rates for FP holdout choice tasks, therefore a PP shelf was also included so to not tip the hat in favor of FP. In the PP shelf, each product profile is defined by the same 7 attributes across all respondents and includes 20 total products throughout the duration of the exercise (Figure 2.3 shows one of these screens that tests 6 Samsung Galaxy products).

Figure 2.3: PP Shelf-Holdout Example



In addition, we must remember that because each of these tasks look like CBC tasks, there is a potential methods effect in favor of CBC vs. ACBC.

The Models

In all 11 cells, we created a hierarchical Bayesian (HB) model with prior variance of 0.5 and 5 degrees of freedom and used point estimates (the default Sawtooth Software settings). In addition, no prohibitions, constraints, or interactions were included in any model.

For the scale factor (response error) involved in the different calibration tasks, and holdout task layouts to not affect the share prediction accuracy criterion (MAE), each model's exponent was tuned to minimize the MAE across all holdout tasks. The holdout tasks were not used in estimating the utilities.

The Results

Comparisons across the methodologies are made within the cells that test 10 attributes, 15 attributes, and 20 attributes. Statistically, the most important comparison is how well the models perform. Here, we compare the Mean Absolute Error (MAE) and how well the model predicts the None category. In addition, price sensitivity curves and Willingness-to-Pay values are compared.

However, it is also just as important to have an enjoyable respondent experience. Therefore, we will also compare the methodologies from the respondent's perspective, examining median time to complete/length of interview (LOI), drop-off percentages, percentage of those who admit to cheating, and respondent evaluations (i.e., easy vs. hard, fun vs. dull).

The 10 Attribute Story

Model Validity

When simulating the shelf and CBC holdout tasks and comparing the simulated shares to the actual holdout shares, we find that ACBC has the lowest MAE (Table 3.1). These simulations include the None parameter (i.e., respondents can opt out of buying a phone).

Table 3.1 Mean Absolute Error Across 10 Attribute Shelf and CBC Holdouts Including the None

Including the None	Shelf-Looking Holdouts			CBC-Looking Holdouts						
	FP	PP	MAE	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	MAE
CBC FP	3.0%	2.3%	2.6%	4.2%	3.8%	6.2%	2.8%	3.6%	4.8%	4.2%
CBC PP	2.0%	1.7%	1.8%	3.1%	3.4%	2.6%	4.1%	2.5%	3.6%	3.2%
ACBC FP	1.2%	1.2%	1.2%	1.9%	3.3%	2.2%	1.8%	2.7%	2.9%	2.5%

However, because the None parameter is computed differently within ACBC vs. CBC (i.e., in ACBC the None is determined from the number of concepts marked as a possibility vs. not a possibility in the screening section), we sought to compare the MAEs when excluding the None option in simulations. After dropping the None, the MAEs for all three methods are comparable (Table 3.2). Again, we should note that shared methods bias would

be expected to favor FP and PP CBC rather than ACBC, because the holdout tasks involved CBC-looking tasks.

Table 3.2 Mean Absolute Error Across 10 Attribute Shelf and CBC Holdouts Excluding the None

Dropping the None	Shelf-Looking Holdouts			CBC-Looking Holdouts						
	FP	PP	MAE	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	MAE
CBC FP	3.7%	3.2%	3.5%	7.8%	6.6%	5.5%	3.6%	6.5%	3.8%	5.7%
CBC PP	2.39%	2.0%	2.2%	5.6%	6.3%	6.0%	9.5%	5.1%	3.2%	6.0%
ACBC FP	2.5%	2.5%	2.5%	5.5%	4.3%	10.4%	1.5%	7.4%	6.2%	5.9%

The None Parameter

Taking a closer look at the None, we find that those respondents in the ACBC sample cells are significantly more likely to choose the None (i.e., respondents would not buy any product on the shelf) (Table 3.3). This could signal a psychological effect of the ACBC exercise, making them less likely to buy an actual product after building their own. Further research should be done to determine if this is category specific, or methodology specific.

Table 3.3: None Proportions within the 10 Attribute Experiments

	CBC FP	CBC PP	ACBC FP
PP Shelf Holdout			
None simulated (Exp=1)	41%	28%	42%
None actual	21%	22%	36%
Difference	+20%	+6%	-6%
FP Shelf Holdout			
None simulated (Exp=1)	40%	39%	44%
None actual	21%	21%	32%
Difference	+19%	+18%	+12%

Price Sensitivity

Price sensitivity is a primary deliverable of choice research. Therefore, we simulated one product versus the None and graphed the share of preference estimated from each model when only changing price.

Figure 3.4: Price Sensitivity of 10 Attributes by Methodology

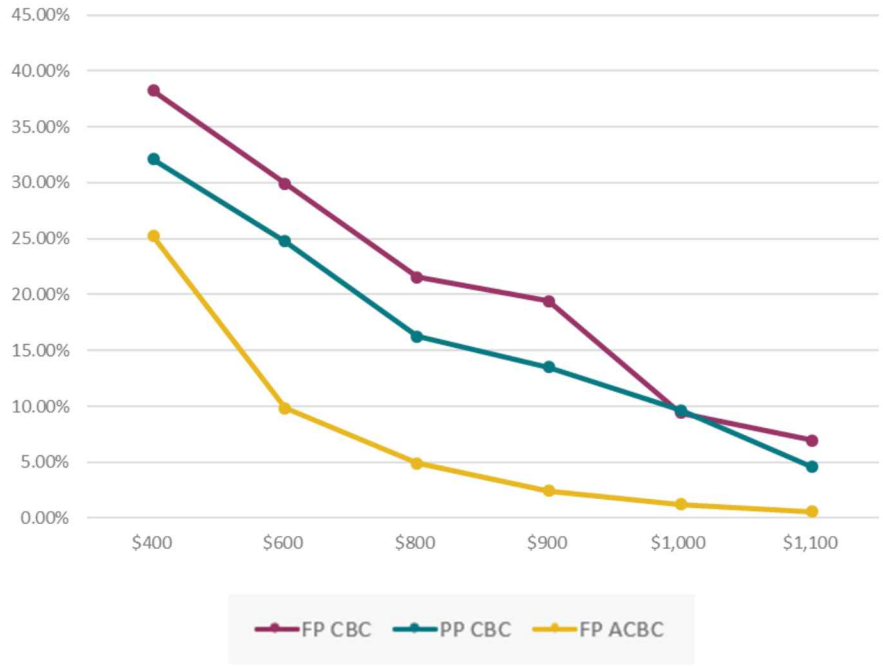


Figure 3.4 shows that FP ACBC seems to conservatively predict choice compared to FP CBC and PP CBC. This is in alignment with the high None proportion found in the ACBC data. FP CBC and PP CBC seem to have parallel curves—even though price was not forced into every product profile for PP CBC.

Willingness to Pay (WTP)

Although there are many caveats to calculating WTP data, the authors wanted to explore any differences in the methodologies tested. Our WTP data is calculated using the simulation approach where two products with the exact same specs are simulated versus the None. By default, their shares will be exactly equal. We then add a feature to the second product that the first product doesn't have, (i.e., increase product 2 from 4G to 5G; product 1 stays at 4G) and find the price for the second product at which shares for both products return to equal.

Figure 3.5: WTP for Brand, Storage, and Generation Attributes by Methodology

	CBC FP	CBC PP	ACBC FP
BRAND			
APPLE	$200	$200	$200
SAMSUNG	$55	$109	$200
GOOGLE	$0	$0	$0
LG	-$181	$11	-$67
MOTOROLA	-$262	-$310	-$130
SONY	-$322	-$238	-$90
STORAGE			
64 GB	-$172	-$222	-$28
128 GB	$0	$0	$0
256 GB	$32	$7	$33
512 GB	$41	$29	$138
1TB	$47	-$11	$117
GENERATION			
4G	$0	$0	$0
5G	$16	$49	$63

The results show that FP ACBC has a WTP that is flatter than both FP CBC and PP CBC. PP CBC seems to align well with FP CBC. This is also found when comparing correlations between utilities for the different methods (Figure 3.6).

Figure 3.6: Correlations of Aggregate Utilities by Methodology

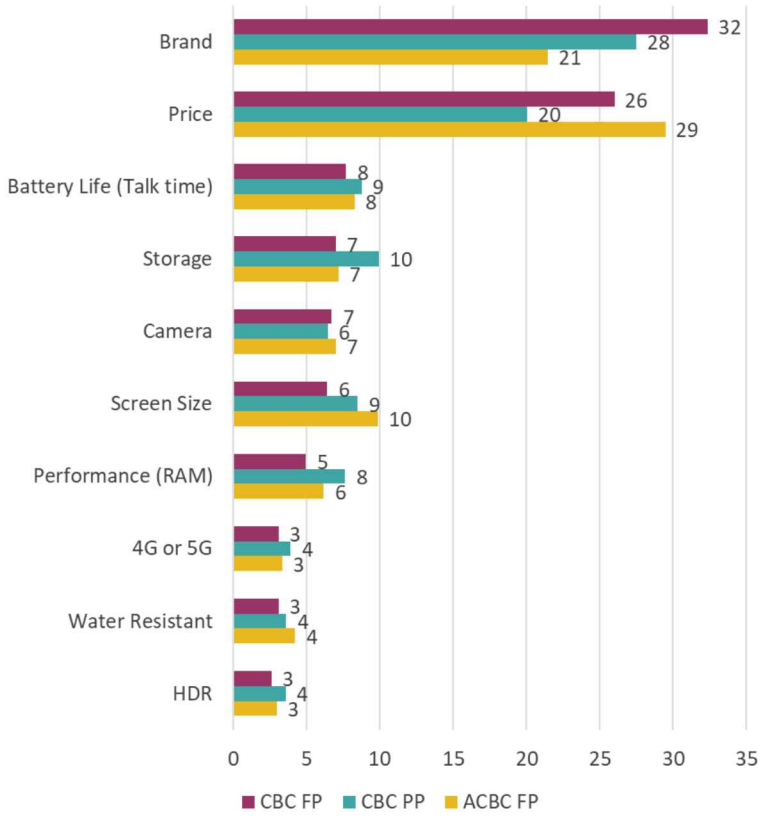
Utility Correlations	CBC FP	CBC PP	ACBC FP
CBC FP	1		
CBC PP	0.94	1	
ACBC FP	0.89	0.92	1

Importance Scores

Importance scores are a typical deliverable, albeit with many potential issues (i.e., lies are in the averages, extremely influenced by the levels tested). While the authors do not believe this is the best way to look at the data, it is interesting to see the differences since all attributes/levels tested across the methods are the same, the importance scores shown are the standard Sawtooth Software calculation of importance scores (i.e., range of HB utilities per attribute per respondent). The authors realize that there are different ways to calculate attribute importance, but wanted to investigate the results based on how a typical conjoint user would use attribute importance.

As expected, FP CBC has the steepest importance scores. Interestingly, FP CBC and PP CBC find Brand as the most important attribute followed by Price, where FP ACBC finds Price more important than Brand (Figure 3.7).

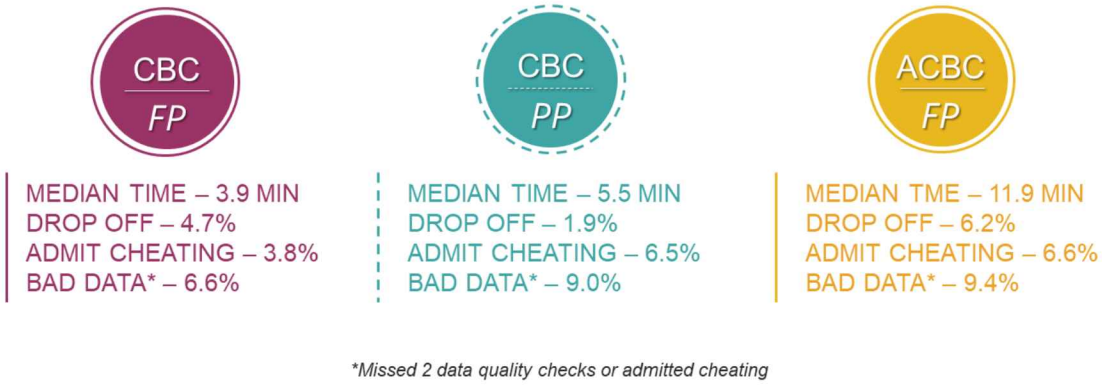
Figure 3.7: Importance Scores by Methodology



Respondent Preference

FP CBC is the quickest exercise to complete, while FP ACBC is the longest (Figure 3.8). FP ACBC also has the highest drop-off percentage, highest likelihood of cheating, and the highest proportion of bad data. Throughout the paper, bad data is defined as anyone who admitted to cheating throughout the exercise or who had two flags in the data (speeding, poor RLH, straight lining, etc.

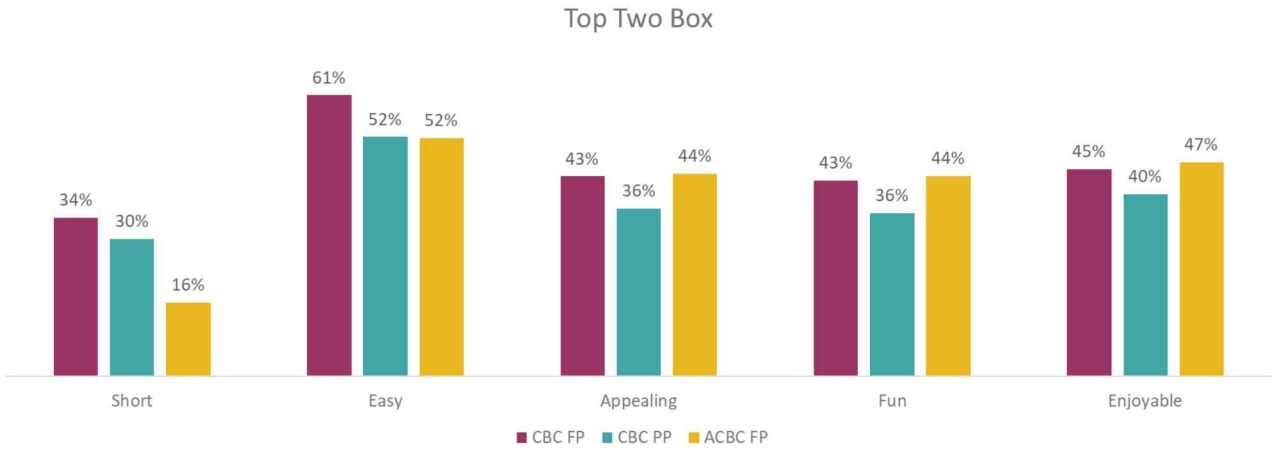
Figure 3.8: Respondent Statistics for 10 Attribute Experiments



*Missed 2 data quality checks or admitted cheating

When rating the different methodologies, FP CBC is the shortest and easiest, while FP ACBC relatively more Enjoyable, Fun, and Appealing (Figure 3.9). Both FP CBC and FP ACBC seem to edge out PP CBC, but the differences are not significant.

Figure 3.9: Respondent Top Two Agreement on 10 Attribute Survey Experience



10 Attribute Conclusion

Overall, FP CBC seems to perform slightly better when testing 10 attributes, from both a respondent and model perspective. (Again, with the caveat that the holdouts would be expected to favor the CBC-looking approaches and put ACBC at a disadvantage.)

15 Attribute Story

Model Validity

Table 4.1 shows that when including the None option in simulations, ACBC has the lowest MAE, particularly PP ACBC. After dropping the None, the MAE for the Adaptive methods are more in line with FP CBC (Table 4.2). PP CBC has the highest MAE, perhaps due to the number of attributes shown out of the total attributes modeled (7/15 < 50%).

Table 4.1: Mean Absolute Error Across 15 Attribute Shelf and CBC Holdouts Including the None

Including the None

	Shelf-Looking Holdouts			CBC-Looking Holdouts							
	FP	PP	MAE	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	MAE
CBC FP	2.4%	3.1%	2.7%	3.1%	1.7%	5.3%	1.9%	3.0%	6.3%	3.6%
CBC PP	2.0%	2.6%	2.3%	6.6%	10.1%	8.2%	8.1%	5.8%	4.5%	7.2%
ACBC FP	1.5%	2.8%	2.2%	2.6%	4.3%	4.1%	3.2%	3.2%	3.0%	3.4%
ACBC PP	1.3%	1.7%	1.5%	3.2%	1.8%	1.5%	3.3%	2.0%	5.5%	2.9%

Table 4.2: Mean Absolute Error Across 15 Attribute Shelf and CBC Holdouts Excluding the None

Dropping the None

	Shelf-Looking Holdouts			CBC-Looking Holdouts							
	FP	PP	MAE	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	MAE
CBC FP	3.0%	3.9%	3.4%	3.4%	4.5%	8.1%	2.6%	4.2%	8.3%	5.2%
CBC PP	2.0%	1.9%	2.0%	11.3%	12.3%	15.9%	11.2%	9.9%	6.5%	11.2%
ACBC FP	2.1%	2.4%	2.3%	1.7%	12.0%	8.7%	7.2%	6.6%	8.1%	7.4%
ACBC PP	2.1%	2.5%	2.3%	4.1%	5.0%	11.4%	8.4%	1.6%	10.7%	6.9%

The None Parameter

Again, we see that those in the ACBC sample cells are significantly more likely to choose the None in the holdout choices (i.e., respondents would not buy any product on the shelf) (Table 4.3). In addition, both PP methods, PP ACBC and PP CBC, seem to align more closely to shelf behavior as seen in the survey tool.

Table 4.3: None Proportions within the 15 Attribute Experiments

	CBC FP	CBC PP	ACBC FP	ACBC PP
PP Shelf Holdout				
None simulated (Exp=1)	44%	22%	62%	38%
None actual	21%	13%	25%	31%
Difference	+23%	+9%	+37%	+6%
FP Shelf Holdout				
None simulated (Exp=1)	37%	10%	48%	28%
None actual	18%	14%	28%	29%
Difference	+19%	-5%	+20%	-2%