# Self-Reflection and Articulated Consumer Preferences

John Hauser
Songting Dong
Min Ding

In publication as: Hauser, John R., Songting Dong, and Min Ding (2014), " Self-Reflection and Articulated Consumer Preferences," *Journal of Product Innovation Management*, 31, 1, 17-32.

This is the pre-peer reviewed version of the article, which has been published in final form at http://doi.org/10.1111/jpim.12077. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.

John R. Hauser is the Kirin Professor of Marketing, MIT Sloan School of Management, Massachusetts Institute of Technology, E62-538, 77 Massachusetts Avenue, Cambridge, MA 02139, (617) 253-2929, hauser@mit.edu. All of John Hauser's papers are available at web.mit.edu/hauser/www.

Songting Dong is a Senior Lecturer in Marketing, Research School of Management, the Australian National University, Canberra, ACT 0200, Australia, +61 (2) 6125-7343, songting.dong@anu.edu.au.

Min Ding is the Smeal Professor of Marketing and Innovation, Smeal College of Business, Pennsylvania State University, University Park, PA 16802-3007; (814) 865-0622, minding@psu.edu.

## Bio for John Hauser

John R. Hauser is the Kirin Professor of Marketing at the MIT Sloan School of Management where he teaches new product development, marketing management, competitive marketing strategy, and research methodology. Among his awards include the Converse Award for contributions to the science of marketing and the Parlin Award and the Churchill Award for contributions to marketing research. He has published extensively and has been fortunate to earn a variety of best-paper awards. He has consulted on product development, sales forecasting, marketing research, voice of the customer, defensive strategy, and R&D management. He is a founder and principal at Applied Marketing Science, Inc., a former trustee of the Marketing Science Institute, a fellow of INFORMS, a fellow and President-elect of the INFORMS Society of Marketing Science, and serves on many editorial boards. He enjoys sailing, NASCAR, opera, and country music.

## Bio for Songting Dong

Songting Dong is a Senior Lecturer in Marketing at the Research School of Management, the Australian National University. He received his Bachelor's and Doctor's degree of Management from Tsinghua University. Part of his PhD training was undertaken at the Smeal College of Business at the Pennsylvania State University. His research interests include designing new methods to measure consumer preferences, decision rules and loyalty, finding insights in customer loyalty, satisfaction, and brand equity, and improving portfolio management in new product development. His papers appeared in journals such as *Journal of Marketing Research* and *International Journal of Research in Marketing*.

## Bio for Min Ding

Min Ding is Smeal Professor of Marketing and Innovation at Smeal College of Business, the Pennsylvania State University. Min received his Ph.D. in Marketing (with a 2nd concentration in Health Care System) from Wharton School of Business, University of Pennsylvania (2001), his Ph.D. in Molecular, Cellular, and Developmental Biology from the Ohio State University (1996), and his B.S. in Genetics and Genetic Engineering from Fudan University (1989). His work endeavors to provide theories and tools that have significant value to the practice of marketing, and more generally, economic development in human context. He enjoys working on problems where potential solutions require substantial creativity and major deviation from the received wisdom. Min received the Maynard Award in 2007, Davidson Award in 2012, and his work has also been voted as Paul Green Award finalists (2006 and 2008) and O'Dell Award finalist (2010). Min is V.P. of membership for the INFORMS Society for Marketing Science (ISMS). He is a diehard trekkie, and author of The Enlightened, a novel.

# Self-Reflection and Articulated Consumer Preferences

## Abstract

Accurate measurement of consumer preferences reduces development costs and leads to successful products. Some product-development teams use quantitative methods such as conjoint analysis or structured methods such as Casemap. Other product-development teams rely on unstructured methods such as direct conversations with consumers, focus groups, or qualitative interviews. All methods assume that measured consumer preferences endure and are relevant for consumers' marketplace decisions. This article suggests that if consumers are not first given tasks to encourage preference self-reflection, unstructured methods may not measure accurate and enduring preferences.

This article provides evidence that consumers learn their preferences as they make realistic decisions. Sufficiently challenging decision tasks encourage preference self-reflection which, in turn, leads to more accurate and enduring measures. Evidence suggests further that if consumers are asked to articulate preferences before self-reflection, then that articulation interferes with consumers' abilities to articulate preferences even after they have a chance to self-reflect.

The evidence that self-reflection enhances accuracy is based on experiments in the automotive and mobile-phone markets. Consumers completed three rotated incentive-aligned preference measurement methods (revealed-preference measures [as in conjoint analysis], a structured method (Casemap), and an unstructured preference-articulation method). The stimuli were designed to be managerially relevant and realistic (53 aspects in automobiles, 22 aspects for mobile phones) so that consumers' decisions approximated *in vivo* decisions. One-to-three weeks later consumers were asked which automobiles (or mobile phones) they would consider. Qualitative comments and response times are consistent with the implications of the measures of predictive ability.

Keywords: *Automotive industry, conjoint analysis, conjunctive rules, consideration sets, enduring preferences, incentive alignment, self-reflection learning, lexicographic rules, revealed preference, self-explication, voice of the customer methods.*

## Consumer's Automotive Preferences Change During Evaluation

Accurate measures of consumer preferences are especially critical in the automotive industry where even modest improvements in accuracy and insights during the design phase can dramatically reduce development costs (estimated to be $1-2 billion or more for a new platform) and increase the odds and magnitude of success. But if consumer preferences change by self-reflection during the buying process, then preferences that are measured prior to that change may lead to designs that do not meet consumer needs. The following vignette illustrates how preferences change through self-reflection.

Maria was happy with her 1995 Ford Probe. It was a sporty and stylish coupe, unique and, as a hatchback, versatile, but it was old and rapidly approaching its demise. She thought she knew her preferences which she stated as a conjunctive consideration rule—a sporty coupe with a sunroof, not black, white or silver, stylish, well-handling, moderate fuel economy, and moderately priced. Typical of automotive consumers she scoured the web, read *Consumer Reports*, and identified her consideration set based on her stated criteria. She learned that most new sporty, stylish coupes had a chop-top style with poor visibility and even worse trunk space. With her growing consumer expertise, she changed her decision rules to retain must-have rules for color, handling, and fuel economy, drop the must-have rule for a sunroof, add must-have rules on visibility and trunk space, and relax her price tradeoffs. The revisions to her preferences were triggered by the choice-set context, but were the result of extensive self-reflection about her preference. She retrieved from memory visualizations of how she used her Probe and how she would likely use the new vehicle. As a consumer, novice to the current automobile market, her stated preferences predicted she would consider a Hyundai Genesis Coupe, a Nissan Altima Coupe, and

a certified-used Infiniti G37 Coupe; as a more-expert automotive consumer her stated preferences predicted she would consider an Audi A5 and a certified-used BMW 335i Coupe.

But the story is not finished. Maria was thrifty and continued to drive her Probe until it succumbed to old age at which time she used the web (e.g., cars.com) to identify her choice set and make a final decision. Her final decision rule was the same rule she stated after gathering information and thinking deeply about her preferences. She bought a sporty coupe without a sunroof, not black, white or silver, stylish, well-handling, moderate fuel economy, and good trunk room (an Audi A5).

Maria's (true) story of thinking deeply about her preferences illustrates that consumers' preferences change as they evaluate information about available automotive products. Such changes are important to product development. Many product development decisions are tied closely to the voice of the customer (e.g., Griffin and Hauser 1993), and there is evidence that listening to the consumer leads to success (e.g., Callahan and Lasry 2004; Cooper and Kleinschmidt 1987; Hauser 2001; Montoya-Weiss and Calantone1994). Methods to obtain information on consumer preferences range from focus groups and qualitative interviews, where consumers are asked directly about their preferences, to more-formal quantitative research such as virtual-customer methods and conjoint analysis (e.g., Dahan and Hauser 2002; Green and Srinivasan 1990; Moore, Louviere, and Verma 1999; Pullman, Moore, and Wardell 2002). "General Motors (GM) alone spends tens of millions of dollars each year searching for new needs combinations and studying needs combinations when they have been identified (Urban and Hauser 2004, p. 73)."

If consumers' preferences change (then endure) after thinking deeply about their preferences (self-reflection), products designed based on preferences stated *prior to self-reflection* may

not reflect consumers' true preferences. Here true preferences mean the preferences consumers use to make decisions after a serious evaluation of the products that are available on the market. It is possible that some preference-measurement methods are more sensitive to self-reflection than others. It is also possible that some methods are best used only after self-reflection is induced.

This article explores whether Maria's story generalizes and whether or not self-reflection affects insights obtained with various preference elicitation methods. Evidence suggests that a commonly-used method, where consumers are asked to articulate their preferences, is sensitive to self-reflection. Articulated preferences change and are more accurate after self-reflection is induced. On the other hand, more-structured methods, such as conjoint-analysis-like procedures and Casemap, themselves induce self-reflection. The primary application draws on an automotive study in which consumers evaluate vehicles on 53 aspects. (Following Tversky 1972, an aspect is a feature-level.) A second application to Hong Kong mobile phones reinforces the results of the automotive study and demonstrates an interference phenomenon. Asking consumers to articulate preferences prior to self-reflection interferes with their ability to articulate preferences after self-reflection.

The following brief review of related theories in the study of consumer psychology motivates the phenomenon of preference changes based on self-reflection.

## Related Theories from Consumer Psychology

A major tenet of modern consumer behavior theory is that consumers construct their decision rules (preferences) based on the choice context (e.g., Bettman, Luce, and Payne 1998, 2008; Lichtenstein and Slovic 2006; Payne, Bettman, and Johnson 1992, 1993; Slovic, Griffin, and Tversky 1990.) For example, under time pressure consumers use simpler decision rules,

place greater importance on a few features, or only consider a few alternatives. Maria's story is consistent with the theory of constructed decision rules—she changed her decision rules (preferences) after seriously evaluating current automotive products. However, there is a subtle difference from the popular interpretation of constructed preference theory. After thinking deeply about her preferences, Maria's decision rule was enduring. It did not change when she made her purchase three months later.

Another tenet is that experts' decision rules are different and more accurate than novices' decision rules (e.g., Alba and Hutchinson 1987, 2000; Brucks 1985; Hansen and Helgeson 1996; Newell, et. al. 2004). Maria's self-reflection helped her gain expertise in her own preferences. Maria gained knowledge about the automotive market through active search. A challenging preference-elicitation method might also enable consumers to think deeply about their preferences.

A third tenet is that consumers, when faced with a choice among many products or based on many features, simplify their decision processes with (a) two-step consider-then-choose processes (Hauser and Wernerfelt 1990; Payne 1976; Roberts and Lattin 1991; Punj and Moore 2009) and (b) simplified decision rules such as conjunctive, lexicographic, or other non-compensatory rules (*supra* citations plus Bröder 2000; Gigerenzer and Goldstein 1996; Martignon and Hoffrage 2002; Thorngate 1980). Automotive choice has many alternatives (200+ make-model combinations) and many features (53 aspects in our empirical example). Maria used both simplifications. Although we do not seek to test these simplifications, our experiments must take them into account.

Finally, there is evidence in psychology that consumers learn their own preferences as they complete intensive tasks that ask them to use or articulate preferences. For example, Betsch, et al. (2001) manipulated task learning through repetition and found that subjects were more like-

ly to maintain a decision routine with 30 repetitions rather than with 15 repetitions. Hansen and Helgeson (1996) demonstrated that learning cues influence naïve decision makers to behave more like experienced decision makers. Garcia-Retamero and Rieskamp (2009) describe experiments where subjects shift from compensatory to conjunctive-like decision rules over seven trial blocks. Hoeffler and Ariely (1999) show that effort and experience improve the stability of compensatory tradeoffs (e.g., sounds that vary on three features). Hard choices reduced violations and increased subjects' confidence in stated preference tradeoffs. Simonson (2008) provides examples where preferences are learned through experience and endure. "Once uncovered, inherent (previously dormant) preferences become active and retrievable from memory (Simonson 2008, p. 162)." These are some of the many citations consistent with learning through self-reflection.

## Data Used to Explore Self-Reflection Learning

The following proposition generalizes Maria's story.

**Self-reflection Learning**. *If consumers are given a sufficiently realistic task and enough time to complete that task, consumers will think deeply about their preferences (and potential consumption of the products). Such self-reflection helps consumers clarify and articulate their preferences. After self-reflection, articulated preferences endure and are more likely to provide accurate insight into the voice of the customer.*

We explore self-reflection by reanalyzing data that were collected to compare new methods to measure and elicit consideration rules (Ding et al. 2011). We focus on the impact of self-reflection and do not repeat the evaluation of the preference-elicitation methods *per se*. As suggested by Payne, Bettman, and Johnson (1993, p. 252) these data were focused on consideration-set decisions. Consideration is an important managerial issue in automotive product design. In one study roughly half of US consumers would not even consider vehicles from General Motors

(GM, Hauser, et al. 2010, p. 485). Ingrassia (2010, p. 163) suggests that a major reason for GM's financial troubles was related to the fact that "many GM brands didn't even make the 'consideration list' of young shoppers." As a result US automakers have invested heavily in methods to understand consumers' decision rules for consideration (e.g., Dzyabura and Hauser 2011).

The data approximated *in vivo* decision making. For example, all measures were incentive-aligned (consumers had a reasonable chance of receiving a $40,000 vehicle based on their answers), the consideration-set decision approximated as closely as feasible marketplace automotive-consideration-set decisions, and the 53-aspect feature set was adapted to our target consumers but drawn from a large-scale national study by a US automaker. The automaker's study, not discussed here, involved many automotive experts and managers and provided essential insight to support the automaker's attempt to emerge from bankruptcy and regain a place in consumers' consideration sets. These characteristics helped assure that the feature-set was realistic and representative of decisions by real automotive consumers. A second study was also incentive aligned and approximated as closely as feasible mobile-phone decisions in Hong Kong. Data which approximate *in vivo* decisions are messier than focused *in vitro* experiments, but new phenomena are more likely to emerge (Greenwald, et al. 1986).

## Task Order Suggests Self-Reflection Learning

*Overview of the Automotive Design*

Respondents were relative novices with respect to automotive purchases, but interested in the category and likely to make a purchase in a year or two. In a set of rotated online tasks, consumers were (1) asked to form consideration sets from a set of 30 realistic automobile profiles chosen randomly from a 53-aspect orthogonal set of automotive features, (2) asked to state their preferences through a structured preference-articulation procedure (Casemap, Srinivasan and

Wyner 1988), and (3) asked to state their consideration rules in an unstructured e-mail to a friend who would act as their agent. Task details are given below. Prior to completing these rotated tasks they were introduced to the automotive features by text and pictures and, as training in the features, asked to evaluate 9 profiles for potential consideration. Consumers completed the training task in less than $1/10^{th}$ the amount of time observed for any of the three primary tasks. One week later consumers were re-contacted and asked to again form consideration sets, but this time from a different randomly-chosen set of 30 realistic automobile profiles.

The study was pretested on 41 consumers and, by the end of the pretests, consumers found the survey easy to understand and representative of their decision processes. The primary sample of 204 consumers agreed. On five-point scales, the tasks were easy to understand (2.01, SD = .91, where 1 = "extremely easy") and easy for consumers to understand that it was in their best interests to tell the true preference (1.81, SE = .85, where 1 = "extremely easy"). Consumers felt they could express their preferences accurately (2.19, SD=.97, where 1 = "very accurately").

*Incentive Alignment* (*Prize Indemnity Insurance*)

Incentive alignment rather than the more-formal term, incentive compatible, is a set of motivating heuristics designed to induce (1) consumers to believe it is in their interests to think hard and tell the truth, (2) it is, as much as feasible, in their interests to do so, and (3) there is no obvious way to improve their welfare by cheating. Instructions were written and pretested carefully to reinforce these beliefs. For previous incentive-aligned preference-elicitation methods see Ding (2007), Ding, Park and Bradlow (2009), and Prelec (2004).

Designing aligned incentives for consideration-set decisions is challenging because consideration is an intermediate stage in the decision process. Kugelberg (2000) used purposefully vague statements that were pretested to encourage consumers to trust that it was in their best in-

terests to tell the truth. For example, if consumers believed they would always receive their most-preferred automobile from a known set, the best response is a consideration set of exactly one automobile.

A common format is a secret set that is revealed after the study is completed. With a secret set, if consumers' preferences screened out too few vehicles, they had a good chance of getting an automobile they did not like. On the other hand, if their preferences screened out too many vehicles none would have remained in the consideration set and their preferences would not have affected the vehicle they received. The size of the secret set, 20 vehicles, was chosen carefully through pretests. Having a restricted set has external validity (Urban and Hauser 2004). The vast majority (80-90%) of US consumers choose automobiles from dealers' inventories (Urban, Hauser, and Roberts [1990] and March 2010 personal communication from a US automaker).

All consumers received a participation fee of $15 when they completed both the initial three tasks and the delayed validation task. In addition, one randomly-drawn consumer was given the chance to receive $40,000 toward an automobile (with cash back if the price was less than $40,000), where the specific automobile (features and price) would be determined by the consumer's answers to one of the sections of the survey (three initial tasks or the delayed validation task). To simulate actual automotive decisions and to maintain incentive alignment, consumers were told that a staff member, not associated with the study, had chosen 20 automobiles from dealer inventories in the area. The secret list was made public after the study.

To implement the incentives, prize indemnity insurance was purchased where, for a fixed fee, the insurance company would pay $40,000 if a consumer won an automobile. One random consumer got the chance to choose two of 20 envelopes. Two of the envelopes contained a win-

ning card; the other 18 contained a losing card. If both envelopes had contained a winning card, the consumer would have received the $40,000 prize. Such drawings are common for radio and automotive promotions. Experience suggests that consumers perceive the chance of winning to be at least as good as the two-in-20 drawing implies (see also the fluency and automated-choice literatures: Alter and Oppenheimer 2008; Frederick 2002; Oppenheimer 2008). In the actual drawing, the consumer's first card was a winner, but, alas, the second card was not.

*Profile-Decision, Structured-Elicitation, and Unstructured-Elicitation Tasks*

*Profile Decision Task* (*Revealed Preferences*). The profile-decision task was designed to be as realistic as possible given the constraints of an online survey. Just as Maria learned and adapted her consideration rules as she searched online to replace her Probe, we wanted consumers to be able to revisit consideration-set decisions as they evaluated the 30 profiles. The computer screen was divided into three areas. The 30 profiles were displayed as icons in a "bullpen" on the left. When a consumer moused over an icon, all features were displayed in a middle area using text and pictures. The consumer could consider, not consider, or replace the profile. All considered profiles were displayed in an area on the right and the consumer could toggle the area to display either considered or not-considered profiles and could, at any time, move a profile among the considered, not-considered, or to-be-evaluated sets. Consumers took the task seriously investing, on average, 7.7 minutes to evaluate 30 profiles. This is substantially more time than the 0.7 minutes consumers spent on the sequential 9-profile training task, even accounting for the larger number of profiles ($p < .01$). Figure 1 provides one example screenshot from each of the tasks. An online appendix provides representative screenshots from the three rotated tasks.

Insert Figure 1 about here.

Table 1 summarizes the features and feature levels. The genesis of these features was the

large-scale study used by a US automaker to test a variety of marketing campaigns and product-development strategies to encourage consumers to consider its vehicles. After the automaker shared its feature list with us, we modified the list of brands and features for our target audience. The $20x7x5^2x4x3^4x2^2$ feature-level design is large by market-research standards, but remained understandable to consumers. To make profiles realistic and to avoid dominated profiles (e.g., Elrod, Louviere and Davey 1992; Johnson, Meyer and Ghose 1989; Toubia, Hauser and Simester 2004), prices were a sum of experimentally-varied levels and feature-based prices chosen to represent market prices at the time of the study (e.g., a price increment is set for a BMW relative to a Scion). Unrealistic profiles were removed if a brand-body-style did not appear in the market. (The resulting sets of profiles had a D-efficiency of .98.)

Insert Table 1 about here.

*Structured-Preference-Articulation Task* (*Casemap*). Consistent with prior theory, data were collected on both compensatory and conjunctive decision rules. The chosen method, Casemap, is used widely. Following Srinivasan (1988), closed-ended questions ask consumers to indicate unacceptable feature levels, indicate their most- and least-preferred level for each feature, identify the most-important critical feature, rate the importance of every other feature relative to the critical feature, and scale preferences for levels within each feature. Prior research suggests that the compensatory portion of Casemap is as accurate as decompositional conjoint analysis, but that consumers are over-zealous in indicating unacceptable levels (e.g., Green, Krieger and Banal 1988; Huber, et al. 1993; Sawtooth 1996; Srinivasan and Park 1997; Srinivasan and Wyner 1988). This was the case with our consumers. Predicted consideration sets were much smaller with Casemap than were observed or predicted with data from the other tasks.

*Unstructured Preference-Articulation* (*E-mail Task*). Consumers were asked to write an

e-mail to a friend who would act as their agent should they win the incentive-alignment lottery. Other than a requirement to begin the e-mail with "Dear Friend," no other restrictions were placed on what they could say. To align incentives they were told that two agents would use the e-mail to select automobiles and, if the two agents disagreed, a third agent would settle ties. (The agents would act only on the consumer's e-mail—no other personal communication.) To encourage trust, agents were audited as in Toubia (2006). Following standard procedures the data were coded by independent judges for compensatory and conjunctive rules, where the latter could include must-have and must-not-have rules (e.g., Griffin and Hauser 1993; Hughes and Garrett 1990; Perreault and Leigh 1989). Example responses and coding rules are available from the authors and published in Ding et al. (2011). Data from all tasks and validations are available from the authors.

*Predictive-Ability Measures*

The validation task occurred roughly one-week after the initial tasks and used the same format as the profile-decision task. Consumers saw a new draw of 30 profiles from the orthogonal design (different draw for each consumer). Standard methods were used to make predictions. Hierarchical Bayes "choice-based-conjoint" logit analyses estimated choices (using calibration data only) based on the profile decision task, an additive model, and the cutoff utility as described below (HB CBC, Lenk, et al. 1996; Rossi and Allenby 2003; Sawtooth 2004).① Standard Casemap analyses estimated choices based on the structured-elicitation task: unacceptable profiles are eliminated and then a compensatory rule with a cutoff utility (as described below) predicted inclusion in the consideration set. A coding methodology estimated choices based on the unstructured preference-articulation task: the stated conjunctive rules eliminated or included profiles and then the coded compensatory rules predicted inclusion in the consideration set. To pre-

dict consideration with compensatory rules a cutoff utility was established with a logit analysis of consideration-set sizes (based on the calibration data only) with stated price range, the number of non-price elimination rules, and the number of non-price compensatory rules as explanatory variables. The cutoff model was applied to all methods in a comparable way.

Predictive ability is measured with the relative Kullback-Leibler divergence (KL, also known as relative entropy). KL divergence measures the expected divergence in Shannon's (1948) information measure between the validation data and a model's predictions and provides an evaluation of predictive ability that is rigorous and discriminates well (Chaloner and Verdinelli 1995; Kullback and Leibler 1951). Relative KL rescales KL divergence relative to the KL divergence between the validation data and a random model (100% is perfect prediction and 0% is no information.) Ding et al. (2011) provide formulae for consideration-set decisions.

KL provides better discrimination than hit rates for consideration-set data because consideration-set sizes tend to be small relative to full choice sets (Hauser and Wernerfelt 1990). For example, if a consumer considers only 30% of the profiles then even a null model of "predict-nothing-is-considered" would get a 70% hit rate. On the other hand, KL is sensitive to both false positive and false negative predictions—it identifies that the "predict-nothing-is-considered" null model contains no information. (For those readers unfamiliar with KL divergence, key analyses using hit rates are repeated at the end of the article. The implications are the same.)

## Comparison Of Predictive Ability Based On Task Order

If self-reflection learning takes place predictive accuracy will depend upon the order(s) in which consumers completed the three tasks. For example, consumers should be able to articulate preferences better after completing an intensive task that causes them to think deeply about their preferences. (The literature, qualitative data, and the authors' experience studying automotive

decisions suggest that consumers form these preferences as they think seriously about their decision process. However, predictive tests alone cannot rule out hypotheses that consumers either retrieve preferences from heretofore obscured memory or are simply better able to articulate preferences. Future *in vitro* research might explore these subtly different hypotheses.)

*Analysis of Predictive Ability versus Task, Task Order, and Interactions*

Initial analysis, available from the authors, suggests that there is a task-order effect, but the effect is for first-in-order versus not-first-in-order. (That is, it does not matter whether a task is second versus third; it only matters that it is first or not first.) Based on this simplification Table 2 summaries the analysis of variance. All predictions are based on models estimated from the calibration data. All validation statistics are based on comparing these predictions to consideration-set decisions one week later.

Task ($p < .01$), task order ($p = .02$), and interactions ($p = .03$) are all significant. The task-based-prediction difference is the methodological issue discussed in Ding et al. (2011). The first-vs.-not-first effect and its interaction with task is the focus of this article. Table 3 provides a simpler summary that isolates the effect as driven by the task order of the unstructured preference-articulation task (the e-mail task).

Insert Tables 2 and 3 about here.

Consumers are better able to articulate preferences (e-mail task) that predict consideration sets one week later if they first complete either a 30-profile consideration-set decision or a Casemap structured preference-articulation ($p < .01$). The information in the predictions one-week later is over 60% larger if consumers have the opportunity for self-reflection learning via either an intensive decision task (30 profiles) or an intensive structured-articulation task (KL =

.151 vs. .093). There are hints of self-reflection learning for Casemap (KL = .068 vs. .082), but such learning, if it exists, is not significant in Table 3 ($p$ = .40). On the other hand, revealed-preference predictions (based on HB CBC) do not benefit if consumers are first asked to articulate preferences with either the e-mail or Casemap tasks. This is consistent with an hypothesis that the bullpen format allows consumers to self-reflect and revisit consideration decisions.

Attribution and self-perception theories were examined as alternative explanations to self-reflection learning. Both theories explain why validation choices might be consistent with stated choices (Folkes 1988; Morwitz, Johnson, and Schmittlein 1993; Sternthal and Craig 1982), but attribution and self-perception do not explain the order effect or the task x order interaction. All consumers completed all three tasks prior to validation. Recency hypotheses can also be ruled out because there is no effect due to second versus third in the task order.

*Summary of the Task-Order Effect for Articulated Preferences*

Table 3 implies the following observations for the automotive consideration decision:

- consumers are better able to articulate preferences to an agent after self-reflection. (Self-reflection came either from substantial consideration-set decisions or the highly-structured Casemap task.)
- the articulated preferences endure (at minimum, their predictive ability endures for at least one week)
- 9-profile, sequential, profile-evaluation training does not provide sufficient self-reflection learning (because substantially more learning was observed in Tables 2 and 3).
- Pretests indicated consumers felt the 9-profile warm-up was sufficient to introduce them to the features, the feature levels, and the task. This suggests that self-reflection learning is learning preferences and is more than simply learning the composition of the market.

If these interpretations are correct, then they have profound implications for research to support product design. To identify enduring preferences, we must first ask respondents to complete a task that enables them to learn by self-reflection. For example, when consumers have the opportunity to self-reflect prior to articulating preferences, we see significant changes toward more rules addressing EPA mileage (more must-not-have low mileage) and fewer rules addressing crash test ratings (fewer must-not-have C3). Consumers change the relative importances of quality (Q5 increases), body types (mid-size SUVs increase), transmission (automatic decreases), and engine type (hybrid decreases). These changes are significant at $p < 0.05$. Other changes in brand, body type, quality, and transmission are marginally significant.

## Qualitative Comments and Response Times

Qualitative data, response times, and decision-rule evolution were not focused systematically on self-reflection learning. Nonetheless, they provide complementary insight.

### *Qualitative Comments*

Consumers were asked to "please give us additional feedback on the study you just completed." Many of these comments were methodological: "*There was a wide variety of vehicles to choose from, and just about all of the important features that consumers look for were listed.*" However, qualitative comments also provided insight on self-reflection learning. Although there were not sufficient data for a formal content analysis, comments suggest that the decision tasks helped consumers to think deeply about imminent car purchases. (Minor spelling and grammar mistakes in the quotes were corrected.)

- *As I went through (the tasks) and studied some of the features and started doing comparisons I realized what I actually preferred.*
- *The study helped me realize which features were more important when deciding to pur-*

*chase a vehicle. Next time I do go out to actively search for a new vehicle, I will take more factors into consideration.*

- *I have not recently put this much thought into what I would consider about purchasing a new vehicle. Since I am a good candidate for a new vehicle, I did put a great deal of effort into this study, which opened my eyes to many options I did not know sufficient information about.*
- *This study made me think a lot about what car I may actually want in the near future.*
- *Since in the next year or two I will actually be buying a car, (the tasks) helped me start thinking about my budget and what things I'll want to have in the car.*
- *(The tasks were) a good tool to use because I am considering buying another car so it was helpful.*
- *I found (the tasks) very interesting, I never really considered what kind of car I would like to purchase before.*

Further comments provide insight on the task-order effect: "*Since I had the tell-an-agent-what-you-think portion first, I was kind of thrown off. I wasn't as familiar with the options as I would've been if I had done this portion last. I might have missed different aspects that could have been expanded on.*" Comments also suggested that the task caused consumers to think deeply: "*I understood the tasks, but sometimes it was hard to know which I would prefer.*" and "*It is more difficult than I thought to actually consider buying a car.*"

*Response Times*

In any set of tasks we expect consumers to learn as they complete the tasks and, hence, we expect response times to decrease with task order. We also expect that the tasks differ in difficulty. Both effects were found for automotive consideration decisions. An ANOVA with response time as the dependent variable suggests that both task ($p < .01$) and task-order ($p < .01$) are significant and that their interaction is marginally significant ($p = .07$). Consistent with self-reflection learning, consumers can articulate their preferences more rapidly (e-mail task) if either

the profile-decision task or the structured-elicitation task comes first. However, general learning cannot be ruled out because response times for all tasks improve (decrease) with task order.

The response time for the one-week-delayed validation task does not depend upon initial task order suggesting that, by the time the consumers complete all three tasks, they have learned their preferences. For those consumers for whom the profile-decision task was not first, the response times for the initial profile-decision task and the delayed validation task are not statistically different ($p$ = .82). This is consistent with an hypothesis that consumers learned their preferences and then applied those rules in both the initial and delayed tasks. Finally, as a face-validity check on random assignment, the 9-profile training times do not vary with task-order.

In summary, order effects, qualitative comments, and response times are consistent with the self-reflection-learning phenomenon—an hypothesis this article advances as a parsimonious explanation. Because articulated decision rules (and feature importances) change as the result of self-reflection, voice-of-the-customer analyses for product development should rely on tasks that allow for and enhance self-reflection learning.

## Self-reflection Learning and Task Interference

The examination of the data from a second study reinforces the lessons from the automotive study. Like the automotive study, these data were collected to examine alternative measurement methods, but reanalyses provide insight on self-reflection learning. The second study focuses on mobile phones in Hong Kong. The number of aspects is moderate, but consideration-set decisions are in the domain where decision heuristics are likely (22 aspects in a $4^5x2^2$ design). The smaller number of aspects makes it feasible to use recently-developed methods to estimate non-compensatory decision rules. (At the time of the experiments no non-compensatory methods were feasible for 53-aspect decision rules.) This second study explores whether the self-

reflection learning results rely on using HB CBC to estimate a compensatory model. The second study also examines whether the self-reflection-learning hypothesis extends to the longer time period. (There was also a validation task toward the end of the initial survey following Frederick's [2005] memory-cleansing task. Because the results based on in-survey validation were almost identical to those based on the delayed validation task, they are not repeated here. Details are available from the authors.)

The mobile-phone data also enable exploration of whether earlier preference-articulation tasks interfere with self-reflection learning. Specifically, for mobile phones the e-mail task always comes after the other two tasks—only the 32-profile profile-decision task (bullpen) and a weakly-structured preference-articulation task were rotated. The weakly-structured task was less structured than Casemap—the weakly-structured task asked consumers to state rules on a preformatted screen, but, unlike Casemap, consumers were not forced to provide rules for every feature and feature level and the rules were not constrained to the Casemap structure.

Other aspects of the mobile-phone study were similar to the automotive study. (1) The profile-decision task used the bullpen format that allowed consumers to self-reflect prior to finalizing their decisions. (2) All decisions were incentive-aligned. One in 30 consumers received a mobile phone plus cash representing the difference between the price of the phone and $HK2500 [$1 = $HK8]. (3) Instructions for the e-mail task were similar and independent judges coded the responses. (4) Features and feature levels were chosen carefully to represent the Hong Kong market and were pretested to be realistic and representative of the marketplace. Table 4 summarizes the mobile-phone features.

Insert Table 4 about here.

The 143 consumers were students at a major university in Hong Kong. In Hong Kong, unlike in the US, mobile phones are unlocked and can be used with any carrier. Consumers, particularly university students, change their mobile phones regularly as technology and fashion advance. After a pretest with 56 consumers to assure that the questions were clear and the tasks not onerous, consumers completed the initial set of tasks at a computer laboratory on campus and, three weeks later, completed the delayed-validation task on any Internet-connected computer. In addition to the chance of receiving a mobile phone (incentive-aligned), all consumers received $HK100 when they completed both the initial tasks and the delayed-validation task.

*Analysis of Task, Task Order, and Interactions* (*Mobile-Phone Study*)

Table 5 summarizes predictive ability for the mobile-phone study (ANOVA results were similar to those in the first study). Two revealed-decision-rule methods were used to estimate decision rules from the calibration profile-decision task. “Revealed preferences (HB CBC)” is as in the automotive study. “Lexicographic decision rules” estimates a conjunctive model of profile decisions using the greedoid dynamic program described in Dieckmann, Dippold and Dietrich (2009), Kohli and Jedidi (2007), and Yee, et al. (2007). (For consideration-set decisions, lexicographic rules with a cutoff give equivalent predictions to conjunctive decision rules and deterministic elimination-by-aspect rules.) Disjunctions-of-conjunctions rules were also estimated using logical analysis of data as described in Hauser, et al. (2010). The logical-analysis-of-data results are basically the same as for the lexicographic rules and provide no additional insight on task-order self-reflection-learning. Details are available from the authors.

Insert Table 5 about here.

The first predictive comparison is based on the two rotated tasks: revealed-preferences

versus weakly-structured preference-articulation predictions. The analysis reproduces and extends the results of the automotive study. Weakly-structured preference-articulation predicts significantly better if it is preceded by a profile-decision task (bullpen) that is sufficiently intense to allow self-reflection. The weakly-structured task provides over 80% more information (KL = .250 vs. .138) when consumers complete the weakly-structured task after the bullpen task rather than before the bullpen task. This is consistent with self-reflection learning. (Self-reflection learning improved Casemap predictions, but not significantly in the automotive study.) This increased self-reflection effect is likely due to the reduction in structure of the weakly-structured task. As in the automotive study, there is no order effect for revealed-preference models, neither for HB CBC nor greedoid-estimated lexicographic decision rules.②

*Interference from the Weakly-Structured Task*

The automotive study demonstrated self-reflection learning, but the mobile-phone study includes a task order in which consumers are asked to articulate preferences prior to self-reflection, then allowed to self-reflect, and then again asked to articulate preferences. It is possible that the first weakly-structured preference-articulation task could interfere with the later post-self-reflection unstructured preference-articulation task. For example, Russo and Schoemaker (1989) provide examples were decision makers do not adjust sufficiently from initial conditions. Russo, Meloy, and Medvec (1998) suggest that pre-decision information is distorted to support brands that emerge as leaders early in a decision process. On the other hand, if self-reflection occurs before both articulation tasks, learning due to self-reflection might endure. For example, Rakow, et al. (2005) give examples where, with substantial training, consumers do not change their decision rules under higher search costs or time pressure.

The last rows of Table 5 examine interference. The data suggest that the predictive ability

of the end-of-survey e-mail task is best if consumers evaluate 32 profiles using the bullpen format before they complete the weakly-structured task. This effect is observed even though, by the time they begin the e-mail task, *all* consumers have evaluated 32 profiles using the bullpen.

The weakly-structured preference-articulation task appears to interfere with (suppress) consumers' abilities to articulate preferences in the e-mail task even though consumers had the opportunity for self-reflection between the weakly-structured and e-mail tasks. Consumers appear to anchor to the preferences they state in the weakly-structured task and this anchor interferes with consumers' abilities to later articulate their preferences.

Fortunately not all predictive ability is lost due to interference. Even when interference occurs, the preferences stated in the e-mail task predict reasonably well. They predict significantly better than rules from the pre-self-reflection weakly-structured task ($p < .01$) and comparably to rules revealed by the bullpen task ($p = .15$). The best predictions overall are by the e-mail task when there is self-reflection, but no interference.

The weakly-structured task does not affect the predictive ability of preferences revealed by the bullpen task. In the bullpen task consumers appear to have sufficient time and motivation for self-reflection. The following diagram summarizes the relative predictive ability observed in the mobile-phone study as follows where $\succ$ implies statistical significance and ~ implies no statistical difference:

$$\begin{matrix} Articulated\ (email) \\ after\ self\ reflection \\ no\ interference \end{matrix} \succ \left\{ \begin{matrix} Articulated\ (email) \\ after\ interference \\ and\ self\ reflection \end{matrix} \right. \sim \left. \begin{matrix} Revealed\ rules \\ allowing\ for \\ self\ reflection \end{matrix} \right\} \succ \begin{matrix} Articulated\ rules \\ (weakly\ structured) \\ before\ self\ reflection \end{matrix}$$

These analyses suggest that product developers who use methods that ask consumers to articulate preferences should give consumers sufficient opportunities to self-reflect before consumers are asked to articulate rules. Product developers must be aware of both self-reflection and

potential interference whenever they attempt to identify the voice of the customer.

As in the automotive study, different recommendations based on the two task orders are observed. Comparing preferences obtained with the weakly-structured task before vs. after self-reflection, changes are found in stated preferences for brands, color, and form. After self-reflection, Hong Kong consumers are less likely to eliminate the rotational form factor and placed more importance on the black color, flip and rotational form factors, and the Motorola, Nokia, and Sony-Ericsson brands. All task-order differences are significant at the $p < 0.05$ level. Many other differences were marginally significant.

Differences are also observed due to interference; interferences shifts articulated preferences towards pre-self-reflection preferences. With interference Hong Kong consumers articulate rules that are less accepting of flip form factors and place lower importances on the Motorola, Nokia, and Sony-Ericsson brands. Nokia and Sony-Ericsson brands were significantly lower ($p < 0.05$); form-factor differences were marginally significant ($p < 0.10$).

*Qualitative Comments and Response Times for the Mobile-Phone Study*

The qualitative data collected for the mobile-phones were less extensive than in the automotive study. However, qualitative comments suggest self-reflection learning (spelling and grammar corrected):

- *It is an interesting study, and it helps me to know which type of phone features that I really like.*
- *It is a worthwhile and interesting task. It makes me think more about the features which I see in a mobile phone shop. It is an enjoyable experience!*
- *Better to show some examples to us (so we can) avoid ambiguous rules in the (profile-decision) task.*
- *I really can know more about my preference of choosing the ideal mobile phone.*

As in the automotive study, consumers completed the bullpen task faster when it came af-

ter the structured-elicitation task ($p$ = .02). However, unlike the automotive study, task order did not affect response times for the weakly-structured preference-articulation task ($p$ = .55) or the e-mail task ($p$ = .39). (The lack of response-time significance could also be due to the fact that English was not a first language for most of the Hong Kong consumers.)

## Hit Rates as a Measure of Predictive Ability

Hit rates are a less discriminating measure of predictive validity than KL divergence, but the hit-rate measures follow the same pattern as the KL measures. Hit rate varies by task order for the automotive e-mail task (marginally significant, $p$ = .08), for the mobile-phone weakly-structured task ($p$ < .01), and for the mobile-phone e-mail task ($p$ = .02). See Table 6.

Insert Table 6 about here.

## Summary

Understanding consumer preferences is key to product development success, especially in automotive markets where launching a new vehicle requires investments in the order of $1-2 billion (Urban and Hauser 2004, p. 72). The automotive and mobile-phone experiments suggest that the accuracy of consumer-preference measures, and the implications for product design, depend upon providing consumers sufficient opportunity to self-reflect prior to asking them to articulate their preferences. Researchers should also be aware of interference. If consumers are asked to articulate preferences prior to self-reflection, then initial articulation errors endure and reduce accuracy.

There is good news for widely applied methods. Revealed-preference methods such as conjoint analysis and highly-structured methods such as Casemap induce self-reflection. These

methods measure enduring preferences.

However, product-development teams often rely on talking directly to consumers or on unstructured methods such as focus groups or qualitative interviews. The experiments in this paper suggest that unstructured methods may not measure enduring preferences unless consumers are given an opportunity to self-reflect. When product development teams rely on unstructured methods, they should use tasks that induce self-reflection. Fortunately, unstructured preference-articulation methods can be accurate in measuring enduring preferences when consumers have a chance to self-reflect (review Tables 3 and 5).

The experiments in this paper also caution product development teams to pay attention to the role of warm-up questions. In the automotive experiment consumers were given detailed descriptions of the features-levels and were asked to evaluate nine profiles. But these warm-up questions were not sufficient to induce self-reflection. Self-reflection, and accurate enduring preferences, required a more substantial task.

This article focused on the implications for product development, but the concept of self-reflection is likely to apply more generally to the study of constructed preferences. There are other methods, other contexts, and other situations in which self-reflection might or might not be important. Self-reflection seems to explain the data and is consistent with prior literature and qualitative comments, but other experiments can articulate, refine, and explore the phenomenon. For example, *in vitro* studies should be able to distinguish whether preferences change through self-reflection (as suggested by the Maria story and the qualitative data) or whether consumers are simply better able to articulate (or retrieve from memory) preferences after self-reflection. Continuous-time Markov process models might provide a means to study dynamic changes in consumer preferences as consumers search for new products (Hauser and Wisniewski 1982).

## Footnotes

1) An additive HB CBC model is sufficiently general to represent both compensatory decision rules and many non-compensatory decision rules (Bröder 2000; Kohli and Jedidi 2007; Olshavsky and Acito 1980; Yee, et al. 2007).
2) The relative KL measures cannot be easily compared between the automotive and mobile phone studies. Predictive ability (KL) depends upon the number of alternatives and the number of aspects, both of which vary between the automotive and mobile phone studies.

## References

Alba, Joseph W. and J. Wesley Hutchinson. 1987. "Dimensions of consumer expertise." *Journal of Consumer Research* 13 (4): 411-54.

_____. 2000. "Knowledge calibration: what consumers know and what they think they know." *Journal of Consumer Research* 27 (2): 123-56.

Alter, Adam L. and Daniel M. Oppenheimer. 2008. "Effects of fluency on psychological distance and mental construal (or why New York is a large city, but New York is a civilized jungle)." *Psychological Science* 19 (2): 161-67.

Betsch, Tilmann, Suzanne Haberstroh, Andreas Glöckner, Thomas Haar, and Klaus Fiedler. 2001. "The effects of routine strength on adaptation and information search in recurrent decision making." *Organizational Behavior and Human Decision Processes* 84: 23-53.

Bettman, James R., Mary Frances Luce and John W. Payne. 1998. "Constructive consumer choice processes." *Journal of Consumer Research* 25 (3): 187-217.

_____. 2008. "Preference construction and preference stability: putting the pillow to rest." *Journal of Consumer Psychology* 18: 170-74.

Bröder, Arndt 2000. "Assessing the empirical validity of the 'take the best' heuristic as a model of human probabilistic inference." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26 (5): 1332-46.

Brucks, Merrie 1985. "The effects of product class knowledge on information search behavior." *Journal of Consumer Research* 12 (1): 1-16.

Callahan, John and Eytan Lasry. 2004. "The importance of customer input in the development of very new products." *R&D Management* 34 (2): 107-120.

Chaloner, Kathryn and Isabella Verdinelli. 1995. "Bayesian experimental design: a review." *Statistical Science* 10 (3): 273-304.

Cooper, Robert G. and Elko J. Kleinschmidt 1987. "New products: what separates winners from losers?" *Journal of Product Innovation Management* 4 (3): 169-184.

Dahan, Ely and John R. Hauser 2002. "The virtual customer." *Journal of Product Innovation Management* 19 (5): 332-53.

Dieckmann, Anja, Katrin Dippold and Holger Dietrich. 2009. "Compensatory versus noncompensatory models for predicting consumer preferences." *Judgment and Decision Making* 4 (3): 200-13.

Ding, Min 2007. "An incentive-aligned mechanism for conjoint analysis." *Journal of Marketing Research* 44 (2): 214-23.

Ding, Min, John R. Hauser, Songting Dong, Daria Dzyabura, Zhilin Yang, Chenting Su, and Steve Gaskin. 2011. "Unstructured direct elicitation of decision Rules." *Journal of Marketing Research* 48 (1): 116-27.

Ding, Min, Yong-Hoon Park, and Eric T. Bradlow. 2009. "Barter markets for conjoint analysis." *Management Science* 55 (6): 1003-17.

Dzyabura, Daria and John R. Hauser. 2011. "Active machine learning for consideration heuristics." *Marketing Science*, 30, (5): 801-19.

Elrod, Terry, Jordan Louviere, and Krishnakumar S. Davey. 1992. "An empirical comparison of ratings-based and choice-based conjoint models." *Journal of Marketing Research* 29 (3): 368-77.

Folkes, Valerie S. 1988. "Recent attribution research in consumer behavior: a review and new directions." *Journal of Consumer Research* 14 (4): 548-65.

Frederick, Shane 2002. "Automated choice heuristics." In *Heuristics and biases: the psychology of intuitive judgment*, ed. T. Gilovich, D. Griffin, and D. Kahneman, 548-58. New York, NY: Cambridge University Press.

_____. 2005. "Cognitive reflection and decision making." *Journal of Economic Perspectives* 19 (4): 25-42.

Garcia-Retamero, Rocio and Jörg Rieskamp. 2009. "Do people treat missing information adaptively when making inferences?" *Quarterly Journal of Experimental Psychology* 62(10): 1991-2013.

Gigerenzer, Gerd and Daniel G. Goldstein. 1996. "Reasoning the fast and frugal way: models of bounded rationality." *Psychological Review* 103 (4): 650-69.

Green, Paul E., Abba M. Krieger, and Pradeep Bansal. 1988. "Completely unacceptable levels in conjoint analysis: a cautionary note." *Journal of Marketing Research* 25 (3): 293-300.

Green, Paul E. and V. Srinivasan. 1990. "Conjoint analysis in marketing: new developments with implications for research and practice." *Journal of Marketing* 54 (4): 3-19.

Greenwald, Anthony G., Anthony R. Pratkanis, Michael R. Leippe, and Michael H. Baumgardner. 1986. "Under what conditions does theory obstruct research progress?" *Psychological Review* 93 (2): 216-29.

Griffin, Abbie and John R. Hauser. 1993. "The voice of the customer." *Marketing Science* 12 (1): 1-27.

Hansen, David E. and James G. Helgeson. 1996. "The effects of statistical training on choice heuristics in choice under uncertainty." *Journal of Behavioral Decision Making* 9: 41-57.

Hauser, John R. 2001. "Metrics thermostat," *Journal of Product Innovation Management*, 18 (3): 134-153.

Hauser, John R., Olivier Toubia, Theodoros Evgeniou, Daria Dzyabura, and Rene Befurt. 2010. "Cognitive simplicity and consideration sets." *Journal of Marketing Research* 47 (3): 485-96.

Hauser, John R. and Birger Wernerfelt. 1990. "An evaluation cost model of consideration sets." *Journal of Consumer Research* 16 (4): 393-408.

Hauser, John R. and Kenneth Wisniewski. 1982. "Dynamic analysis of consumer response to marketing strategies," *Management Science*, 28, (5): 455-486.

Hoeffler, Steve and Dan Ariely. 1999. "Constructing stable preferences: a look into dimensions of experience and their impact on preference stability." *Journal of Consumer Psychology* 8(2): 113-39.

Huber, Joel, Dick R. Wittink, John A. Fiedler, and Richard Miller. 1993. "The effectiveness of alternative preference elicitation procedures in predicting choice." *Journal of Marketing Research* 30 (1): 105-14.

Hughes, Marie Adele and Dennis E. Garrett. 1990. "Intercoder reliability estimation approaches in marketing: a generalizability theory framework for quantitative data." *Journal of Marketing Research* 27 (2): 185-95.

Ingrassia, Paul 2010. *Crash course: the American automobile industry's road from glory to disaster*, New York, NY: Random House.

Johnson, Eric J., Robert J. Meyer, and Sanjoy Ghose. 1989. "When choice models fail: compensatory models in negatively correlated environments." *Journal of Marketing Research* 26 (3): 255-90.

Kohli, Rajeev, and Kamel Jedidi. 2007. "Representation and inference of lexicographic preference models and their variants." *Marketing Science* 26 (3): 380-99.

Kugelberg, Ellen 2004. *Information scoring and conjoint analysis*. Department of Industrial Economics and Management, Royal Institute of Technology, Stockholm, Sweden.

Kullback, Solomon, and Richard A. Leibler. 1951. "On information and sufficiency." *Annals of Mathematical Statistics* 22: 79-86.

Lenk, Peter J., Wayne S. DeSarbo, Paul E. Green, Martin R. Young. 1996. "Hierarchical Bayes conjoint analysis: recovery of partworth heterogeneity from reduced experimental designs." *Marketing Science* 15 (2): 173-91.

Lichtenstein, Sarah and Paul Slovic. 2006. *The construction of preference*. Cambridge, UK: Cambridge University Press.

Martignon, Laura and Ulrich Hoffrage. 2002. "Fast, frugal, and fit: simple heuristics for paired comparisons." *Theory and Decision* 52: 29-71.

Montoya-Weiss, Mitzi, and Roger Calantone. 1994. "Determinants of new product performance: a review and meta–analysis." *Journal of Product Innovation Management*, 11, 397–417.

Moore, William L., Jordan J. Louviere, and Rohit Verma. 1999. "Using conjoint analysis to help design product platforms." *Journal of Product Innovation Management* 16 (1): 27-39.

Morwitz, Vicki G., Eric Johnson, David Schmittlein. 1993. "Does measuring intent change behavior?" *Journal of Consumer Research* 20 (1): 46-61.

Newell, Ben R., Tim Rakow, Nicola J. Weston and David R. Shanks. 2004. "Search strategies in decision-making: the success of success." *Journal of Behavioral Decision Making* 17: 117-37.

Olshavsky, Richard W. and Franklin Acito. 1980. "An information processing probe into conjoint analysis." *Decision Sciences* 11 (3): 451-70.

Oppenheimer, Daniel M. 2008. "The secret life of fluency." *Trends in Cognitive Sciences* 12 (6): 237-41.

Payne, John W. 1976. "Task complexity and contingent processing in decision making: an information search." *Organizational Behavior and Human Performance* 16: 366-87.

Payne, John W., James R. Bettman and Eric J. Johnson. 1992. "Behavioral decision research: a constructive processing perspective." *Annual Review of Psychology* 43: 87-131.

_____. 1993. *The Adaptive Decision Maker*. Cambridge UK: Cambridge University Press.

Perreault, William D., Jr. and Laurence E. Leigh. 1989. "Reliability of nominal data based on qualitative judgments." *Journal of Marketing Research* 26 (2): 135-48.

Prelec, Dražen 2004. "A Bayesian truth serum for subjective data." *Science* 306 (15): 462-66.

Pullman, Madeline. E., William. L. Moore, and Don G. Wardell. 2002. "A comparison of quality

function deployment and conjoint analysis in new product design." *Journal of Product Innovation Management* 19 (5): 354-64.

Punj, Girish and Robert Moore. 2009. "Information search and consideration set formation in a web-based store environment." *Journal of Business Research* 62: 644-50.

Rakow, Tim, Ben R. Newell, Kathyrn Fayers and Mette. Hersby. 2005. "Evaluating three criteria for establishing cue-search hierarchies in inferential judgment." *Journal of Experimental Psychology-Learning Memory and Cognition* 31 (5): 1088-104.

Roberts, John H., and James M. Lattin. 1991. "Development and testing of a model of consideration set composition." *Journal of Marketing Research* 28 (4): 429-40.

Rossi, Peter E., and Greg M. Allenby. 2003. "Bayesian statistics and marketing." *Marketing Science* 22 (3): 304-28.

Russo, J. Edward, Margaret G. Meloy, and Victoria H. Medvec. 1998. "Predecisional distortion of product information." *Journal of Marketing Research*, 35 (4): 438-52.

Russo, J. Edward and Paul J. H. Schoemaker. 1989. *Decision traps: the ten barriers to brilliant decision-making & how to overcome them*, New York, NY: Doubleday.

Sawtooth Software, Inc. 1996. *ACA system: adaptive conjoint analysis manual*, Sequim, WA: Sawtooth Software, Inc.

_____. 2004. *The CBC hierarchical Bayes technical paper*. Sequim, WA: Sawtooth Software, Inc.

Shannon, Claude E. 1948. "A mathematical theory of communication." *Bell System Technical Journal* 27 (July & October): 379–423 and 623–56.

Simonson, Itamar 2008. "Will I like a 'medium' pillow? another look at constructed and inherent preferences." *Journal of Consumer Psychology* 18: 155-69.

Slovic, Paul, Dale Griffin and Amos Tversky. 1990. "Compatibility effects in judgment and choice." In *Insights in decision making: a tribute to Hillel J. Einhorn*, ed. R. M. Hogarth, 5-27. Chicago: University of Chicago Press.

Srinivasan, V. 1988. "A conjunctive-compensatory approach to the self-explication of multiattributed preferences." *Decision Sciences* 19 (2): 295-305.

Srinivasan, V. and Chan Su Park. 1997. "Surprising robustness of the self-explicated approach to customer preference structure measurement." *Journal of Marketing Research* 34 (2): 286-91.

Srinivasan, V. and Gordon A. Wyner. 1988. "Casemap: computer-assisted self-explication of multiattributed preferences." In *Handbook on New Product Development and Testing*, ed. W. Henry, M. Menasco, and K. Takada, 91-112. Lexington, MA: D. C. Heath.

Sternthal, Brian and C. Samuel Craig. 1982. *Consumer behavior: an information processing perspective*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Thorngate, Warren 1980. "Efficient decision heuristics." *Behavioral Science* 25 (3): 219-25.

Toubia, Olivier 2006. "Idea generation, creativity, and incentives." *Marketing Science* 25 (5): 411-25.

Toubia, Olivier, John R. Hauser, and Duncan Simester. 2004. "Polyhedral methods for adaptive choice-based conjoint analysis." *Journal of Marketing Research* 41 (1): 116-31.

Tversky, Amos 1972. "Elimination by aspects: a theory of choice." *Psychological Review* 79 (4): 281-99.

Urban, Glen L. and John R. Hauser. 2004. "'Listening-in' to find and explore new combinations of customer needs." *Journal of Marketing* 68 (2): 72-87.

Urban, Glen L., John R. Hauser, and John H. Roberts. 1990. "Prelaunch forecasting of new automobiles: models and implementation." *Management Science* 36 (4): 401-21.

Yee, Michael, Ely Dahan, John R. Hauser and James Orlin. 2007. "Greedoid-based noncompensatory inference." *Marketing Science* 26 (4): 532-49.

**Table 1. US Automotive Features and Feature Levels**

| Feature | Levels |
|---|---|
| Brand | Audi, BMW, Buick, Cadillac, Chevrolet, Chrysler, Dodge, Ford, Honda, Hyundai, Jeep, Kia, Lexus, Mazda, Mini-Cooper, Nissan, Scion, Subaru, Toyota, Volkswagen |
| Body type | Compact sedan, compact SUV, crossover vehicle, hatchback, mid-size SUV, sports car, standard sedan |
| EPA mileage | 15 mpg, 20 mpg, 25 mpg, 30 mpg, 35 mpg |
| Glass package | None, defogger, sunroof, both |
| Transmission | Standard, automatic, shiftable automatic |
| Trim level | Base, upgrade, premium |
| Quality of workmanship rating | Q3, Q4, Q5 (defined to respondents) |
| Crash test rating | C3, C4, C5 (defined to respondents) |
| Power seat | Yes, no |
| Engine | Hybrid, internal-combustion |
| Price | Varied from $16,000 to $40,000 based on five manipulated levels plus market-based price increments for the feature levels (including brand) |

## Table 2. Analysis Of Variance: Task-Order And Task-Based Predictions

| | | KL Divergence [a] | |
|---|---|---|---|
| **ANOVA** | **df** | **F** | **Significance [b]** |
| **Task order** | 1 | 5.3 | **.02** |
| **Task-based predictions** | 2 | 12.0 | **<.01** |
| **Interaction** | 2 | 3.5 | **.03** |

| **Effect** | **Beta** | ***t*** | **Significance [b]** |
|---|---|---|---|
| **First in order** | -.004 | -0.2 | .81 |
| Not First in order | *na* [c] | *na* | *na* |
| **E-mail task** [d] | .086 | 6.2 | **<.01** |
| Casemap [d] | .018 | 1.3 | .21 |
| Revealed preferences [e] | *na* [c] | *na* | *na* |
| **First-in-order x E-mail task** [f] | -.062 | -2.6 | **.01** |
| First-in-order x Casemap [f] | -.018 | -.8 | .44 |
| First-in order x Revealed preferences [f] | *na* [c] | *na* | *na* |

[a] Relative Kullback-Leibler Divergence, an information-theoretic measure.

[b] Bold font if significant at the 0.05 level or better.

[c] *na* = set to zero for identification. Other effects are relative to this task order, task, or interaction.

[d] Predictions based on stated preferences with estimated compensatory cut-off.

[e] HB CBC estimation based on profile decision task

[f] Second-in-order x Task coefficients are set to zero for identification and, hence, not shown.

## Table 3. Predictive Ability Automotive Choice (One-Week Delay)

| Task-Based Predictions [a] | Task Order [b] | KL Divergence [c] | *t* | Significance |
|---|---|---|---|---|
| **Revealed preferences** (HB CBC based on profile-decision task) | First | .069 | 0.4 | .67 |
| | Not First | .065 | – | – |
| **Casemap** (Structured preference-articulation task) | First | .068 | -0.9 | .40 |
| | Not First | .082 | – | – |
| **E-mail to an agent** (Unstructured preference-articulation task) | **First** | **.093** | **-2.6** | **.01** |
| | **Not First** | **.151** | – | – |

[a] All predictions based on calibration data only. All predictions evaluated on delayed validation task.

[b] Bold font if significant at the 0.05 level or better between first in order vs. not first in order.

[c] Relative Kullback-Leibler Divergence, an information-theoretic measure.

**Table 4. Hong Kong Mobile Phone Features and Levels**

| Feature | Levels |
|---|---|
| Brand | Motorola, Lenovo, Nokia, Sony-Ericsson |
| Color | Black, blue, silver, pink |
| Screen size | Small (1.8 inch), large (3.0 inch) |
| Thickness | Slim (9 mm), normal (17 mm) |
| Camera resolution | 0.5 Mp, 1.0 Mp, 2.0 Mp, 3.0 Mp |
| Style | Bar, flip, slide, rotational |
| Price | Varied from $HK1,000 to $HK2,500 based on four manipulated levels plus market-based price increments for the feature levels (including brand). |

## Table 5. Predictive Ability (Three-Week Delay), Hong Kong Mobile Phones

| Task-Based Predictions [a] | Task Order [b] | KL Divergence [c] | *t* | Significance |
|---|---|---|---|---|
| **Revealed preferences** (HB CBC based on profile-decision task) | First | .251 | 1.0 | .32 |
| | Not first | .225 | – | – |
| **Lexicographic decision rules** (Machine-learning estimation based on profile-decision task) | First | .236 | 0.3 | .76 |
| | Not first | .225 | – | – |
| **Weakly-structured task** (Structured, but less structured than Casemap) | **First** | **.138** | **-3.5** | **< .01** |
| | **Not first** | **.250** | – | – |
| **E-mail to an agent** [d] (Unstructured preference-articulation task) | **Weakly-structured first** [e] | **.203** | **-2.7** | **<.01** |
| | **Profile decisions first** | **.297** | – | – |

[a] All predictions based on calibration data only. All predictions evaluated on delayed validation task.

[b] Bold font if significant at the 0.05 level or better between first in order vs. not first in order.

[c] Relative Kullback-Leibler Divergence, an information-theoretic measure.

[d] For mobile phones the e-mail task occurred after the profile-decision and weakly-structured tasks.

[e] The weakly-structured preference-articulation task was rotated with the profile-decision task.

## Table 6. Predictive Ability Confirmation: Hit Rate

| Task-Based Predictions [a] | Task Order [b, c] | Hit Rate | *t* | Significance |
|---|---|---|---|---|
| **AUTOMOTIVE CHOICE** | | | | |
| **Revealed preferences** (HB CBC based on profile-decision task) | First | .697 | 0.4 | .69 |
| | Not First | .692 | – | – |
| **Casemap** (Structured preference-articulation task) | First | .692 | -1.4 | .16 |
| | Not First | .718 | – | – |
| **E-mail to an agent** (Unstructured preference-articulation task) | ***First*** | ***.679*** | ***-1.8*** | ***.08*** |
| | ***Not First*** | ***.706*** | – | – |
| **FURTHER STUDY (MOBILE PHONES)** | | | | |
| **Revealed preferences** (HB CBC based on profile-decision task) | First | .800 | 0.6 | .55 |
| | Not first | .791 | – | – |
| **Lexicographic decision rules** (Machine-learning estimation based on profile-decision task) | First | .776 | 1.2 | .25 |
| | Not first | .751 | – | – |
| **Weakly-structured task** (Structured, but less structured than Casemap) | **First** | **.703** | **-2.9** | **<.01** |
| | **Not first** | **.768** | – | – |
| **E-mail to an agent** [e] (Unstructured preference-articulation task) | **Weakly-structured first** [e] | **.755** | **-2.3** | **.02** |
| | **Profile decisions first** | **.796** | – | – |

[a] All predictions based on calibration data only. All predictions evaluated on delayed validation task.

[b] Bold font if significant at the 0.05 level or better between first in order vs. not first in order.

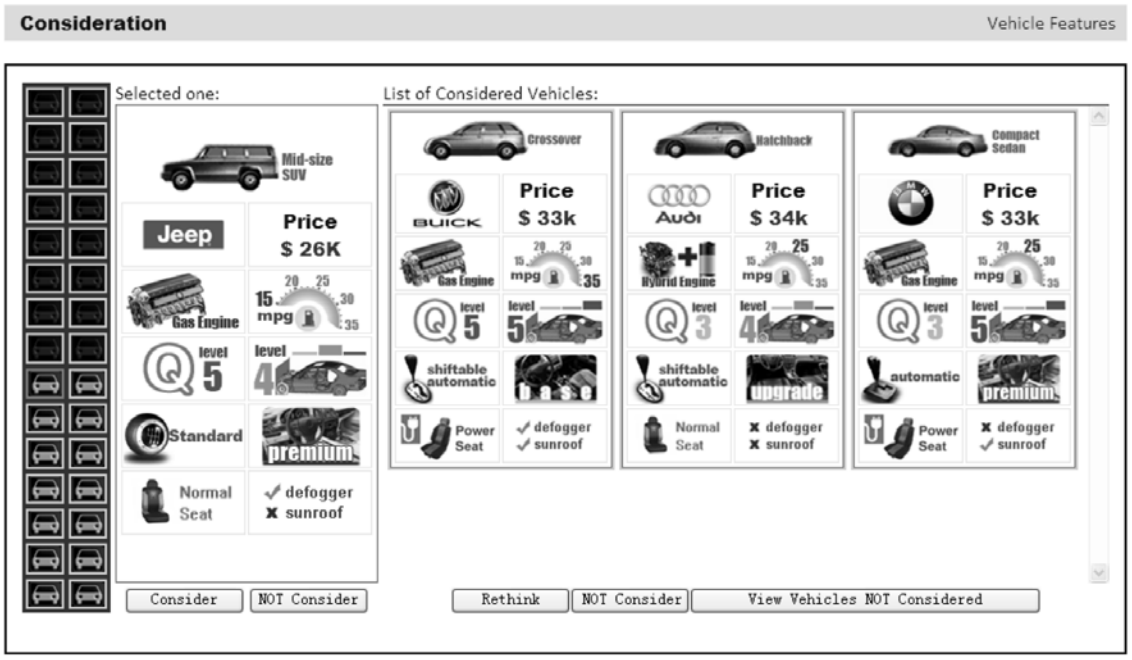
[c] Bold italics font if significantly better at the 0.10 level or better between first in order vs. not first in order.

[d] In the mobile-phone study the weakly-structured task was rotated with the profile-decision task.

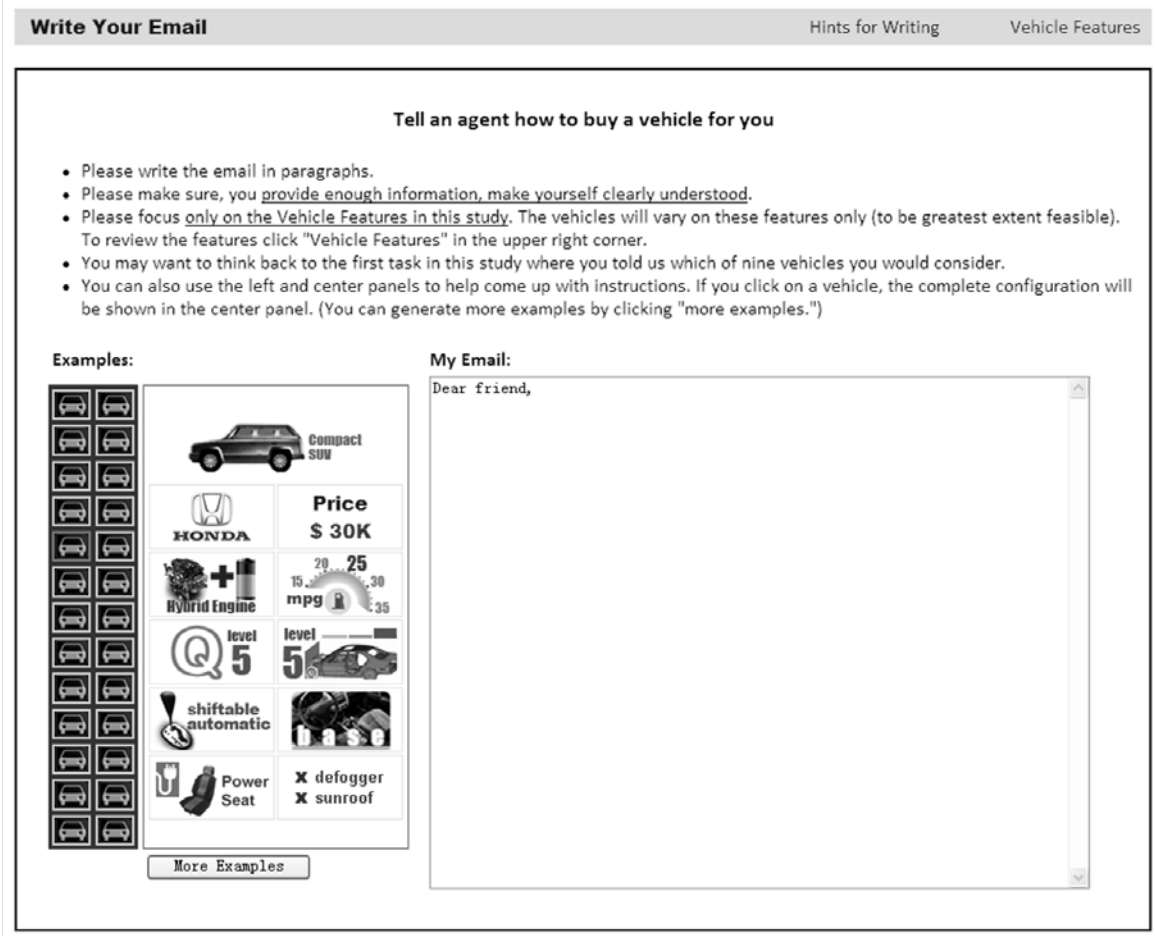
[e] In the mobile-phone study e-mail task occurred after the profile-decision and weakly-structured tasks.

**Figure 1. Example Screenshots from Three Preference Measurement Methods**

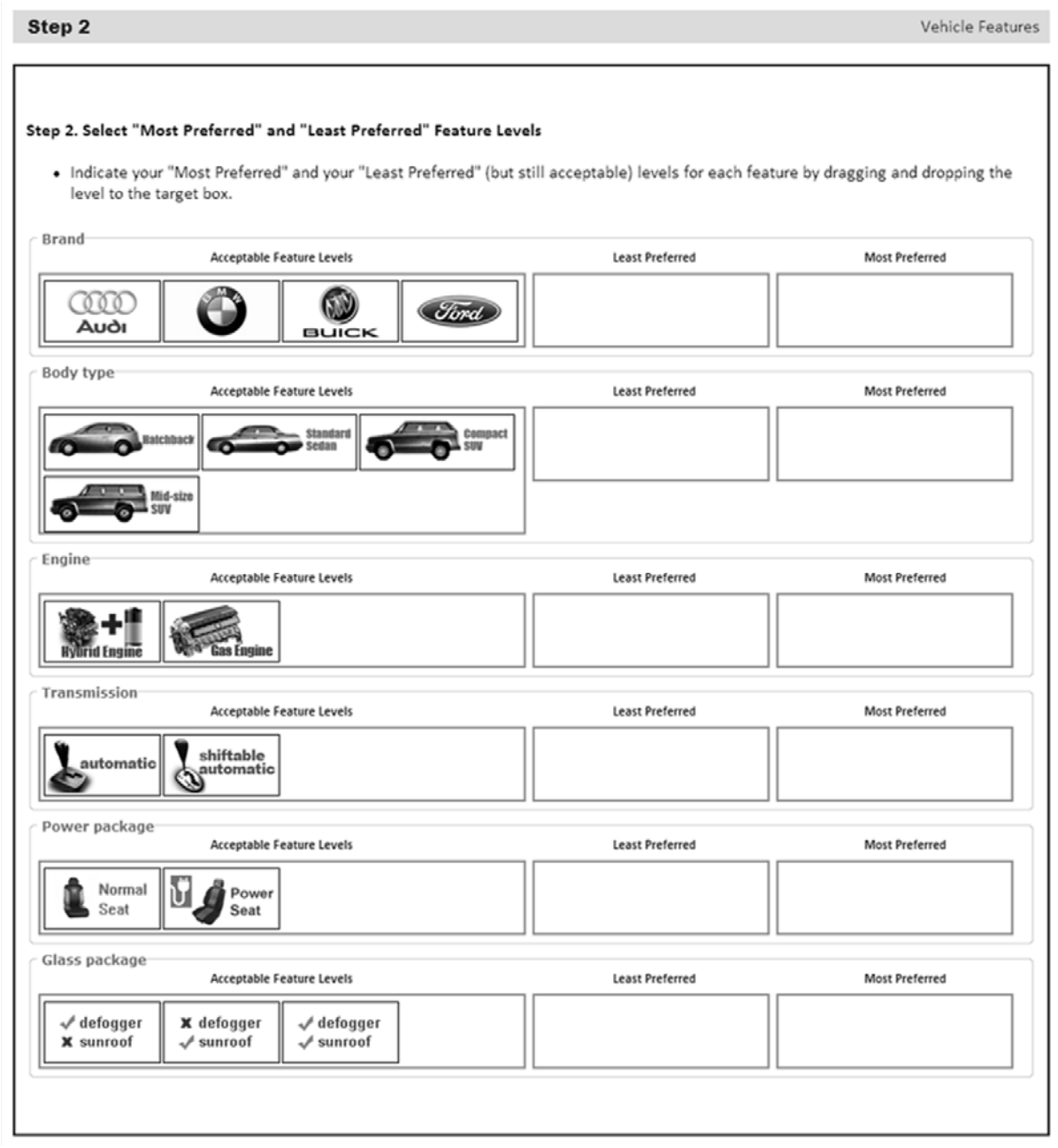
**(Originals in color. More complete screenshots available in an online appendix.)**




(a) Revealed-preference (bullpen) task. Consumers indicate which profiles they will consider.




(b) Unstructured preference-elicitation task. Consumers write an e-mail to an agent.




(c) Structured preference elicitation (Casemap). Step 2 of 5. Other steps include indicating unacceptable features, selecting a critical feature, stating importances for features, and stating preferences for levels of features.