

# **Learning to Construct Decision Rules**

JOHN HAUSER

SONGTING DONG

MIN DING

June 2010

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.

Songting Dong is a Lecturer in Marketing, Research School of Business, the Australian National University, Canberra, ACT 0200, Australia, +61 (2) 6125-7343, dongst@gmail.com.

Min Ding is Smeal Research Fellow and Professor of Marketing, Smeal College of Business, Pennsylvania State University, University Park, PA 16802-3007; phone: (814) 865-0622, mind-ing@psu.edu.

Keywords: *Conjoint analysis, conjunctive rules, consideration sets, constructive decision rules, enduring decision rules, expertise, incentive alignment, introspection, learning, lexicographic rules, non-compensatory decision rules, revealed preference, self-explication.*

## Abstract

Data suggest that, as consumers make sufficiently many decisions, they introspect and form decision rules that differ from rules they describe prior to introspection. Articulated introspection-based rules predict future decisions significantly better and are more stable and less likely to be influenced by context. The observed effects derive from two studies in which respondents made consideration-set decisions, answered structured questions about decision rules, and wrote an e-mail to a friend who would act as their agent. The tasks were rotated. Weeks later respondents made consideration-set decisions but from a different set of product profiles. Product profiles were complex (53 automotive aspects in the first study) and chosen to represent the marketplace. Decisions were incentive-aligned and the decision format allowed consumers to revise their consideration set as they evaluated profiles. Delayed validation, qualitative comments, and response times provide insight on how decision-rule introspection affects and is affected by consumers' stated decision rules.

*One of the main ideas that has emerged from behavioral decision research (is that) preferences and beliefs are actually constructed—not merely revealed—in the elicitation process.*

– Paul Slovic, Dale Griffin, and Amos Tversky (1990)

*Experts process information more deeply ...*

– Joseph W. Alba and J. Wesley Hutchinson (1987)

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 (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 conjunctive criteria. She learned that most new (2010) sporty, stylish coupes had a chop-top style with poor visibility and even worse trunk space. With her growing consumer expertise, she changed her conjunctive 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 decision rule were triggered by the choice-set context, but were the result of extensive introspection. 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 consideration rule 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 consideration rule predicted she would consider an Audi A5 and a certified-used BMW 335i Coupe. But our 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. This paper provides insight

on whether we should expect Maria to construct a new consideration rule based on the new choice-set context or to retain her introspected consideration rule in the new context.

Maria's (true) story illustrates major tenets of modern consumer decision theory. (1) Maria simplified her decision process with a two-stage consider-then-choose process (e.g., Beach and Potter 1992; Hauser and Wernerfelt 1990; Payne 1976; Payne, et. al. 1993; Roberts and Latin 1991; Montgomery and Svenson 1976; Punj and Moore 2009; Simonson, Nowlis, and Lemon 1993). Conjunctive consideration rule(s) were used to screen alternatives rather than make the final choice. (2) Her consideration rules were constructed based on the available choice alternatives even though she began with an *a priori* consideration rule (e.g., Bettman, Luce, and Payne 1998, 2008; Lichtenstein and Slovic 2006; Morton and Fasolo 2009; Payne, Bettman, and Johnson 1988, 1992, 1993; Payne, Bettman, and Schkade 1999; Rabin 1998; Slovic, Griffin, and Tversky 1990.) (3) Although Maria was not under time pressure either initially or when her Probe expired, the large number of alternatives (200+ make-model combinations) and the large number of automotive features caused her to use conjunctions in her decision heuristic (*supra* citations plus Bröder 2000; Gigerenzer and Goldstein 1996; Klein and Yadav 1989; Martignon and Hoffrage 2002; Thorngate 1980). (4) As Maria became a more-expert automotive consumer she modified her heuristic consideration rules (Alba and Hutchinson 1987; Betsch, et. al. 2001; Bröder and Schiffer 2006; Brucks 1985; Hensen and Helgeson 1996; Newell, et. al. 2004).

These major tenets are flexible, but as typically interpreted, they predict differently the consideration rule that Maria used three months after her initial search when her mechanic told her she must replace her Probe. Constructed-decision-rule theories, if applied to consideration decisions, suggest Maria would reconstruct her decision rules three months hence, possibly affected by anchoring, compromise effects, asymmetric dominance, and other context effects (Hu-

ber, Payne, and Puto 1982; Huber and Puto 1983; Kahneman and Tversky 1979; Simonson and Tversky 1992; Thaler 1985). As Payne, Bettman, and Schkade (1999, 245) summarize: “(a) major tenet of the constructive perspective ... (is that) preferences are generally constructed at the time the valuation question is asked.”

In contrast the learning and expertise literatures suggest that greater training (expertise) leads to decisions that are more likely to endure (Betsch, et al. 1999; 2001; Bröder and Newell 2008; Bröder and Schiffer 2006; Garcia-Retamero and Rieskamp 2009; Hensen and Helgeson 1996, 2001; Newell, Weston, and Shanks 2004). Even the classic Bettman and Zins (1977, 79) research on constructed decision rules found a roughly equal distribution of constructed rules, implementation of rules from memory, and preprocessed choices. Betsch, et al. (1999, 152) suggest that “a growing body of research provides ample evidence that future choices are strongly predicted by past behavior (behavioral routines).” And Hensen and Helgeson (1996, 42) opine that: “Training ... may result in different choice strategies from those of the untrained.” The expertise and learning theories would likely predict that Maria’s more-expert introspected decision rule endured. They predict Maria would be less likely to reconstruct her decision rule and that the effect of context would diminish.

In this paper we interpret evidence that informs the tension between typical interpretations of construction and expertise. Evidence suggests that, when faced with an important and potentially novel decision, consumers like Maria learn their consideration rules through introspection. Self-reported consideration rules predict later decisions much better when they are elicited after substantial incentive-aligned consideration tasks than when they are elicited before such tasks. (Respondents had a reasonable chance of getting a car (plus cash) worth \$40,000, where the specific car was based on the respondent’s stated consideration rules.) Evidence from

a second study (mobile phones) reinforces the introspection-learning hypothesis and suggests that measurement context effects during the learning phase endure. In both studies, qualitative comments, response times, and changes in decision rules are consistent with an introspection-learning hypothesis.

Our data suggest further that the task training, common in behavioral experiments, is not sufficient to induce introspection learning. It is likely that many published constructed-decision-rule experiments are in domains where respondents were not given sufficient opportunity for introspection learning.

### **OUR APPROACH**

Our paper has the flavor of an effects paper (Deighton, et al. 2010). The data were collected to compare new methods to measure and elicit consideration rules (Anonymous 2010). As suggested by Payne, Bettman, and Johnson (1993, 252) these data were focused on consideration-set decisions—a domain where decision heuristics are likely. Serendipitously, the data inform issues of decision rule construction and introspection learning (this paper).

The data were expensive to obtain because the measurement approximated *in vivo* decision making. For example, in the first study all measures were incentive-aligned, 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. (An aspect is a feature-level as defined by Tversky 1972.) 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.

The 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).

Because we are interpreting effects to form hypotheses we forego the standard hypothesis-then-test format. Instead we describe first the two studies and then interpret new analyses as evidence suggesting introspection learning. Ancillary data (qualitative comments, response times, decision rule evolution) provide further insight. We then re-interpret evidence from various literatures in light of introspection learning and, from these literatures, we expand the introspection learning hypothesis. We leave further exploration to future *in vitro* experiments.

For clarity we state the basic hypothesis: Introspection learning implies that, if consumers are given a sufficiently realistic task and enough time to complete that task, consumers will introspect their decision rules (and potential consumption of the products). Introspection helps consumers clarify and articulate their decision rules. The clarified decision rules endure and are more stable and less likely to be influenced by context.

## **TASK ORDER SUGGESTS CONSUMERS LEARN CONSIDERATION RULES**

### Overview of the Study Design (First Study, Automobiles)

Student respondents are relative novices with respect to automotive purchases, but interested in the category and likely to make a purchase upon graduation. In a set of rotated online tasks, respondents 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 elicitation 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. Respondents completed the training task in less than 1/10<sup>th</sup> the amount of time observed for any of the three primary tasks. One week later respondents 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 respondents and, by the end of the pretests, respondents found the survey easy to understand and representative of their decision processes. The 204 respondents agreed. On five-point scales, the tasks were easy to understand (2.01, SD = .91, where 1 = “extremely easy”) and easy for respondents’ to understand that it was in their best interests to tell us their true preference (1.81, SE = .85, where 1 = “extremely easy”). Respondents 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) respondents 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 (Ding 2007; Ding, Grewal and Liechty 2005; Ding, Park and Bradlow 2009; Kugelberg 2004; Park, Ding and Rao 2008; Prelec 2004; Toubia, Hauser and Garcia 2007; Toubia, et al. 2003).

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 respondents to trust that it was in their best in-

terests to tell the truth. For example, if respondents 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 which is revealed after the study is completed. With a secret set, if respondents' decision rules screened out too few vehicles, they had a good chance of getting an automobile they did not like. On the other hand, if their decision rules screened out too many vehicles none would have remained in the consideration set and their decision rules would not have affected the vehicle they received. We chose the size of the secret set, 20 vehicles, 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 respondents received a participation fee of \$15 when they completed both the initial three tasks and the delayed validation task. In addition, one randomly-drawn respondent 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 respondent'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, respondents 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 we purchased prize indemnity insurance where, for a fixed fee, the insurance company would pay \$40,000 if a respondent won an automobile. One random respondent got the chance to choose two of 20 envelopes. Two of the envelopes contained a winning card; the other 18 contained a losing card. If both envelopes had contained a winning

card, the respondent would have received the \$40,000 prize. Such drawings are common for radio and automotive promotions. Experience suggests that respondents 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 respondent's first card was a winner, but, alas, the second card was not.

### Consideration-Set-Decision, Structured-Elicitation, and Unstructured-Elicitation Tasks

*Consideration-Set Decision Task.* We designed the consideration-set-decision task 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 respondents 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 respondent moused over an icon, all features were displayed in a middle area using text and pictures. The respondent could consider, not consider, or replace the profile. All considered profiles were displayed in an area on the right and the respondent 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. Respondents took the task seriously investing, on average, 7.7 minutes to evaluate 30 profiles. This is substantially more time than the 0.7 minutes respondents spent on the sequential 9-profile training task, even accounting for the larger number of profiles ( $p < .01$ ).

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  $20 \times 7 \times 5^2 \times 4 \times 3^4 \times 2^2$  feature-level design is large by industry standards, but remained understandable to respondents. To make profiles realistic and to avoid dominated profiles (e.g., Elrod, Louviere and Davey 1992; Green, Helsen, and Shandler 1988; Johnson, Meyer and Ghose 1989; Hauser, et. al. 2010; Toubia, et al. 2003; Toubia, et al. 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., we used a price increment for a BMW relative to a Scion). We removed unrealistic profiles 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-Elicitation Task.* The structured task collects data on both compensatory and conjunctive decision rules. We chose an established method, Casemap, that is used widely and that we have used in previous studies. Following Srinivasan (1988), closed-ended questions ask respondents 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 respondents are over-zealous in indicating unacceptable levels (Akaah and Korgaonkar 1983; Bateson, Reibstein and Boulding 1987; Green 1984; Green and Helsen 1989; Green, Krieger and Banal 1988; Huber, et al. 1993; Klein 1986; Leigh, MacKay and Summers 1984, Moore and Semenik 1988; Sawtooth 1996; Srinivasan and Park 1997; Srinivasan and Wyner 1988). This was the case with our respondents. Predicted consideration sets were

much smaller with Casemap than were observed or predicted with data from the other tasks.

*Unstructured-Elicitation Task.* Respondents were asked to write an e-mail to a friend who would act as their agent should they win the 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 we told them 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 respondent’s e-mail—no other personal communication.) To encourage trust, agents were audited (Toubia 2006). Following standard procedures the data were coded 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; Wright 1973). Example responses and coding rules are available from the authors and published in Anonymous (2010). 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 consideration-set-decision task. Respondents saw a new draw of 30 profiles from the orthogonal design (different draw for each respondent). For the structured-elicitation task we used standard Casemap analyses: first eliminating unacceptable profiles and then using the compensatory portion with a cutoff as described below. For the unstructured-elicitation task we used the stated conjunctive rules to eliminate or to include profiles and then the coded compensatory rules. To predict consideration with compensatory rules we needed to establish a cutoff. We established the cutoff 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.

As a benchmark, we used the data from the calibration consideration-set-decision task to predict consideration sets in the validation data. We make predictions using standard hierarchical-Bayes, choice-based-conjoint-like, additive logit analysis estimated with data from the calibration consideration-set-decision task (HB CBC, Hauser, et al. 2010; Lenk, et al. 1996; Rossi and Allenby 2003, Sawtooth 2004; Swait and Erdem 2007). An additive HB CBC model is sufficiently general to represent both compensatory decision rules and many non-compensatory decision rules (Bröder 2000; Jedidi and Kohli 2005; Kohli and Jedidi 2007; Olshavsky and Acito 1980; Yee, et al. 2007). In the first study we could not use non-compensatory revealed-preference methods because such methods are not computationally feasible for a 53-aspect design (Boros, et al. 1997; Gilbride and Allenby 2004, 2006; Hauser, et al. 2010; Jedidi and Kohli 2005; Yee, et al. 2007). Non-compensatory revealed-preference methods are feasible for the  $4^5 \times 2^2$  feature-set in the second study.

To measure predictive ability we report 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 Verdinielli 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.) Anonymous (2010) provide formulae for consideration-set decisions. See also Hauser, et al. (2010) for a related measure.

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 respondent 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 interested in hit rates, we repeat key analyses at the end of the paper. The implications are the same.)

### **COMPARISON OF PREDICTIVE ABILITY BASED ON TASK ORDER**

If introspection learning takes place we expect predictive accuracy to depend upon the order(s) in which respondents completed the three tasks. For example, consumers should be able to articulate decision rules better after completing an intensive task that causes them to think deeply about their decision rules. (From the literature, qualitative data, and many years studying automotive decisions, we believe consumers form these decision rules as they introspect the decision process, but the predictive tests cannot rule out an hypothesis that consumers retrieve decision rules from heretofore obscured memory.)

#### Analysis of Predictive Ability versus Task, Task Order, and Interactions (First Study)

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 summarizes the analysis of variance. All predictions are based on the calibration data. All validations are based on 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 Anonymous (2010). The first-vs.-not-first effect and its interaction with task is interesting theoretically relative to the introspection-learning hypothesis. Table 3 provides a simpler summary that isolates the effect as

driven by the task order of unstructured-elicitation (the e-mail task).

---

Insert tables 2 and 3 about here.

---

Respondents are better able to describe decision rules that predict consideration sets one week later if they first complete either a 30-profile consideration-set decision or a structured set of queries about their decision rules ( $p < .01$ ). The information in the predictions one-week later is over 60% larger if respondents have the opportunity for such introspection learning via either an intensive decision task (30 profiles) or an intensive structured-elicitation task (KL = .151 vs. .093). There are hints of introspection 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-decision-rule predictions (based on HB CBC) do not benefit if respondents are first asked to self-state decision rules with either the e-mail or Casemap tasks. This is consistent with an hypothesis that the bull-pen format allows respondents to introspect and revisit consideration decisions and thus learn their decision rules during the consideration-set-decision task.

#### Commentary on the Task Order Effect for the E-mail Task

We believe that the results in table 3 indicate more than just a methodological finding.

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

- respondents are better able to articulate stable decision rules to an agent after introspection. (Introspection came either from substantial consideration-set decisions or the highly-structured Casemap task.)
- the articulated decision rules endure (at minimum, their predictive ability endures for at least one week)
- 9-profile, sequential, profile-evaluation training does not provide sufficient introspection



learning (because substantially more learning was observed in tables 2 and 3).

- Pretests indicated respondents felt the 9-profile warm-up was sufficient to introduce them to the features, the feature levels, and the task. This suggests that introspection learning is more than simply learning the composition of the market.

We examined attribution and self-perception theories as alternative explanations to introspection 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 they do not explain the order effect or the task  $\times$  order interaction. In our data all respondents completed all three tasks prior to validation. We can also rule out recency hypotheses because there is no effect due to second versus third in the task order.

Our respondents were mostly novices with respect to automobile purchases. We believe most respondents did not have well-formed decision rules prior to our study but, like Maria, they learned their decision rules through introspection as they made consideration decisions (or completed the highly-structured Casemap task) that determined the automobile they would receive (if they won the lottery). If this hypothesis is correct, then it has profound implications for research on constructed decision rules. Most experiments in the literature do not begin with substantial introspective training and, likely, are in a domain when respondents are still on the steep part of the learning curve.

If most constructed-decision experiments are in the domain where respondents are still learning their decision rules, then there are at least two possibilities with very different implications. (1) Context affects decision rules during learning and the context effect endures even if consumers later have time for extensive introspection. If this is the case, then it is important to understand the interaction of context and learning. Alternatively, (2) context might affect deci-

sion rules during learning, but the context effect might be overwhelmed *in vivo* when consumers have sufficient time for introspection. If this is the case, then there are research opportunities to define the boundaries of context-effect phenomena.

We next examine ancillary data from the automotive study to explore the reasonableness of the introspection-learning hypothesis. We then describe a second study which reinforces the introspection-learning hypothesis and addresses issues of whether context-like effects endure.

### **QUALITATIVE SUPPORT, RESPONSE TIMES, AND DECISION-RULE EVOLUTION**

The first study was a methodological study designed to examine the predictive ability of unstructured elicitation. Qualitative data, response times, and decision-rule evolution were obtained, but not focused systematically on introspection learning. Nonetheless, we can examine unsolicited comments for further insight.

#### Qualitative Comments in the First Study

Respondents 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 introspection learning. These comments suggest that the decision tasks helped respondents to think deeply about imminent car purchases. (We corrected minor spelling and grammar in the quotes.)

- *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 purchase 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 respondents 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 the First Study

In any set of tasks we expect respondents 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 introspection learning, respondents can articulate their decision rules more rapidly (e-mail task) if

either the consideration-set-decision task or the structured-elicitation task comes first. However, we cannot rule out general learning because response times for all tasks improve 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 respondents complete all three tasks, they have learned their decision rules. For those respondents for whom the consideration-set-decision task was not first, the response times for the initial consideration-set-decision task and the delayed validation task are not statistically different ( $p = .82$ ). This is consistent with an hypothesis that respondents learned their decision rules 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.

#### Decision Rule Evolution in First Study

Although the total number of stated decision rules does change as the result of task order, the rules change in feature-level focus. When respondents have the opportunity to introspect prior to the unstructured elicitation (e-mail) task, consumers are more discriminating—more features include both must-have and must-not-have conjunctive rules. On average, we see a change toward more rules addressing EPA mileage (more must-not-have low mileage) and quality ratings (more must-have Q5) and fewer rules addressing crash test ratings (fewer must-not-have C3) and transmission (fewer must have automatic). More respondents want midsized SUVs.

In summary, ancillary analyses of qualitative comments, response times, and decision rules are consistent with the introspection-learning hypothesis—an hypothesis that we believe is a parsimonious explanation.

## **SECOND STUDY: INTROSPECTION LEARNING AND TASK CONTEXT EFFECTS**

We now examine data from a second study. Like the first study, these data were also collected to examine alternative measurement methods, but re-analyses provide insight on introspection learning effects. 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  $4^5 \times 2^2$  design). The smaller number of aspects makes it feasible to use non-compensatory revealed-decision-rule methods allowing us to explore whether the revealed-decision-rule results are unique to HB CBC. These data have the advantage that the delayed validation task occurs three weeks after the initial tasks. We can examine whether the introspection-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 second study also enables us to explore the effect of context during introspection learning. Specifically, the e-mail task always comes after the other two tasks—only the 32-profile consideration-set-decision task (bullpen) and a structured-elicitation task were rotated. This rotation protocol provides insight on whether measurement context during learning affects respondents' articulation of decision rules and their match to subsequent behavior. In addition, the structured elicitation task was less structured than Casemap and more likely to interfere with unstructured elicitation (e-mail task). In the second study the structured task asked respondents to state rules on a preformatted screen, but, unlike Casemap, respondents were not forced to provide rules for every feature and feature level and the rules were not constrained to the Casemap conjunctive/compensatory structure.

Other aspects of the second study were similar to the first study. (1) The consideration-set-decision task used the bullpen format that allowed respondents to introspect consideration-set decisions prior to finalizing their decisions. (2) All decisions were incentive-aligned. One in 30 respondents 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 carefully to be realistic and representative of the marketplace. Table 4 summarizes the mobile-phone features.

---

Insert table 4 about here.

---

The 143 respondents 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 respondents to assure that the questions were clear and the tasks not onerous, respondents 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 respondents received \$HK100 when they completed both the initial tasks and the delayed-validation task.

#### Analysis of Task, Task Order, and Interactions (Second Study)

Table 5 summarizes predictive ability for the second study (ANOVA results were similar to those in the first study). We used two revealed-decision-rule methods estimated from the calibration consideration-set-decision task. “Revealed decision rules (HB CBC)” is as in the first

(automotive) study. “Lexicographic decision rules” estimates a conjunctive model of consideration-set 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. See Anonymous 2010.) We also estimated disjunctions-of-conjunctions rules using logical analysis of data (LAD) as described in Hauser, et al. (2010). The LAD results are basically the same as for the lexicographic rules and provide no additional insight on the task-order introspection-learning hypothesis. Details are available from the authors.

---

Insert table 5 about here.

---

We first compare the two rotated tasks: revealed-decision-rule predictions versus structured-elicitation predictions. The analysis reproduces and extends the first-study analysis. Structured elicitation predicts significantly better if it is preceded by a consideration-set-decision task (bullpen) that is sufficiently intense to allow introspection. The structured elicitation task provides over 80% more information ( $KL = .250$  vs.  $.138$ ) when respondents complete the elicitation task after the bullpen task rather than before the bullpen task. This is consistent with self-learning through introspection. Unlike in the previous study (automotive), the introspection learning effect (mobile phones) is significant for the structured task. (Introspection learning improved Casemap predictions, but not significantly in the first study.) This increased effect is likely due to the reduction in structure of the structured task. The less-structured structured task’s sensitivity to introspection learning (for mobile phones) is more like that of the e-mail task than the Casemap task (for automobiles). As in the first study, there is no order effect for revealed-

decision-rule models, neither for HB CBC nor greedoid-estimated lexicographic decision rules.

The revealed-decision-rule models provide more information for mobile-phone consideration than for automotive consideration, likely because there are more observations per parameter for the mobile-phone aspects. The set of aspects for mobile phones is complex, but in the “sweet spot” of HB CBC. Additive decision rules (HB CBC), which can represent both compensatory and many non-compensatory decision rules, do as well as lexicographic decision rules. This is consistent with prior research (Dieckmann, et al. 2009; Kohli and Jedidi 2007; Yee, et al. 2007), and with a count of the stated decision rules: 78.3% of the respondents asked their agents to use a mix of compensatory and non-compensatory decision rules.

#### A Context Effect (Interference from the Structured Task)

The task-order effect for the end-of-survey e-mail task is a new effect not tested in the first study. The predictive ability of the e-mail task is best if respondents evaluate 32 profiles using the bullpen format before they complete a structured elicitation task. We observe this effect even though, by the time they begin the e-mail task, *all* respondents have evaluated 32 profiles using the bullpen. The order effect for structured elicitation (stated-rule format in Table 5) is consistent with introspection learning: respondents who evaluate 32 profiles before structured elicitation are better able to articulate decision rules. However, initial measurement interference from the structured task (stated-rule format) endures to the unstructured elicitation task (e-mail-to-an-agent format in Table 5) even though the unstructured e-mail task follows memory-cleansing and other tasks.

Table 5 suggests that the structured task interferes with (suppresses) respondents’ abilities to state unstructured rules even though respondents had the opportunity for introspection be-



tween the structured and unstructured tasks. (That is, the e-mail task does better if in the order {introspection, stated-rule, e-mail} than in the order {stated-rule, introspection, e-mail}.) Respondents appear to anchor to the decision rules they state in the structured task and this anchor interferes with respondents' abilities to introspect, learn, and then articulate their decision rules.

However, not all predictive ability is lost. Even with interference, the decision rules stated in the post-introspection e-mail task predict reasonably well. They predict significantly better than rules from the pre-introspection structured task ( $p < .01$ ) and comparably to rules revealed by the bullpen task ( $p = .15$ ).

On the other hand, the structured task does not affect the predictive ability of decision rules revealed by the bullpen task. In the bullpen task respondents have sufficient time and motivation for introspection. We summarize the relative predictive ability as follows where  $>$  implies statistical significance and  $\sim$  implies no statistical difference:

$$\text{Stated rules (email) after introspection no interference} > \left\{ \begin{array}{l} \text{Stated rules (email) after interference} \\ \text{and introspection} \end{array} \sim \left\{ \begin{array}{l} \text{Revealed rules} \\ \text{allowing for} \\ \text{introspection} \end{array} \right\} > \left\{ \begin{array}{l} \text{Stated rules} \\ \text{(structured)} \\ \text{before introspection} \end{array} \right.$$

From the literature we know that context effects (compromise, asymmetric dominance, object anchoring, etc.) affect decision rules. It is reasonable to hypothesize that context also affects stated decision rules, interferes with introspection learning, and suppresses the ability of respondents to describe their decision rules to an agent. If this hypothesis is true, context will have less of an effect on decision rules if consumers first complete a task that allows them to introspect.

We now return to Maria. Before her extensive search she stated consideration-set decision rules. Had she been forced to form a consideration set from which to purchase an automobile, her rules would have led her to a Hyundai Genesis Coupe, a Nissan Altima Coupe, and a

certified-used Infiniti G37 Coupe. Her search process revealed information and encouraged introspection causing her to develop new decision rules. The new consideration-set rules implied an Audi A5 and a certified-used BMW 335i Coupe. Three months later she used the introspected rules to choose an Audi A5 from the set of available vehicles (a sporty coupe without a sunroof, not black, white or silver, stylish, well-handling, moderate fuel economy, and good truck room).

#### Qualitative Support, Response Times, And Decision-Rule Evolution for the Second Study

The qualitative data collected for the second study (mobile-phones) were not as extensive as the qualitative data collected in the first study (automotive). However, qualitative comments suggest introspection 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 (consideration-set-decision) task.*
- *I really can know more about my preference of choosing the ideal mobile phone.*

As in the first study, respondents completed the bullpen task faster when it came after the structured-elicitation task ( $p = .02$ ). However, unlike the first study, task order did not affect response times for the structured-elicitation 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 respondents.)

Task order affected the types of decision rules for mobile phones. If the bullpen task preceded structured elicitation, respondents stated more rules in the e-mail task (more must-have and more compensatory rules). For example, more respondents required Motorola, Nokia, or So-

ny-Ericsson, more required a black phone, and more required a flip or rotational style. (Also fewer rejected a rotational style.)

### 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 structured task ( $p < .01$ ), and for the mobile-phone e-mail task ( $p = .02$ ). See table 6.

### Summary

Introspection learning is consistent with our data and, we believe, parsimonious.

## SUPPLEMENTAL INTERPRETATIONS AND RELATED LITERATURES

### Introspection Learning

Based on one of the author's thirty years of experience studying automotive consumers, Maria's story rings true. The qualitative quotes and the quantitative data are consistent with an introspection-learning hypothesis which we restate here:

**Introspection Learning Hypothesis.** *If consumers are given a sufficiently realistic task and enough time to complete that task, consumers will introspect their decision rules (and potential consumption of the products). Introspection helps consumers clarify and articulate their decision rules. The clarified decision rules endure and are more stable and less likely to be influenced by context.*

The automotive and the mobile-phone studies randomize rather than manipulate systematically the context of the consideration-set decision. However, given that task-order effects endure to validation one or three weeks later, it appears that either introspection learning endures

and leads to stable decision rules or that respondents introspect similar decision rules when allowed sufficient time.

We hypothesize that introspected decision rules would be much less susceptible to context-based construction than decision rules observed prior to introspection. Many of the results in the literature would diminish sharply. For example we hypothesize that time pressure would influence rules less after respondents had a chance to form decision rules through introspection. We believe that diminished decision-rule construction is important to theory and practice because introspection learning is likely to represent marketplace consider-then-choose behavior in many categories. Recognizing that the two studies were not designed specifically to test introspection learning, we supplement our analyses by examining related research in the learning, expertise, and constructed-decision-rule literatures.

#### Evidence from Learning, Expertise, and Constructed-Decision-Rule Literatures

In the decision-rule learning literature, Betsch, et al. (2001) manipulated task learning through repetition and found that respondents were more likely to maintain a decision routine with 30 repetitions rather than with 15 repetitions. (They used a one-week delay between induction and validation.) Hensen and Helgeson (1996) demonstrated that learning cues influence naïve decision makers to behave more like experienced decision makers. Hensen and Helgeson (1996, 2001) found that experienced decision makers use different decision rules. Garcia-Retamero and Rieskamp (2009) describe experiments where respondents shift from compensatory to conjunctive-like decision rules over seven trial blocks. Newell, et al. (2004, 132) summarize evidence that learning reduces decision-rule incompatibility. Incompatibility was reduced from 60-70% to 15% when respondents had 256 learning trials rather than 64 learning trials.

Even in children, more repetition results in better decision rules (John and Whitney 1986). As Bröder and Newell (2008, 212) summarize: “the decision how to decide ... is the most demanding task in a new environment.”

Hoeffler and Ariely (1999) show that effort and experience improve the stability of compensatory tradeoffs (e.g., sounds that vary on three features). For sounds they find no effect due to the number of choices made, but found that hard choices reduced violations. Making a choice increased respondents confidence in stated preference tradeoffs—an effect consistent with introspection learning.

In the expertise literature, Alba and Hutchison (1987) hypothesize that experts use different rules than novices and Alba and Hutchison (2000) hypothesize that confident experts are more accurate. For complex usage situations Brucks (1985) presents evidence that consumers with greater objective or subjective knowledge (expertise) spend less time searching inappropriate features, but examine more features. We believe that introspection learning is a means to earn expertise.

In the constructed-decision-rule literature Bettman and Park (1980) find significant effects on between-brand processing due to prior knowledge. Bettman and Zins (1977) suggest that about a third of the choices they observed were based on rules stored in memory and Bettman and Park (1980) find significant effects due to prior knowledge. Simonson (2008) argues that some preferences are learned through experience, that such preferences endure, and that “once uncovered, inherent (previously dormant) preferences become active and retrievable from memory (Simonson, 162).” Bettman, et al. (2008) counter that such enduring preferences may be influenced by constructed decision rules. We hypothesize that the Simonson-vs.-Bettman-et-al. debate depends upon the extent to which respondents have time for (prior) introspection learning.

Does Context Endure? Does the Observed Interference Generalize to other Context Effects?

The interference observed in the second study (mobile phones) has precedent in the literature. Russo and Schoemaker (1989) provide many examples where decision makers do not adjust sufficiently from initial conditions and Rabin (1998) reviews behavioral literature to suggest that learning does not eliminate decision biases. Russo, Meloy, and Medvec (1998) suggest that pre-decision information is distorted to support brands that emerge as leaders in a decision process. Bröder and Schiffer (2006) report experiments where respondents maintain their decision strategies in a stock market game even though payoff structures change and decision strategies are no longer optimal. Rakow, et al. (2005) give examples where, with substantial training, respondents do not change their decision rules under higher search costs or time pressure. Alba and Hutchison (1987, 423) hypothesize that experts are less susceptible to time pressure and information complexity.

There is some evidence that the introspection-learning might endure in seminal experiments in the constructed-decision-rule literature. For example, on the first day of their experiments, Bettman, et al. (1988, Table 4) find that time pressure caused the pattern of decision rules to change ( $-.11$  vs.  $-.17$ ,  $p < .05$ , where more negative means more attribute-based processing). However, Bettman, et al. did not observe a significant effect on the second day after respondents had made decisions under less time pressure on the first day ( $.19$  vs.  $.20$ ,  $p$  not available). Indeed the second-day-time-pressure patterns were more consistent with no-time-pressure patterns—consistent with an hypothesis that respondents formed their decision rules on the first day and continued to use them on the second day.

In a related set of experiments, Amir and Levav (2008) examine context effects during a learning task (e.g., six successive choices of laptop computers that varied on two features). Con-

text during learning makes respondents more likely to be influenced by context in choices that were made after a filler task. (Respondents were not allowed to revisit choice decisions—a protocol that may have inhibited introspection learning.)

It should not surprise us when initial conditions endure. Although introspection learning appears to improve respondents' abilities to articulate decision rules, our analyses suggest that decision-context manipulations are likely to be important during the learning phase and that they endure at least to the end-of-survey e-mail task. Simply asking structured questions to elicit decision rules for mobile phones diminished the ability of respondents to later articulate unstructured decision rules even after introspection. We expect that other phenomena will affect decision rule selection during decision-rule learning and that this effect will influence how consumers learn through introspection.

## **METHODOLOGICAL ISSUES**

### **Training**

Researchers are careful to train subjects in the experiment's decision task, but that training may not be sufficient for introspection learning. For example, Luce, Payne, and Bettman's (1999) experiments in automotive choice used two warm-up questions, but, our data suggest that 9 profiles were not sufficient for introspection learning. Similarly, in deliberation-without-attention experiments (Dijksterhuis, et al. 2006; Lassiter, et al. 2009), respondents read information about products and were asked for preferences with minimal opportunity for introspection. Decision tasks that encourage introspection might resolve the controversy. To the extent that most researchers provide substantially less training than the bullpen task, context-dependent constructed-decision-rule experiments might be focused on the domain where respondents are still

learning their decision rules. It would be interesting to see which experimental context effects diminish after a 30-profile (automotive) or 32-profile (mobile phone) bullpen task.

### Unstructured Elicitation

Think-aloud protocols, process tracing, and many related direct-observation methods are common in the consumer-behavior literature. Our data suggest that unstructured elicitation is a good means to identify decision rules that predict future behavior, especially if respondents are first given the opportunity for introspection learning.

### Conjoint Analysis and Contingent Valuation

In both tables 3 and 5 the best predictions are obtained from unstructured elicitation (after a bullpen task). This methodological result is not a panacea because the combined tasks take longer than revealed methods (HB CBC, greedoid methods, and LAD can be applied to the bullpen data alone) and because unstructured elicitation requires qualitative coding. Contingent valuation in economics is controversial (Diamond and Hausman 1994; Hanemann 1994; Hausman 1993), but tables 3 and 5 suggest that it can be made more accurate if respondents are first given a task that allows for introspection learning.

## SUMMARY

Our data were collected to explore new methods to measure preference. Features, choice alternatives, and formats were chosen to approximate *in vivo* decisions. To test the methods fairly, they were rotated randomly. Methodological issues are published elsewhere, but through a new lens, the two studies highlight interesting task-order effects to provide serendipitously a



fresh perspective on issues in the constructed-decision-rule, learning, and expertise literatures.

The introspection-learning hypothesis begins with established hypotheses from the constructed-decision-rule literature: Novice consumers, novice to either the product category or the task, construct their decision rules as they evaluate products. During the construction process consumers are influenced context. The introspection-learning hypothesis adds that, if consumers make sufficiently many decisions, they will introspect and form enduring decision rules that may be different from those they would state or use prior to introspection. The hypothesis states further that subsequent decisions will be based on the enduring rules and that the enduring decision rules are less likely to be influenced by context. (It remains an open question how long the rules endure.) This may seem like a subtle change, but the implications are many:

- constructed decision rules are less likely if consumers have a sufficient opportunity for prior introspection
- constructed decision rules are less likely for more-expert consumers
- many results in the literature are based on domains where consumers are still learning,
- but, context effects during learning may endure
- context effects are less likely to be observed *in vivo*.

The last statement is based in our belief that many important *in vivo* consumer decisions, such as automotive consideration decisions, are made after introspection. The introspection-learning hypothesis and most of its implications can also be explored *in vitro*. As this is not our primary skill, we leave *in vitro* experiments to future research.

## REFERENCES

**(Reduced font and spacing in references to save paper.)**

- Akaah, Ishmael P. and Pradeep K. Korgaonkar (1983), "An Empirical Comparison of the Predictive Validity of Self-explicated, Huber-hybrid, Traditional Conjoint, and Hybrid Conjoint Models," *Journal of Marketing Research*, 20, (May), 187-197.
- Alba, Joseph W. and J. Wesley Hutchinson (1987), "Dimensions of Consumer Expertise," *Journal of Consumer Research*, 13, (March), 411-454.
- \_\_\_\_\_ (2000), "Knowledge Calibration: What Consumers Know and What They Think They Know," *Journal of Consumer Research*, 27, (September), 123-156.
- 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, 161-167.
- Amir, On and Jonathan Levav (2008), "Choice Construction versus Preference Construction: The Instability of Preferences Learned in Context," *Journal of Marketing Research*, 45, (April), 145-158.
- Anonymous (2010), "Unstructured Direct Elicitation of Decision Rules," *Journal of Marketing Research*, forthcoming.
- Bateson, John E. G., David Reibstein, and William Boulding (1987), "Conjoint Analysis Reliability and Validity: A Framework for Future Research," *Rev. of Marketing*, Michael Houston, Ed., 451-481.
- Beach, Lee Roy and Richard E. Potter (1992), "The Pre-choice Screening of Options," *Acta Psychologica*, 81, 115-126.
- Betsch, Tilmann, Babette Julia Brinkmann, Klaus Fiedler and Katja Breining (1999), "When Prior Knowledge Overrides New Evidence: Adaptive Use Of Decision Strategies And The Role Of Behavioral Routines," *Swiss Journal of Psychology*, 58, 3, 151-160.
- Betsch, Tillman, Susanne 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, (December), 187-217.
- \_\_\_\_\_ (2008), "Preference Construction and Preference Stability: Putting the Pillow to Rest," *Journal of Consumer Psychology*, 18, 170-174.
- Bettman, James R. and C. Whan Park (1980), "Effects of Prior Knowledge and Experience and Phase of the Choice Process on Consumer Decision Processes: A Protocol Analysis," *Journal of Consumer Research*, 7, 234-248.
- Bettman, James R. and Michel A. Zins (1977), "Constructive Processes in Consumer Choice," *Journal of*

- Consumer Research*, 4, (September), 75-85.
- Boros, Endre, Peter L. Hammer, Toshihide Ibaraki, and Alexander Kogan (1997), "Logical Analysis of Numerical Data," *Mathematical Programming*, 79, (August), 163-190.
- 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, 1332-1346.
- Bröder, Arndt and Ben R. Newell (2008), "Challenging Some Common Beliefs: Empirical Work Within the Adaptive Toolbox Metaphor," *Judgment and Decision Making*, 3, 3, (March), 205-214.
- Bröder, Arndt and Stefanie Schiffer (2006), "Adaptive Flexibility and Maladaptive Routines in Selecting Fast and Frugal Decision Strategies," *Journal of Experimental Psychology: Learning, Memory and Cognition*, 34, 4, 908-915.
- Brucks, Merrie (1985), "The Effects of Product Class Knowledge on Information Search Behavior," *Journal of Consumer Research*, 12, 1, (June), 1-16.
- Chaloner, Kathryn and Isabella Verdinelli (1995), "Bayesian Experimental Design: A Review," *Statistical Science*, 10, 3, 273-304.
- Deighton, John, Debbie MacInnis, Ann McGill and Baba Shiv (2010), "Broadening the Scope of Consumer Research," Editorial by the editors of the *Journal of Consumer Research*, February 16.
- Diamond, Peter A. and Jerry A. Hausman (1994), "Contingent Valuation: Is Some Number Better Than No Number? (Contingent Valuation Symposium)," *Journal of Economic Perspectives* 8, 4, 45-64
- Dieckmann, Anja, Katrin Dippold and Holger Dietrich (2009), "Compensatory versus Noncompensatory Models for Predicting Consumer Preferences," *Judgment and Decision Making*, 4, 3, (April), 200-213.
- Dijksterhuis, Ap, Maarten W. Bos, Loran F. Nordgren, and Rick B. van Baaren (2006), "On Making the Right Choice: The Deliberation-Without-Attention Effect," *Science*, 311, (17 Feb), 1005-1007.
- Ding, Min (2007), "An Incentive-Aligned Mechanism for Conjoint Analysis," *Journal of Marketing Research*, 54, (May), 214-223.
- Ding, Min, Rajdeep Grewal, and John Liechty (2005), "Incentive-Aligned Conjoint Analysis," *Journal of Marketing Research*, 42, (February), 67-82.
- Ding, Min, Young-Hoon Park, and Eric T. Bradlow (2009) "Barter Markets for Conjoint Analysis" *Management Science*, 55 (6), 1003-1017.
- Elrod, Terry, Jordan Louviere, and Krishnakumar S. Davey (1992), "An Empirical Comparison of Ratings-Based and Choice-based Conjoint Models," *J. of Marketing Research* 29, 3, (Aug), 368-377.
- Folkes, Valerie S. (1988), "Recent Attribution Research in Consumer Behavior: A Review and New Directions," *Journal of Consumer Research*, 14, 4, (March), 548-565.

- Frederick, Shane (2002), "Automated Choice Heuristics," in Thomas Gilovich, Dale Griffin, and Daniel Kahneman, eds., *Heuristics and Biases: The Psychology of Intuitive Judgment*, New York, NY: Cambridge University Press, 548-558.
- \_\_\_\_\_ (2005), "Cognitive Reflection and Decision Making." *J. 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), 650-669.
- Gilbride, Timothy J. and Greg M. Allenby (2004), "A Choice Model with Conjunctive, Disjunctive, and Compensatory Screening Rules," *Marketing Science*, 23 (3), 391-406.
- \_\_\_\_\_ (2006), "Estimating Heterogeneous EBA and Economic Screening Rule Choice Models," *Marketing Science*, 25, 5, (September-October), 494-509.
- Green, Paul E., (1984), "Hybrid Models for Conjoint Analysis: An Expository Review," *Journal of Marketing Research*, 21, 2, (May), 155-169.
- Green, Paul E. and Kristiaan Helsen (1989), "Cross-Validation Assessment of Alternatives to Individual-Level Conjoint Analysis: A Case Study," *Journal of Marketing Research*, 26, (August), 346-350.
- Green, Paul E., Kristiaan Helsen, and Bruce Shandler (1988), "Conjoint Internal Validity Under Alternative Profile Presentations," *Journal of Consumer Research*, 15, (December), 392-397.
- Green, Paul E., Abba M. Krieger, and Pradeep Bansal (1988), "Completely Unacceptable Levels in Conjoint Analysis: A Cautionary Note," *Journal of Marketing Research*, 25, (August), 293-300.
- 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, 216-229.
- Griffin, Abbie and John R. Hauser (1993), "The Voice of the Customer," *Marketing Science*, 12, 1, (Winter), 1-27.
- Hanemann, W. Michael (1994), "Valuing the environment through contingent valuation. (Contingent Valuation Symposium)," *Journal of Economic Perspectives* 8, 4, (Fall), 19-43.
- Hauser, John R., Olivier Toubia, Theodoros Evgeniou, Daria Dzyabura, and Rene Befurt (2010), "Cognitive Simplicity and Consideration Sets," *Journal of Marketing Research*, 47, (June), 485-496.
- Hauser, John R. and Birger Wernerfelt (1990), "An Evaluation Cost Model of Consideration Sets," *Journal of Consumer Research*, 16 (March), 393-408.
- Hausman, Jerry A., Ed. (1993), *Contingent Valuation: A Critical Assessment*, Amsterdam, The Netherlands: North Holland Press.
- Hensen, David E. and James G. Helgeson (1996), "The Effects of Statistical Training on Choice Heuris-

- tics in Choice under Uncertainty,” *Journal of Behavioral Decision Making*, 9, 41-57.
- \_\_\_\_\_ (2001), “Consumer Response to Decision Conflict from Negatively Correlated Attributes: Down the Primrose Path of Up Against the Wall?,” *Journal of Consumer Psychology*, 10, 3, 159-169.
- Hoeffler, Steve and Dan Ariely (1999), “Constructing Stable Preferences: A Look into Dimensions of Experience and their Impact on Preference Stability,” *J. of Consumer Psychology*, 8, 2, 113-139.
- Huber, Joel, John W. Payne, and Christopher Puto (1982), "Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis," *Journal of Consumer Research*, 9 (June), 90-8.
- Huber Joel and Christopher Puto (1983), "Market Boundaries and Product Choice: Illustrating Attraction and Substitution Effects," *Journal of Consumer Research*, 10 (June), 31-44.
- 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, (February), 105-114.
- 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, (May), 185-195.
- Jedidi, Kamel and Rajeev Kohli (2005), “Probabilistic Subset-Conjunctive Models for Heterogeneous Consumers,” *Journal of Marketing Research*, 42 (4), 483-494.
- John, Deborah Roedder and John C. Whitney, Jr. (1986), “The Development of Consumer Knowledge in Children: A Cognitive Structure Approach,” *Journal of Consumer Research*, 12, (Mar) 406-417.
- Johnson, Eric J., Robert J. Meyer, and Sanjoy Ghose (1989), “When Choice Models Fail: Compensatory Models in Negatively Correlated Environments,” *J. of Marketing Research*, 26, (Aug), 255-290.
- Kahneman, Daniel and Amos Tversky (1979), “Prospect Theory: An Analysis of Decision under Risk,” *Econometrica*, 47, 2, (March), 263-291.
- Klein, Noreen M. (1986), “Assessing Unacceptable Attribute Levels in Conjoint Analysis,” *Advances in Consumer Research*, 14, 154-158.
- Klein, Noreen M. and Manjit S. Yadav (1989), “Context Effects on Effort and Accuracy in Choice: An Enquiry into Adaptive Decision Making,” *Journal of Consumer Research*, 15, (March), 411-421.
- Kohli, Rajeev, and Kamel Jedidi (2007), “Representation and Inference of Lexicographic Preference Models and Their Variants,” *Marketing Science*, 26 (3), 380-399.
- 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.

- Lassiter, G. Daniel, Matthew J. Lindberg, Claudia Gonzáles-Vellejo, Francis S. Bellezza, and Nathaniel D. Phillips (2009), "The Deliberation-Without-Attention Effect: Evidence for an Artifactual Interpretation," *Psychological Science*, 20, 6, 671-675.
- Leigh, Thomas W., David B. MacKay, and John O. Summers (1984), "Reliability and Validity of Conjoint Analysis and Self-Explicated Weights: A Comparison," *Journal of Marketing Research*, 21, 4, (November), 456-462.
- 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), 173-191.
- Lichtenstein, Sarah and Paul Slovic (2006), *The Construction of Preference*, Cambridge, UK: Cambridge University Press.
- Luce, Mary Frances, John W. Payne, and James R. Bettman (1999), "Emotional Trade-off Difficulty and Choice," *Journal of Marketing Research*, 36, 143-159.
- Martignon, Laura and Ulrich Hoffrage (2002), "Fast, Frugal, and Fit: Simple Heuristics for Paired Comparisons," *Theory and Decision*, 52, 29-71.
- Montgomery, Henry and Ola Svenson (1976), "On Decision Rules and Information Processing Strategies for Choices among Multiattribute Alternatives," *Scandinavian Jour. of Psychology*, 17, 283-291.
- Moore, William L. and Richard J. Semenik (1988), "Measuring Preferences with Hybrid Conjoint Analysis: The Impact of a Different Number of Attributes in the Master Design," *Journal of Business Research*, 16, 3, (May), 261-274.
- Morton Alec and Barbara Fasolo (2009), "Behavioural Decision Theory for Multi-criteria Decision Analysis: A Guided Tour," *Journal of the Operational Research Society*, 60, 268-275.
- Morwitz, Vicki G., Eric Johnson, David Schmittlein (1993), "Does Measuring Intent Change Behavior?" *Journal of Consumer Research*, 20, 1, (June) 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, 117-137.
- Olshavsky, Richard W. and Franklin Acito (1980), "An Information Processing Probe into Conjoint Analysis," *Decision Sciences*, 11, (July), 451-470.
- Oppenheimer, Daniel M. (2008), "The Secret Life of Fluency," *Trends in Cognitive Sciences*, 12, 6, (June), 237-241.
- Park, Young-Hoon, Min Ding, Vithala R. Rao (2008) "Eliciting Preference for Complex Products: Web-Based Upgrading Method," *Journal of Marketing Research*, 45 (5), p. 562-574
- Payne, John W. (1976), "Task Complexity and Contingent Processing in Decision Making: An Information Search," *Organizational Behavior and Human Performance*, 16, 366-387.

- Payne, John W., James R. Bettman and Eric J. Johnson (1988), "Adaptive Strategy Selection in Decision Making," *J. of Experimental Psychology: Learning, Memory and Cognition*, 14, 3, 534-552.
- \_\_\_\_\_ (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.
- Payne, John W., James R. Bettman and David A. Schkade (1999), "Measuring Constructed Preferences: Toward a Building Code," *Journal of Risk and Uncertainty*, 19, 1-3, 243-270.
- Perreault, William D., Jr. and Laurence E. Leigh (1989), "Reliability of Nominal Data Based on Qualitative Judgments," *Journal of Marketing Research*, 26, (May), 135-148.
- Platt, John R. (1964), "Strong Inference," *Science*, 146, (October 16), 347-53.
- Prelec, Dražen (2004), "A Bayesian Truth Serum for Subjective Data," *Science*, 306, (Oct 15), 462-466.
- Punj, Girish and Robert Moore (2009), "Information Search and Consideration Set Formation in a Web-based Store Environment," *Journal of Business Research*, 62, 644-650.
- Rabin, Matthew (1998), "Psychology and Economics," *Journal of Economic Literature*, 36, 11-46.
- Rakow, Tim, Ben R. Newell, Kathryn 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, 1088-1104.
- Roberts, John H., and James M. Lattin (1991), "Development and Testing of a Model of Consideration Set Composition," *Journal of Marketing Research*, 28 (November), 429-440.
- Rossi, Peter E., Greg M. Allenby (2003), "Bayesian Statistics and Marketing," *Marketing Science*, 22, 3, 304-328.
- 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," *ACA 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 623-656.
- Simonson, Itamar (2008), "Will I like a "Medium" Pillow? Another Look at Constructed and Inherent Preferences," *Journal of Consumer Psychology*, 18, 155-169.
- Simonson, Itamar, Stephen Nowlis, and Katherine Lemon (1993), "The Effect of Local Consideration Sets on Global Choice Between Lower Price and Higher Quality," *Marketing Sci.*, 12, 357-377.
- Simonson, Itamar and Amos Tversky (1992), "Choice in Context: Tradeoff Contrast and Extremeness Aversion," *Journal of Marketing Research*, 29, 281-295.

- Slovic, Paul, Dale Griffin and Amos Tversky (1990), "Compatibility Effects in Judgment and Choice," in Robin M. Hogarth (Ed.), *Insights in Decision Making: A Tribute to Hillel J. Einhorn*, Chicago: University of Chicago Press, 5-27.
- Srinivasan, V. (1988), "A Conjunctive-Compensatory Approach to The Self-Explication of Multiattributed Preferences," *Decision Sciences*, 19, (Spring), 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, (May), 286-291.
- Srinivasan, V. and Gordon A. Wyner (1988), "Casemap: Computer-Assisted Self-Explication of Multiattributed Preferences," in W. Henry, M. Menasco, and K. Takada, Eds, *Handbook on New Product Development and Testing*, Lexington, MA: D. C. Heath, 91-112.
- Sternthal, Brian and C. Samuel Craig (1982), *Consumer Behavior: An Information Processing Perspective*, Englewood Cliffs, NJ: Prentice-Hall, Inc.
- Swait, Joffre and Tülin Erdem (2007), "Brand Effects on Choice and Choice Set Formation Under Uncertainty," *Marketing Science* 26, 5, (September-October), 679-697.
- Thaler, Richard (1985), "Mental Accounting and Consumer Choice," *Marketing Science*, 4, 3, (Summer), 199-214.
- Thorngate, Warren (1980), "Efficient Decision Heuristics," *Behavioral Science*, 25, (May), 219-225.
- Toubia, Olivier (2006), "Idea Generation, Creativity, and Incentives," *Marketing Science*, 25, 5, (September-October), 411-425.
- Toubia, Olivier, Duncan I. Simester, John R. Hauser, and Ely Dahan (2003), "Fast Polyhedral Adaptive Conjoint Estimation," *Marketing Science*, 22 (3), 273-303.
- Toubia, Olivier, John R. Hauser and Rosanna Garcia (2007), "Probabilistic Polyhedral Methods for Adaptive Choice-Based Conjoint Analysis: Theory and Application," *Marketing Science*, 26, 5, 596-610.
- Toubia, Olivier, John R. Hauser, and Duncan Simester (2004), "Polyhedral Methods for Adaptive Choice-based Conjoint Analysis," *Journal of Marketing Research*, 41, 1, (February), 116-131.
- Tversky, Amos (1972), "Elimination by Aspects: a Theory of Choice," *Psych. Review*, 79,4, 281-299.
- Urban, Glen L. and John R. Hauser (2004), "'Listening-In' to Find and Explore New Combinations of Customer Needs," *Journal of Marketing*, 68, (April), 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, (April), 401-421.
- Wright, Peter (1973), "The Cognitive Processes Mediating Acceptance of Advertising," *Journal of Marketing Research*, 10, (February), 53-62.
- Yee, Michael, Ely Dahan, John R. Hauser and James Orlin (2007) "Greedoid-Based Noncompensatory Inference," *Marketing Science*, 26, 4, (July-August), 532-549.



**TABLE 1****US AUTOMOTIVE FEATURES AND FEATURE LEVELS IN FIRST STUDY**

---

| 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, shifttable 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 VARIATION: TASK-ORDER AND TASK-BASED PREDICTIONS

| ANOVA   | KL Divergence <sup>a</sup> |           |                           |
|---|----------------------------|-----------|---------------------------|
|   | df                         | F         | Significance <sup>b</sup> |
| <b>Task order</b>                                     | 1                          | 5.3       | <b>.02</b>                |
| <b>Task-based predictions</b>                         | 2                          | 12.0      | <b>&lt;.01</b>            |
| <b>Interaction</b>                                    | 2                          | 3.5       | <b>.03</b>                |
| Effect  | Beta                       | t         | Significance <sup>b</sup> |
| <b>First in order</b>                                 | -.004                      | -0.2      | .81                       |
| Not First in order                                    | <i>na</i> <sup>c</sup>     | <i>na</i> | <i>na</i>                 |
| <b>E-mail task <sup>d</sup></b>                       | .086                       | 6.2       | <b>&lt;.01</b>            |
| Casemap <sup>d</sup>                                  | .018                       | 1.3       | .21                       |
| Revealed decision rules <sup>e</sup>                  | <i>na</i> <sup>c</sup>     | <i>na</i> | <i>na</i>                 |
| <b>First-in-order x E-mail task <sup>f</sup></b>      | -.062                      | -2.6      | <b>.01</b>                |
| First-in-order x Casemap <sup>f</sup>                 | -.018                      | -.8       | .44                       |
| First-in order x Revealed decision rules <sup>f</sup> | <i>na</i> <sup>c</sup>     | <i>na</i> | <i>na</i>                 |

<sup>a</sup> Relative Kullback-Leibler Divergence, an information-theoretic measure.

<sup>b</sup> Bold font if significant at the 0.05 level or better.

<sup>c</sup> *na* = set to zero for identification. Other effects are relative to this task order, task, or interaction.

<sup>d</sup> Predictions based on stated decision rules with estimated compensatory cut-off.

<sup>e</sup> HB CBC estimation based on consideration-set decision task

<sup>f</sup> Second-in-order x Task coefficients are set to zero for identification and, hence, not shown.

**TABLE 3****PREDICTIVE ABILITY (ONE-WEEK DELAY), FIRST STUDY (AUTOMOTIVE)**

| <b>Task-Based Predictions<sup>a</sup></b>   | <b>Task Order<sup>b</sup></b> | <b>KL Divergence<sup>c</sup></b> | <b>t</b>    | <b>Significance</b> |
|---|-------------------------------|----------------------------------|-------------|---------------------|
| <b>Revealed decision rules</b><br>(HB CBC based on consideration-set decision task) | First                         | .069                             | 0.4         | .67                 |
|   | Not First                     | .065                             | –           | –                   |
| <b>Casemap</b><br>(Structured elicitation task)                                     | First                         | .068                             | -0.9        | .40                 |
|   | Not First                     | .082                             | –           | –                   |
| <b>E-mail to an agent</b><br>(Unstructured elicitation task)                        | <b>First</b>                  | <b>.093</b>                      | <b>-2.6</b> | <b>.01</b>          |
|   | <b>Not First</b>              | <b>.151</b>                      | –           | –                   |

<sup>a</sup> All predictions based on calibration data only. All predictions evaluated on delayed validation task.

<sup>b</sup> Bold font if significant at the 0.05 level or better between first in order vs. not first in order.

<sup>c</sup> Relative Kullback-Leibler Divergence, an information-theoretic measure.

**TABLE 4****HONG KONG MOBILE PHONE FEATURES AND FEATURE 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), SECOND STUDY (MOBILE PHONES)

| Task-Based Predictions <sup>a</sup>   | Task Order <sup>b</sup>       | KL Divergence <sup>c</sup> | t           | Significance    |
|---|-------------------------------|----------------------------|-------------|-----------------|
| <b>Revealed decision rules</b><br>(HB CBC based on consideration-set decision task)                           | First                         | .251                       | 1.0         | .32             |
|   | Not first                     | .225                       | –           | –               |
| <b>Lexicographic decision rules</b><br>(Machine-learning estimation based on consideration-set decision task) | First                         | .236                       | 0.3         | .76             |
|   | Not first                     | .225                       | –           | –               |
| <b>Stated-rule format</b><br>(Structured elicitation task)  | <b>First</b>                  | <b>.138</b>                | <b>-3.5</b> | <b>&lt; .01</b> |
|   | <b>Not first</b>              | <b>.250</b>                | –           | –               |
| <b>E-mail to an agent</b> <sup>d</sup><br>(Unstructured elicitation task)                                     | <b>SET first</b> <sup>e</sup> | <b>.203</b>                | <b>-2.7</b> | <b>&lt;.01</b>  |
|   | <b>Decision first</b>         | <b>.297</b>                | –           | –               |

<sup>a</sup> All predictions based on calibration data only. All predictions evaluated on delayed validation task.

<sup>b</sup> Bold font if significant at the 0.05 level or better between first in order vs. not first in order.

<sup>c</sup> Relative Kullback-Leibler Divergence, an information-theoretic measure.

<sup>d</sup> For mobile phones e-mail task occurred after the consideration-set-decision and structured-elicitation tasks.

<sup>e</sup> The structured-elicitation task (SET) was rotated with the consideration-set decision task.

**TABLE 6**  
**PREDICTIVE ABILITY CONFIRMATION: HIT RATE**

| <b>FIRST STUDY (AUTOMOTIVE)</b>   |                                  |                 |             |                     |
|---|----------------------------------|-----------------|-------------|---------------------|
| <b>Task-Based Predictions<sup>a</sup></b>   | <b>Task Order<sup>b, c</sup></b> | <b>Hit Rate</b> | <b>t</b>    | <b>Significance</b> |
| <b>Revealed decision rules</b><br>(HB CBC based on consideration-set decision task)                           | First                            | .697            | 0.4         | .69                 |
|   | Not First                        | .692            | –           | –                   |
| <b>Casemap</b><br>(Structured elicitation task)   | First                            | .692            | -1.4        | .16                 |
|   | Not First                        | .718            | –           | –                   |
| <b>E-mail to an agent</b><br>(Unstructured elicitation task)  | <b><i>First</i></b>              | <b>.679</b>     | <b>-1.8</b> | <b>.08</b>          |
|   | <b><i>Not First</i></b>          | <b>.706</b>     | –           | –                   |
| <b>SECOND STUDY (MOBILE PHONES)</b>   |                                  |                 |             |                     |
| <b>Revealed decision rules</b><br>(HB CBC based on consideration-set decision task)                           | First                            | .800            | 0.6         | .55                 |
|   | Not first                        | .791            | –           | –                   |
| <b>Lexicographic decision rules</b><br>(Machine-learning estimation based on consideration-set decision task) | First                            | .776            | 1.2         | .25                 |
|   | Not first                        | .751            | –           | –                   |
| <b>Stated-rule format<sup>d</sup></b><br>(Structured elicitation task)  | <b>First</b>                     | <b>.703</b>     | <b>-2.9</b> | <b>&lt;.01</b>      |
|   | <b>Not first</b>                 | <b>.768</b>     | –           | –                   |
| <b>E-mail to an agent<sup>e</sup></b><br>(Unstructured elicitation task)                                      | <b>SET first<sup>d</sup></b>     | <b>.755</b>     | <b>-2.3</b> | <b>.02</b>          |
|   | <b>Decision first</b>            | <b>.796</b>     | –           | –                   |

<sup>a</sup> All predictions based on calibration data only. All predictions evaluated on delayed validation task.

<sup>b</sup> Bold font if significant at the 0.05 level or better between first in order vs. not first in order.

<sup>c</sup> Bold italics font if significantly better at the 0.10 level or better between first in order vs. not first in order.

<sup>d</sup> In the second study the structured-elicitation task (SET) was rotated with the consideration-set decision task.

<sup>e</sup> In the second study e-mail task occurred after the consideration-set-decision and structured-elicitation tasks.