Sawtooth Software

RESEARCH PAPER SERIES

Task Order Effects in Menu-Based Choice

Bryan K. Orme
Sawtooth Software, Inc.

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

Task Order Effects in Menu-Based Choice[1]

Bryan Orme, Sawtooth Software

April 2010

Background

Menu-Based Choice (MBC) studies ask respondents to select from zero to multiple options from a menu, such as for restaurant choices or buying options on a new vehicle. A simple menu-based choice task is shown in Figure 1.

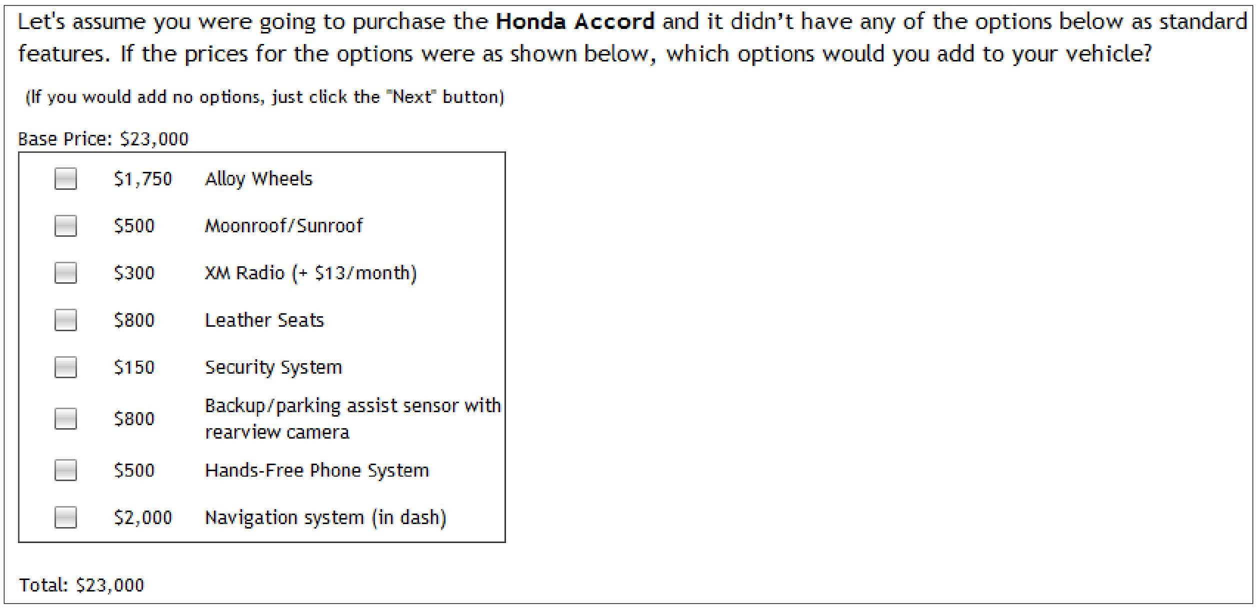

Figure 1

As respondents click ("buy") features on the menu, the total price shown at the bottom increases. When respondents finish configuring what they would purchase, they click the Next button and move to the next question. Examples of MBC include articles by Liechty *et al.* (2001), Baaken and Bond (2004), and Cohen and Liechty (2007), as well as tutorials at Sawtooth Software and SKIM training events (Orme, 2006, 2007a, 2007b, 2009).

One can estimate models and build full simulators that project purchases given any possible combination of prices, and these can be somewhat complex to build. The questions that this paper seeks to answer can be approached with simple counting analysis.

[1] Special thanks to Rich Johnson (Sawtooth Software), Keith Chrzan (Maritz), Chris Moore (GfK), Kees van der Wagt (SKIM), and Jack Horne (Market Strategies International) for providing helpful comments regarding this paper. Chris and Jack additionally examined their own recent menu-based datasets, and reported their findings regarding task order effects.

Among other things, researchers can use MBC exercises to gauge price sensitivity for each feature on the menu. The prices for each menu item can be changed from respondent to respondent, and from task to task. For example, the price for Alloy Wheels could take on 4 possible values ($1500, $1750, $2000, $2500) much like levels in a conjoint analysis experiment. Unique price ranges (alternative-specific prices) can be specified for each of the items in the study. If the prices for features in the menu are manipulated in an uncorrelated fashion (such as using a randomized, near-orthogonal design), the researcher can estimate the price sensitivity for each feature independent of the others. Cross-elasticities (substitution effects) can also be estimated. Bundling vs. *a la carte* strategies may also be studied.

Effect of Task Position

Menu-Based Choice experiments have been described in the literature and at our conferences for about 10 years now. A question that I have yet to see answered in these articles is whether a respondent's answers to the first menu task are very different from the answers to later tasks. Each respondent could be asked to complete just one menu. But, researchers typically ask respondents to complete multiple menus (where each new menu reflects changes in prices, or other aspects). Do estimated price sensitivities or preferences for certain items on the menu shift from early to later tasks due to learning effects?

The most relevant article to this question that I can think of is available in the Technical Papers Library on our website, entitled, "How Many Questions Should You Ask in Choice-Based Conjoint Studies?" For that paper, Rich Johnson and I examined about 20 commercial CBC (Choice-Based Conjoint) datasets, and we analyzed how respondents' preferences changed from early tasks to later tasks. We presented the results at the 1996 ART/Forum, and were fortunate to win "best presentation." Although CBC tasks aren't exactly like MBC tasks, there would seem to be many similarities. Respondents deal with prices changing from task to task and select options on the screen. Respondents inevitably apply the experiences learned from earlier choice tasks in evaluating later choice tasks.

One of the handy aspects of randomized experiments is that (across respondents) we can examine the preferences (or Counts scores) for levels, aggregating choices for one task at a time. We can compare preferences from the first task to those from the second task, etc. After examining results for about 20 commercial CBC data sets in our 1996 paper, Rich and I drew some conclusions regarding task effects and CBC:

- Respondents answer later tasks more rapidly than earlier tasks, with the biggest drop-off in time to complete from the first to the second task.
- Respondents are more internally consistent in later tasks than earlier tasks.
- Respondents are more likely to choose "None" as the interview progresses.
- Respondents become more price-sensitive in later tasks.

After looking at the differences in reliability and preferences from very early tasks to later ones, we suggested, "…later tasks are better predictors of results from the total interview, and estimates from the first two tasks are least like the total. This may support the practice of including warm-up tasks that are excluded from the analysis…"

The notion that later tasks are different from earlier tasks for CBC has fueled some debate in the industry. Are the earlier tasks (perhaps just the first one) the most trusted? Or, are later ones more valid? We think the answer probably depends on the product category and buying situation. Many product categories involve a search process (such as via the Internet) wherein the buyer becomes aware of different prices offered by different brands, channels, and suppliers. Essentially, multiple choice scenarios are evaluated, and through iterations of search the buyer becomes more informed and knowledgeable about different prices, terms, and availability. Also, many product categories are purchased repeatedly on a periodic basis, where prices, brand availability, and product specifications may change over time. It would seem that repeated choice tasks might faithfully reflect these situations and be compatible with the idea of realistic learning behavior for many real-world purchases. Also, since MBC experiments typically require larger sample sizes than other conjoint methods to obtain adequate precision, asking multiple tasks can save a great deal of money in fielding costs.

Results of Two Menu-Based Choice Experiments

In this article, we'll refer to two MBC choice experiments we conducted as Internet surveys using our SSI Web package, with Western Wats (Opinion Outpost) panelists. The first one involved 681 respondents, and showed a fast-food menu (Orme 2006). Each respondent completed eight menu-based tasks, involving selections of value meals vs. *a la carte* options (Figure 2). Burgers, salads, fries, drinks, healthy sides, and desserts were on the menu. The second study involved 806 respondents making choices of options for new cars, collected in March 2010 (Figure 1).

Menu Scenario #1: Please imagine you pulled into a fast-food restaurant to order dinner for just yourself. If this were the menu, what (if anything) would you purchase?

☐ **Deluxe Hamburger Value Meal**	☐ **Chicken Sandwich Value Meal**	☐ **Fish Sandwich Value Meal**
-Deluxe Hamburger -Medium fries -Medium drink $3.99	-Chicken Sandwich -Medium fries -Medium drink $5.59	-Fish Sandwich -Medium fries -Medium drink $3.99

(Only order sandwiches, fries or drinks from this area if you did not pick a value meal above)

Sandwiches:
☐ Deluxe Hamburger $1.99
☐ Chicken Sandwich $3.59
☐ Fish Sandwich $1.99

Fries:
☐ Small $0.79
☐ Medium $1.49
☐ Large $1.69

Drinks:
☐ Small $0.99
☐ Medium $1.69
☐ Large $2.19

Salads:
☐ Cobb dinner salad $4.79
☐ Grilled chicken salad $4.39

Healthy Sides:
☐ Carrots/Celery with Ranch dressing $1.19
☐ Apple slices/Grapes with dipping sauce $0.99

Desserts:
☐ Apple/Cherry/Berry pie $0.99
☐ Cookies $1.19

☐ I wouldn't buy anything from this menu.
I'd drive to a different restaurant, or do something else for dinner.

Figure 2

Internet panelists tend to be quite adept at completing web-based surveys, as they voluntarily do this often. Certainly, some of the panelists have completed complex grids, conjoint tasks, or even menu-based tasks, in previous questionnaire experiences. Thus, we'd expect smaller learning effects from experienced panelists than from respondents who rarely complete market research surveys. Even so, we noticed some strong learning effects in both menu-based studies.

With the car options study, we measured the time to complete each of the eight repeated menu tasks (as shown in Figure 1). Not surprisingly, the time to complete the task decreased, as shown in Figure 3 below:

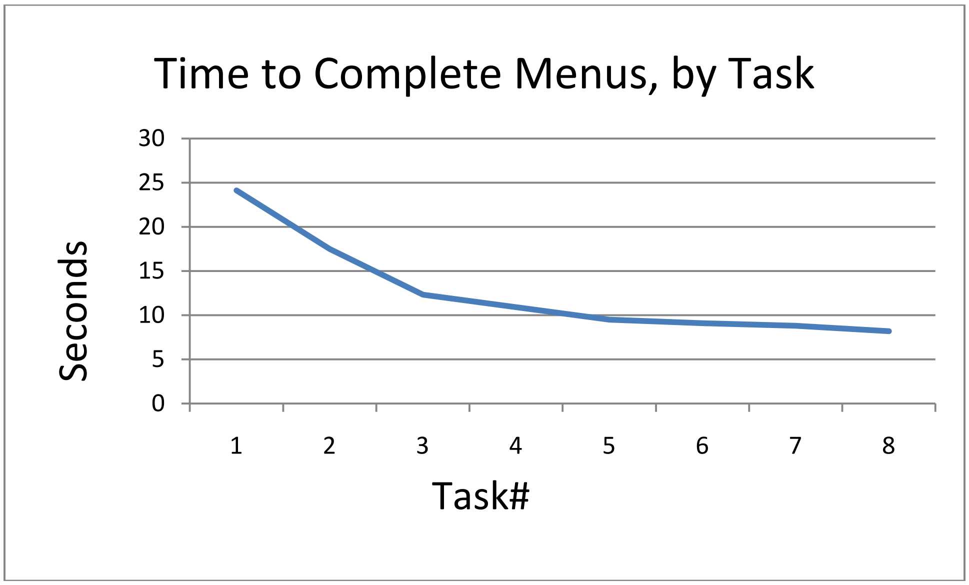

Figure 3

The car option menu was quite focused and not very complicated (Figure 1). We first asked respondents to select three vehicles in their consideration set, and how much they expected to pay for each of those vehicles. Then, we inserted the first mention with a price $2,000 below the expected price in the menu (Figure 1). We repeated the task eight times, where the only things that changed from menu to menu (task to task) were the prices for the features. As expected, time to complete decreased in later tasks. The first task took 24 seconds median time to complete, and the last tasks took about 8-9 seconds.

When fielding such a menu-based task, one naturally wonders whether many respondents consistently select the same options from the menu, irrespective of price changes. For our sample of 806 respondents, 16% of them (130 respondents) made the same choices all eight times, irrespective of price changes. 25% (204 respondents) made identical choices for the 5th through 8th tasks. For the 130 respondents who repeated their answers all eight times, only 26 of them selected no options (a blank menu), which we prompted respondents that they could do. My view of these data is that it generally reflects engaged respondents.

We can also count the percent of times respondents chose different options for the automobiles, and summarize them by task (Figure 4).

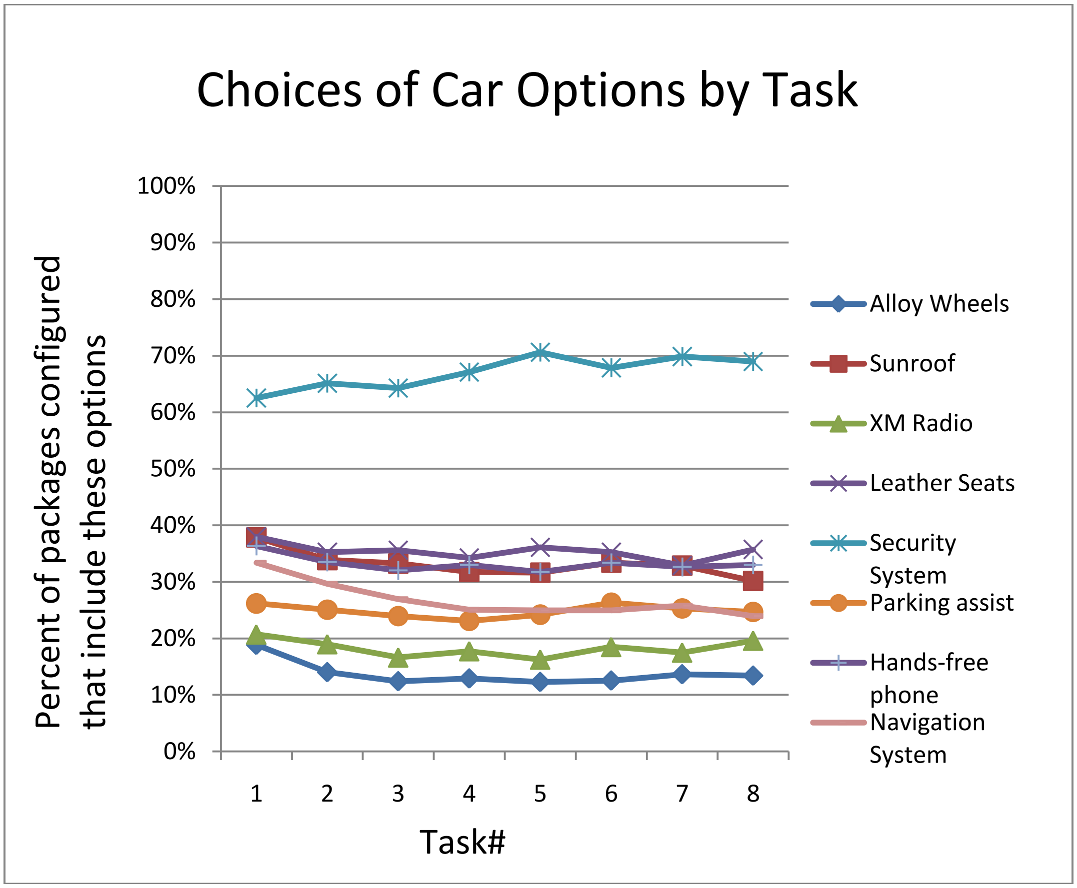

Figure 4

With 806 respondents, the probabilities for each data point above have a margin of error as large as +/-3.5%. There seems to be a slight upward trend for adding a Security System to the vehicle, and a very slight downward trend for the other options. But, the shifts in preference are not very large, suggesting a great degree of stability in terms of aggregate preferences for the options, across tasks.

Figure 5 shows the choices for menu items for the fast-food menu study, by task.

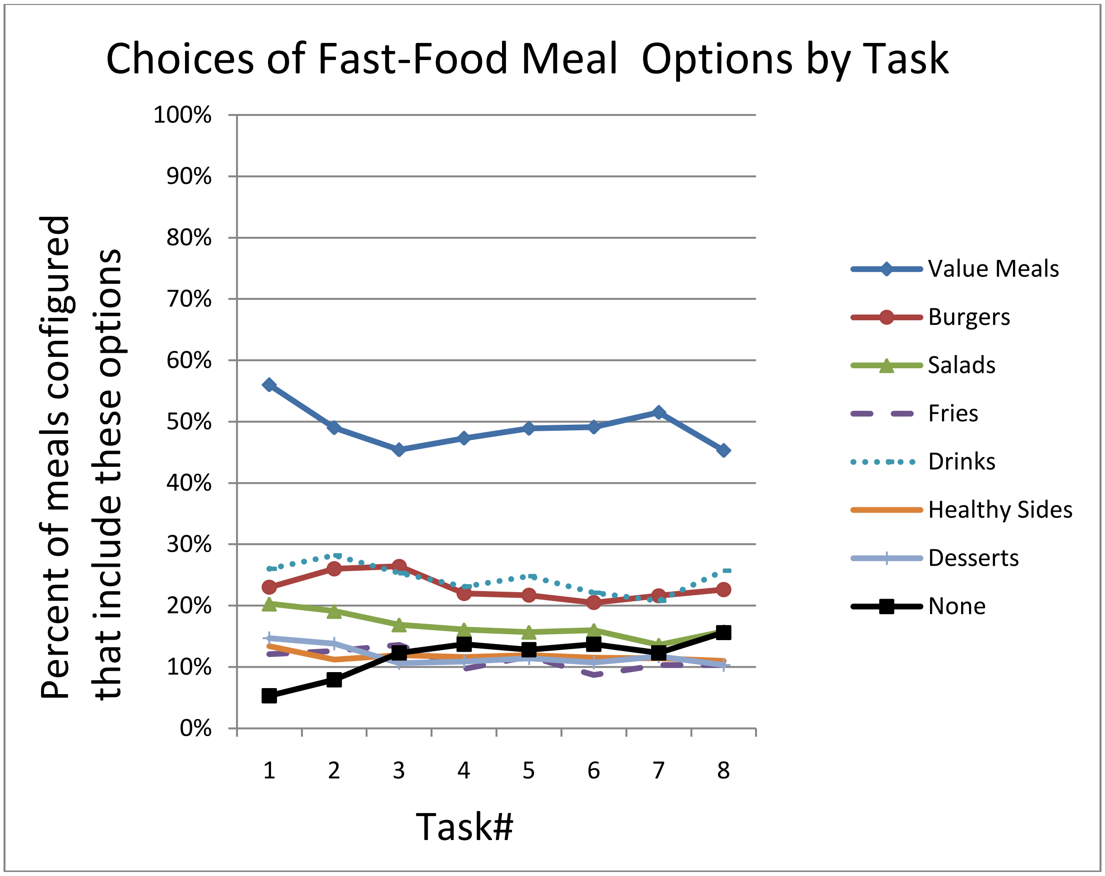

Figure 5

With the fast-food menu study, a key difference is that we offered a None option. Respondents could opt to purchase nothing from the menu. In the car menu scenario, we told respondents that they should assume they were buying this vehicle, and were to add any options that were not in the base model. The clear trend from the fast-food menu selections is the increase in None choices, from about 5% to triple that amount (15%) by the last task. And, as None trends upward, choice likelihood for other items on the menu, naturally, trend lower.[2]

When respondents learn from previous tasks that items can appear at lower prices, they likely became more selective in later menus and are more likely to reject the entire menu (if the prices shown for the desired items are viewed as too high).

Returning to the car options study (that didn't include a None), we may summarize the total number of options selected, by task, in Figure 6.

[2] Chris Moore of GfK NOP, reported to me via email that for a recent MBC study his firm conducted (for restaurant meals), the None percentage remained very constant across all 10 tasks (n=1602). He notes that 95% of his respondents had visited the restaurant in question in the last year and thus had experienced the real menu (which the survey imitated), and importantly that there was a CBC exercise with similar menu options and prices preceding the MBC task. Those previous experiences (especially the CBC) may have reduced or eliminated the None learning effect that we observed in our two data sets (which had no preceding conjoint-like warm-up exercise).

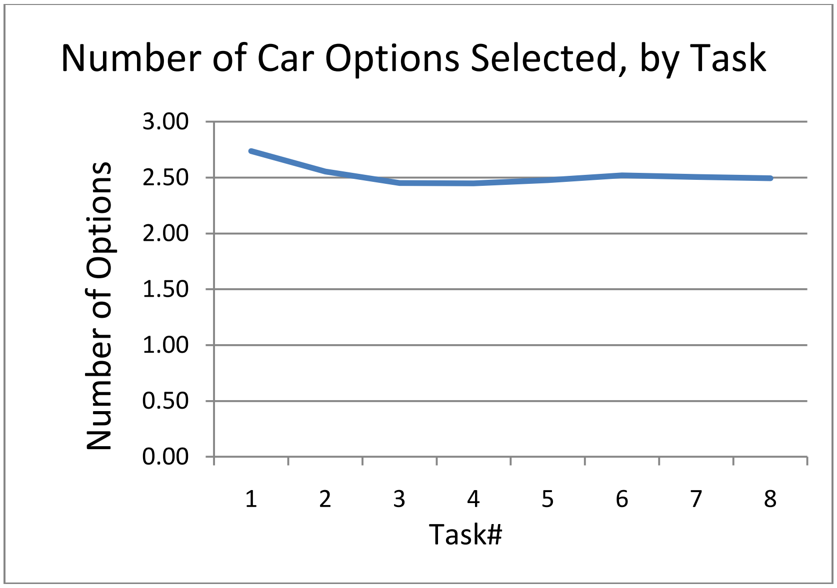

Figure 6

There is a very slight decrease in the number of options "purchased" from task 1 to task 3, but after that, it flatlines. We also examined the total price of the options "purchased" by task, in Figure 7.

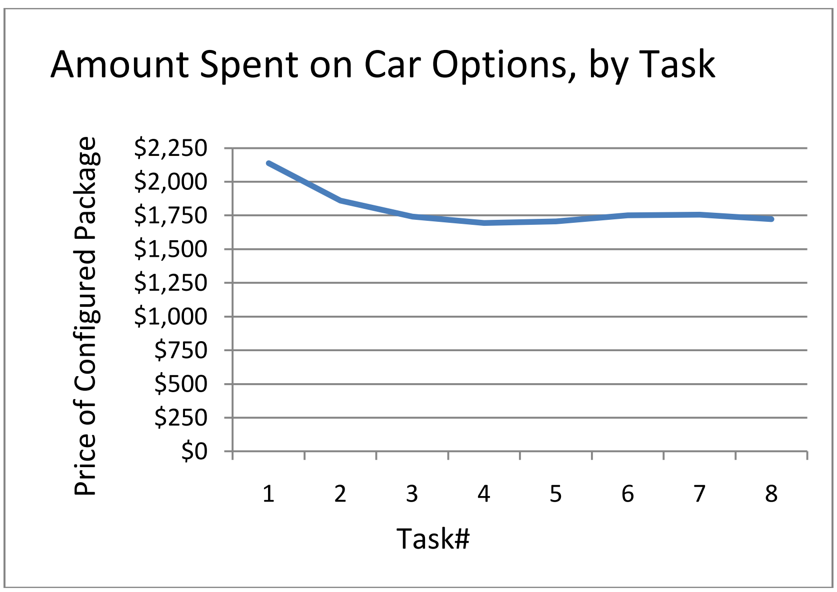

Figure 7

The amount spent trends down from tasks 1 to 3, and then flatlines. The reduction in spend from the first task to the last task is 1- $1,723/$2,138, or 19%. Could it be possible that respondents are becoming more price sensitive in later tasks, and tending to trade into options that appear "on deal"?

We examined the average price elasticity of demand for the items in the menu, by task, for both the car options study and the fast-food study. To do this, we used Counting analysis. In counting analysis, we simply compute the percent of times each item was chosen, when shown at different prices. For example, Alloy Wheels could appear at $1500, $1750, $2000, or $2500. The percent of times Alloy Wheels was selected at each of the prices was:

Table 1

Alloy Wheel Selection Probability by Price

$1,500	17.4%
$1,750	15.4%
$2,000	11.6%
$2,500	10.6%

We can compute the Price Elasticity of Demand by taking the natural log of each of the columns above, and running a univariate regression. The beta is the price elasticity.

We averaged the prices and the choice probabilities for all items on the menu, and computed the average price elasticity, by each task. We rescaled the elasticities within each study so that task 1 had an elasticity of -1.

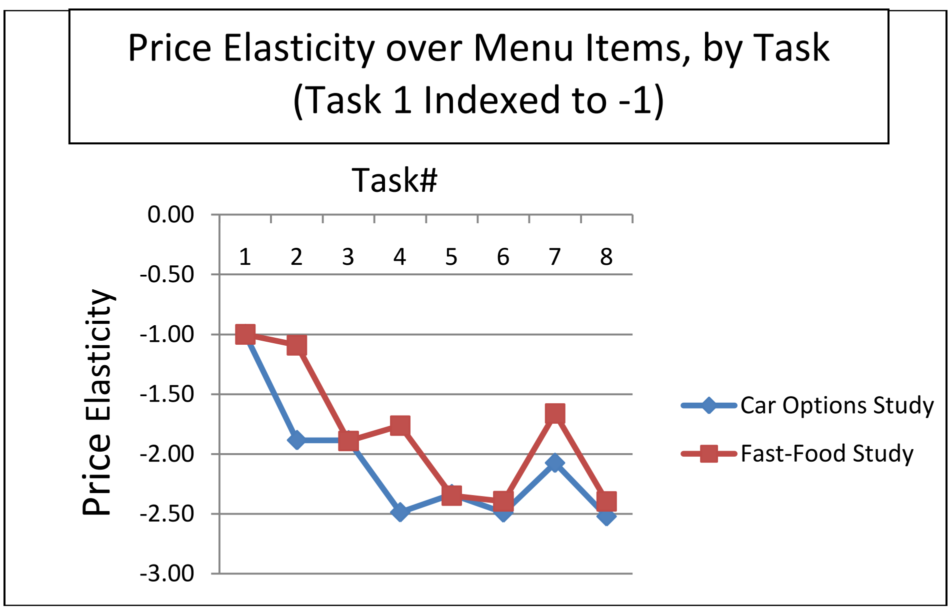

Figure 8

It is clear that the price elasticity increases dramatically in the first few tasks.[3] By the fifth task, the price elasticity has increased by a factor of nearly 2.5, relative to task 1.

[3] Jack Horne examined a recent BYO (configurator) choice task his firm (Market Strategies International) had conducted, and reported to me via email that he also found the slope of price increased dramatically for later tasks. For his dataset, the relative price slope (averaged across 5 features) by task position was -1.0, -1.9, -2.0,

With conjoint tasks and CBC tasks, there is a lot more noise (response error) in the first task(s) compared to later tasks. With an increase in response error, the differences in aggregate choice probabilities will appear to be damped. This is often referred to as "scale factor" in the choice literature. The only variables that we changed in these two menus were the prices for the items, so we cannot distinguish whether the increase in price sensitivity is due to less error (and larger scale factor) or a true increase in price sensitivity from earlier tasks to later tasks due to learning effects. Even so, the implications for Menu-Based Choice studies are clear: *if you use just one choice task, the slope of the estimated price curves will be much flatter than if using later tasks*.

Many CBC researchers have felt that respondents tend to be less price sensitive when making hypothetical choices within the questionnaire than if making real choices in the real world. Indeed, it has been shown that as CBC tasks are made more realistic, with respondents perceiving that there are real economic and utility consequences for their choices, the price sensitivity increases (Ding and Huber, 2009).

A clear weakness of our research is that we do not know the true elasticity of the items in either of our experiments, so we cannot know whether the earlier MBC tasks are more valid than the later. There is published evidence (based on CBC studies using multiple tasks) that CBC produces price sensitivities for packaged goods that are somewhat *more* price sensitive with respect to economic models based on scanner sales data (Rogers and Renken, 2003). But, many experienced conjoint researchers believe that respondents tend to display *not enough* price sensitivity when making hypothetical choices. There indeed is a huge difference in estimated price sensitivity between the first few and later MBC tasks, and I suspect that that later MBC tasks are probably more valid than earlier ones.

Respondent Perception of MBC Studies

It is well-known that CBC studies can be tedious for respondents. Are MBC studies also taxing? In our car options study, we included a second, more complex menu exercise (again with 8 tasks), that we have not analyzed for this current article (Figure 9).

-2.0, -2.5, -3.0, -3.1 for seven tasks. Chris Moore (GfK) examined his Restaurant Meals menu tasks, and found no changes in price sensitivity across tasks (but all respondents had earlier completed a full CBC exercise with features and prices prior to the menu-based task).

If you were deciding between the following three cars, and the prices were as shown, which car would you select, and which options would you add to it?

Honda Accord		BMW 5-Series		Hyundai Accent	
Base Price: $25,000		Base Price: $35,000		Base Price: $23,000	
$1,500	Alloy Wheels	$2,500	Alloy Wheels	$1,750	Alloy Wheels
$900	Moonroof/Sunroof	$700	Moonroof/Sunroof	$500	Moonroof/Sunroof
$400	XM Radio (+ $13/month)	$600	XM Radio (+ $13/month)	$500	XM Radio (+ $13/month)
$1,600	Navigation system (in dash)	$1,000	Navigation system (in dash)	$2,000	Navigation system (in dash)
Total: $25,000		Total: $35,000		Total: $23,000	

Figure 9

We displayed three cars on the screen (vehicles in each respondent's consideration set) at expected base prices, and varied the options' prices for each of the three cars, as well as the base price of each vehicle. Respondents were asked to choose which car they would most likely purchase, and which of four options they would configure for that chosen vehicle. Additional profiling questions and usage questions were asked. In all (two MBC sections with 16 total tasks plus additional survey questions), the survey lasted a median time of about 8 minutes for the panelists.

After completing the MBC exercises (16 total tasks), we asked respondents their perceptions of the questionnaire on a five-point scale. The mean responses are reported in Table 2.

The survey was at times monotonous and boring. (5=Strongly Agree, 1=Strongly Disagree)	2.0
I'd be very interested in taking another survey just like this in the future. (5=Strongly Agree, 1=Strongly Disagree)	4.6
The survey format made it easy for me to give realistic answers that reflect exactly what I'd do if buying a <insert product>. (5=Strongly Agree, 1=Strongly Disagree)	4.3
The way the <insert product> were presented made me want to slow down and make careful choices. (5=Strongly Agree, 1=Strongly Disagree)	4.1
How would you compare your overall experience with this survey compared to other internet surveys you have completed? (5=far better, 4=better, 3=about the same, 2=worse, 1=FAR worse).	4.1

Table 2

Even after completing 16 repetitive menu tasks, where the only thing that changed across the tasks was price, respondents still tended to rate the experience quite positively.

Some Conclusions:

Researchers in MBC studies may wish to provide a few warm-up tasks, to familiarize respondents with the nature of the exercise. These tasks may be discarded. If you include these tasks in the analysis, the estimated sensitivities to factors such as price may be considerably damped.

The choice of features/options (both number and general composition) across successive tasks remains relatively consistent.

If a None option is included, None usage will likely increase in later tasks.

Our panelists seem willing and able to complete multiple menu-based choice tasks. In our two studies, the only things that changed between menus were the prices. One would expect this to potentially be quite monotonous, but respondents reported relatively high satisfaction for a menu-based choice survey that included 16 menu tasks. Other MBC studies may break up the monotony by changing other aspects of the menu, such as composition of the menu and layout.

Much can be learned regarding respondent preferences within MBC using the simple method of Counts analysis. If more sophisticated analysis is required, models may be estimated leading to choice simulators. Approaches for doing this are described in Liechty *et al.* (2001), Baaken and Bond (2004), and Cohen and Liechty (2007), as well as tutorials at Sawtooth Software and SKIM training events (Orme, 2006, 2007a, 2007b, 2009). We plan to continue our investigation of methods for analyzing MBC, as menu-based experiments seem to offer some unique benefits for many kinds of market research problems, but can be more complex to analyze than traditional CBC.

References

Bakken, David G. and Megan Kaiser Bond (2004), "Estimating Preferences for Product Bundles vs. a la carte Choices," Sawtooth Software Conference Proceedings, pp 123-134.

Cohen, Steven H. and John C. Liechty (2007), "Have it Your Way: Menu-based conjoint analysis helps marketers understand mass customization," Marketing Research Magazine, Fall 2007, pp 28-34.

Ding, Min, and Joel Huber, "When Is Hypothetical Bias a Problem in Choice Tasks, and What Can We Do About It?" Sawtooth Software Conference Proceedings, 2009, pp 263-272.

Johnson, Rich, and Bryan Orme (1996), "How Many Questions Should You Ask in Choice-Based Conjoint Studies?" 1996 ART/Forum, available for download from http://www.sawtoothsoftware.com/download/techpap/howmanyq.pdf.

Liechty, John, Venkatram Ramaswamy, and Steven H. Cohen (2001), "Choice Menus for Mass Customization: An Experimenal Approach for Analyzing Customer Demand with an Application to a Web-Based Information Service," Journal of Marketing Research, May 2001, pp 183-196.

Orme, Bryan (2006), "Advanced Applications of CBC/Web Software," Tutorial presented at the SKIM/Sawtooth Software Training, Munich, Germany, October, 2006.

Orme, Bryan (2007a), "Advanced Applications of CBC/Web Software," Tutorial presented at the SKIM/Sawtooth Software Training, Dusseldorf, Germany, May, 2007.

Orme, Bryan (2007b), "Advanced CBC/Web Applications for Power Users," Tutorial presented at the Sawtooth Software Conference, Santa Rosa, California, October, 2007.

Orme, Bryan (2009), "Turbo CBC: A Power User's Forum on Discrete Choice Modeling," Turbo CBC Workshop, Anaheim, California, October, 2009.

Rogers, Greg and Tim Renken (2003), "Validation and Calibration of Choice-Based Conjoint for Pricing Research," Sawtooth Software Conference Proceedings, pp 209-215.