



Marketing Science

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

Measuring Consumer Preferences Using Conjoint Poker

Olivier Toubia, Martijn G. de Jong, Daniel Stieger, Johann Füller,

To cite this article:

Olivier Toubia, Martijn G. de Jong, Daniel Stieger, Johann Füller, (2012) Measuring Consumer Preferences Using Conjoint Poker. Marketing Science 31(1):138-156. <https://doi.org/10.1287/mksc.1110.0672>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2012, INFORMS

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

Measuring Consumer Preferences Using Conjoint Poker

Olivier Toubia

Columbia Business School, New York, New York 10027, ot2107@columbia.edu

Martijn G. de Jong

Erasmus School of Economics, Erasmus University, 3000 DR Rotterdam, The Netherlands,
mgdejong@ese.eur.nl

Daniel Stieger, Johann Füller

Department of Strategic Management, Marketing and Tourism, University of Innsbruck,
A-6020 Innsbruck, Austria {daniel.stieger@atv.at, johann.fueller@hyve.de}

We develop and test an incentive-compatible Conjoint Poker (CP) game. The preference data collected in the context of this game are comparable to incentive-compatible choice-based conjoint (CBC) analysis data. We develop a statistical efficiency measure and an algorithm to construct efficient CP designs. We compare incentive-compatible CP to incentive-compatible CBC in a series of three experiments (one online study and two eye-tracking studies). Our results suggest that CP induces respondents to consider more of the profile-related information presented to them compared with CBC.

Key words: conjoint analysis; product; measurement and inference; experimental economics

History: Received: August 6, 2009; accepted: July 8, 2011; Eric Bradlow and then Preyas Desai served as the editor-in-chief and Robert Meyer served as associate editor for this article. Published online in *Articles in Advance* December 20, 2011.

1. Introduction

Conjoint analysis is one of the most widely used quantitative market research methods (Bradlow 2005, Green and Srinivasan 1990, Wittink and Cattin 1989). By estimating how much a sample of consumers value a specific set of features, this method allows forecasting how these consumers (and, by approximation, the entire market) are likely to react to any hypothetical set of new products. This enables product developers to optimize the design and the pricing of single products or of entire product lines (e.g., Kohli and Sukumar 1990, Mahajan et al. 1982).

However, one key limitation of preference measurement methods such as conjoint analysis is the potential lack of motivation experienced by respondents (e.g., Ding et al. 2005, Liechty et al. 2005, Netzer et al. 2008). There is a growing concern that the amount of effort and attention spent by consumers when filling out preference measurement questionnaires is lower than when making real-life purchasing decisions, and practitioners have called for preference measurement methods that increase respondents' level of involvement. For example, Johnson (2008, p. 4) writes, "Although respondents do seem to use simplification strategies when filling out questionnaires, they probably work harder when making important real-life

choices. So simplification in answering questions is not a good thing. We learn less than we might if we pushed respondents to use deeper processing. We need to find ways to do that."

One of the significant recent contributions to the preference measurement literature has been the introduction of *incentive-compatible* mechanisms, which offer additional motivation to respondents to provide truthful input. For example, Ding et al. (2005) showed that the external validity of choice-based conjoint (CBC) analysis is dramatically increased when the responses given by consumers have an impact on their compensation. In particular, the authors asked consumers to make a series of choices, such that each respondent had a positive probability of receiving his or her preferred alternative from each choice set after the end of the experiment. Ding (2007) extended this method to situations in which the researcher has access to only a few alternative products and therefore is unable to offer all the alternatives in all the choice sets as potential rewards. Whereas Ding (2007) required estimating how much each respondent would be willing to pay for each potential reward product (based on that respondent's choices), Dong et al. (2010) showed that similar predictive performance may be achieved by simply inferring

the respondent's rank ordering of the set of possible reward products.

Researchers have also recently developed incentive-compatible preference measurement tasks that deviate further from traditional conjoint analysis. For example, Ding et al. (2009) proposed an online incentive-compatible preference measurement mechanism inspired by barter markets. Park et al. (2008) proposed a mechanism that relies on allowing participants to upgrade products, where incentive-compatibility is achieved by using the Becker–De Groot–Marschak (BDM) procedure (i.e., the participant states his or her willingness to pay for an upgrade, a random price is generated, and the transaction is realized at that price if and only if it is lower than or equal to the stated willingness to pay).

The objective of incentive compatibility is to induce truth telling, which requires participant involvement and attention. (Note that involvement and attention are not *sufficient* conditions for truth telling.) However, current incentive compatible preference measurement methods may not increase involvement to the level of real-life purchasing decisions. For example, consider a consumer making a decision on which laptop computer to purchase with his or her own money. He or she may pay more attention to the information relevant to that decision compared with a typical CBC context in which he or she is asked to make many consecutive choices between laptop computers, where each choice has only a probabilistic link to a real outcome and where this outcome involves prize money (as opposed to his or her own money).

In this paper we explore additional ways to increase involvement and attention in preference measurement while maintaining incentive compatibility. We develop and test an incentive-compatible “Conjoint Poker” (CP) game inspired by regular poker. In this game, each card represents a product defined by a combination of features. “Hands” are defined similarly to regular poker (e.g., a pair is a hand in which two products have one feature in common). The preference data revealed by respondents during this game are comparable to incentive-compatible CBC data. We develop a statistical efficiency measure and an algorithm to construct optimal CP designs. We compare incentive-compatible CP to incentive-compatible CBC in a series of three experiments. Our first study, a between-subjects online experiment, provides indirect evidence that CP participants consider more of the profile-related information presented to them compared with CBC participants. We then conduct two eye-tracking studies that provide convergent, direct evidence for this effect.

The rest of this paper is structured as follows. In §2 we introduce CP, develop a model to analyze data from the game, a measure of statistical efficiency for

CP designs and an algorithm for constructing efficient CP designs. We report the results of our experiments in §3 and conclude in §4.

2. An Incentive-Compatible Poker Game for Preference Measurement

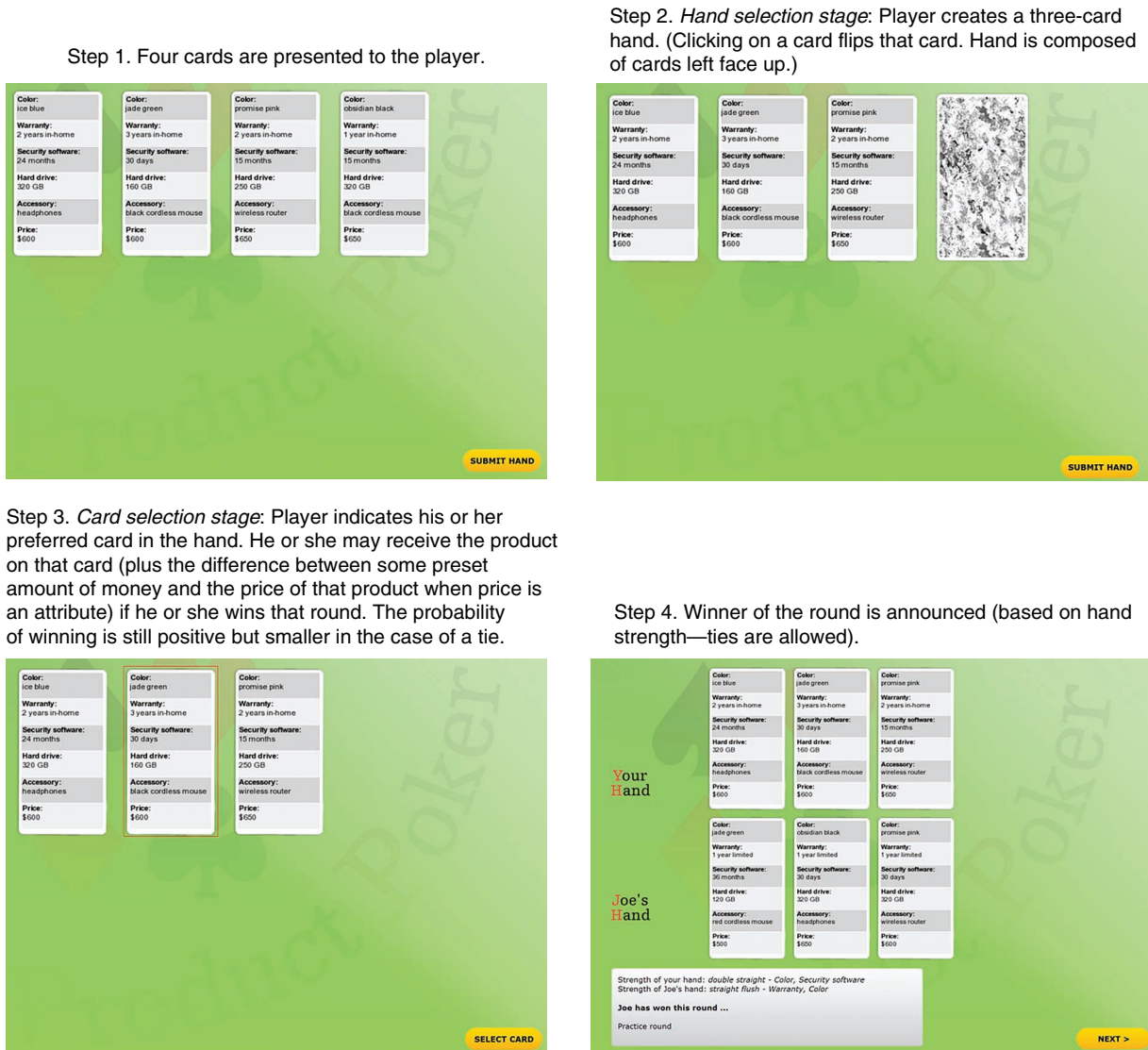
In this section we describe an incentive-compatible Conjoint Poker game. Different versions of the game may be developed; we have developed four-card and seven-card versions. The seven-card version, inspired by Texas Hold'em, is described in the electronic companion (available as part of the online version that can be found at <http://mktsci.journal.informs.org/>). We focus here on the four-card version, which is loosely based on three-card poker.

2.1. Overview

Instead of using traditional playing cards defined by two attributes (number and color) with 13 and 4 levels, respectively, cards in this game represent product profiles described by any number A of attributes ($A = 6$ in our experiments), where each attribute a has L_a levels. See Figure 1 for an example of a card. Although our implementation is online, the game may also be played off-line using physical cards. In its online format, the game may be played between multiple consumers or against the computer. Allowing for multiple players raises several issues with respect to design efficiency, estimation, and learning and information diffusion among respondents (Ding et al. 2009). We leave the investigation of multiple-player versions to future research and focus in this paper on versions of the game in which each respondent plays against the computer only.

Figure 1 A Conjoint Poker Card

Color:	Obsidian black
Warranty:	2 years in-home
Security software:	30 days
Hard drive:	320 GB
Accessory:	Black cordless mouse
Price:	\$550

Figure 2 One Round of Conjoint Poker

The game is played in *rounds*. Each round has two stages, a *hand selection stage* and a *card selection stage*. See Figure 2 for an illustration of the steps involved in each round. In the hand selection stage, each player is asked to form a three-card hand from a set of four cards. As in regular poker, a hand is a set of cards that have a specific pattern. Hands have different strengths, and the winner in each round is the player with the strongest hand (ties are allowed). The strength of a hand is based on the probability that this hand would be achievable from a random set of four cards (drawn without replacement from the set of all cards). Because the number of attributes and levels in CP do not typically match those in regular poker (two attributes with 13 and 4 levels, respectively), the various types of hands are defined a little differently, and the probabilities corresponding to each hand need to be computed. These probabilities

are used to determine the relative strengths of the different hands (such that less likely hands have higher strength) and will also be used in the choice model developed below. See Appendix A for details on the computation of these probabilities. We use the following six types of hands, listed from weakest to strongest (see Figure 3 for an illustration):

—One pair (weakest): two cards have the same level on one attribute.

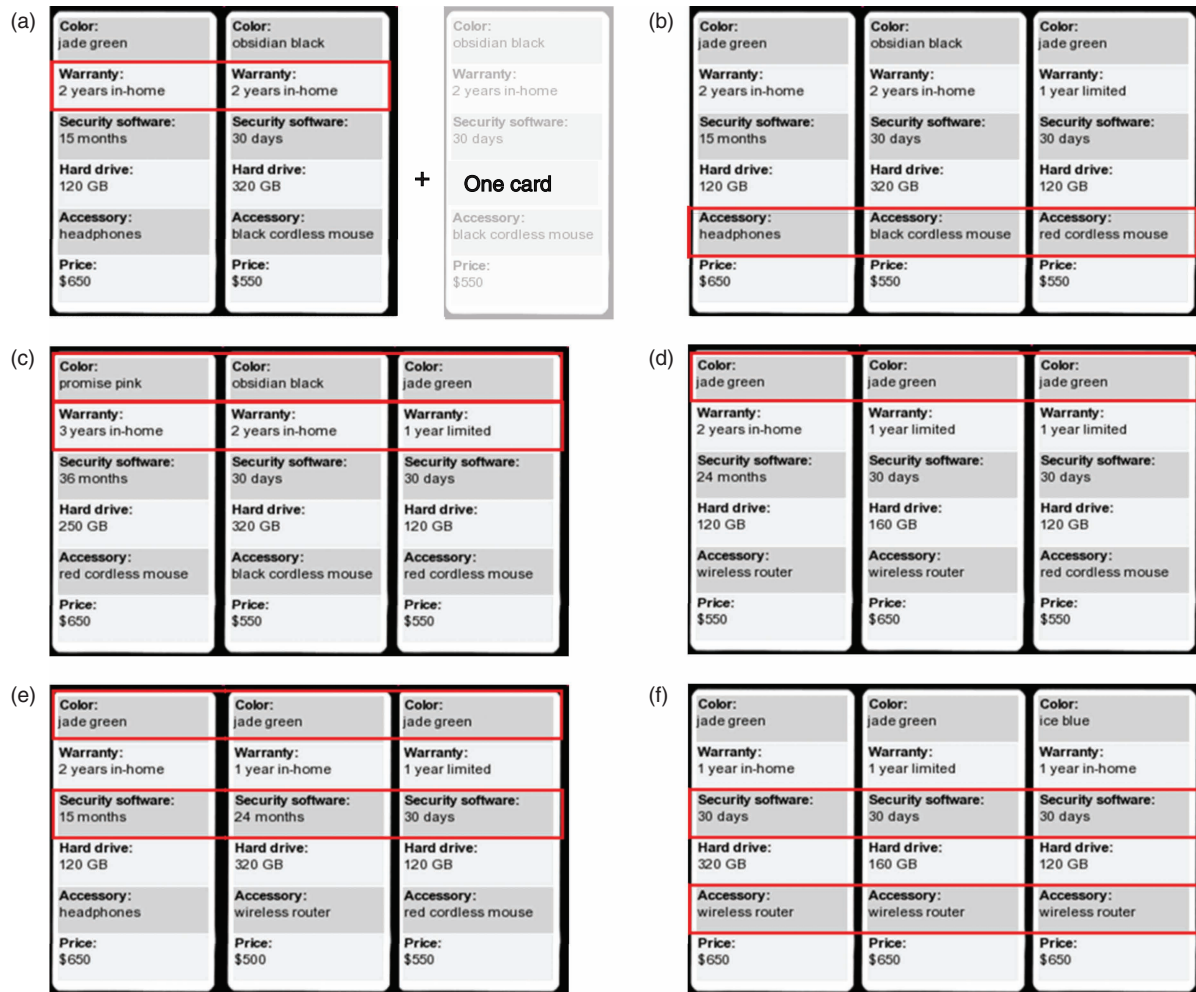
—Straight: all three cards have different levels on one attribute.

—Double straight: all three cards have different levels on two attributes.

—Flush: all three cards have the same level on one attribute.

—Straight flush: all three cards have the same level on one attribute and different levels on another.

Figure 3 Examples of Hands (from Weakest to Strongest): One Pair (a), Straight (b), Double Straight (c), Flush (d), Straight Flush (e), and Double Flush (f)



—Double flush (strongest): all three cards have the same level on two attributes.

In the card selection stage, each player is asked to indicate his or her preferred card from that hand.¹ This information is used to provide incentives to respondents, as described next.

2.2. Incentives

At the end of the experiment, one player is selected randomly, and one of the rounds played by that player is selected randomly. If the player won that round, then he or she wins the product on his or her preferred card. If price is one of the attributes, then the player also receives the difference between a preset amount of money and the price of that product. If the player was tied for best in that round, he or she receives the reward with a probability equal to 1 divided by the number of players in the tie. If the

player lost that round, then he or she receives nothing (except for any potential nominal fee paid to all respondents). In cases in which the experiment also involves an external validity task, there is a positive probability that the incentives will be based on that task instead (see the setup of our experiments in §3).²

For the sake of argument, let us assume risk neutrality in the hand selection stage (this assumption will be relaxed in our final model). We define the utility of a card as the utility of the product on that card plus the utility of the difference between the preset amount of money and the price of that product, if applicable. Given the incentive structure, the expected utility derived by a respondent from a given round if playing a given hand, h , is then proportional to the

¹ In our implementation, players were asked to choose one card. Other implementations may introduce a “no-choice” option.

² This mechanism is based on the “random lottery procedure,” which has been widely used and validated in experimental economics (Starmer and Sugden 1991). See Ding et al. (2005, 2009) for other applications of this procedure to incentive-compatible conjoint analysis.

probability of winning that round with h , multiplied by the utility of the respondent's preferred card in h : *Expected utility if play hand $h \propto$ Probability of winning the round with $h \times$ Utility of preferred card in h .*

This implies a trade-off. On the one hand, players should play strong hands to increase the probability that they will receive a prize. On the other hand, players should also play hands that contain cards that they like in order to increase the utility from the potential prize. This implies that the optimal strategy is neither to always play the strongest hand nor to play a hand that contains the card with the highest utility in each round. In particular, it may be optimal to play a hand that is not the strongest hand but contains a high-utility card or to play a hand that does not contain one's favorite card but has a higher chance of winning the round.

The above expression was provided only to illustrate the basic trade-off faced by CP participants. In reality, we do not expect all consumers to be risk neutral in the hand selection stage. In the next section, we develop a choice model that captures this trade-off without assuming risk neutrality.

2.3. Consumer Choice Model

We now propose a choice model that captures the hand selection and card selection stages of CP. Let i index consumers, r index the rounds in the game, h index the possible hands available to players, and j index the profiles (i.e., cards). When there are four cards per round and when each hand consists of three cards, each player has a choice between four possible hands in each round. We index these four hands by $h \in \{1, 2, 3, 4\}$ and the four cards by $j \in \{1, 2, 3, 4\}$, where $h = 1$ corresponds to the hand made of profiles $\{1, 2, 3\}$, $h = 2$ corresponds to profiles $\{2, 3, 4\}$, etc. With a slight abuse of notation, we write $j \in h$ if card j is present in hand h . Let $h_{i,r}$ be the hand selected by consumer i in round r in the hand selection stage, and let $j_{i,r}$ be the profile selected by that consumer in the card selection stage among the profiles in hand $h_{i,r}$. To estimate partworths from the choices made by consumers, we construct a likelihood function for $\Pr(\{h_{i,r}, j_{i,r}\})$. We have

$$\Pr(\{h_{i,r}, j_{i,r}\}) = \Pr(h_{i,r}) \times \Pr(j_{i,r} | h_{i,r}). \quad (1)$$

The second component, $\Pr(j_{i,r} | h_{i,r})$, corresponds to the card selection stage in which consumer i chooses one profile from a set and may be modeled simply using logistic probabilities:

$$\Pr(j | h_{i,r}) = \begin{cases} \frac{\exp(x_j \beta_i)}{\sum_{j' \in h_{i,r}} \exp(x_{j'} \beta_i)} & \text{if } j \in h_{i,r}, \\ 0 & \text{otherwise,} \end{cases} \quad (2)$$

where β_i corresponds to the partworths for consumer i , and x_j is an appropriately coded row vector that captures the attribute levels in profile j .

The other term in Equation (1), $\Pr(h_{i,r})$, corresponds to the hand selection stage, which is unique to CP. Here also, the consumer needs to make a choice between a set of possible alternatives (i.e., hands). In the case of risk neutrality, the expected utility from choosing a hand is proportional to product of the probability of winning with that hand and the utility of the preferred card in that hand. Note that whereas a utility intercept is not identified from the card selection stage alone in the absence of a "no-choice" option, such an intercept is identified from the hand selection stage because of the possibility of not winning anything in the round. In particular, if we normalize the utility of not winning anything (i.e., of losing the round) to 0, then the expected utility obtained by a risk-neutral consumer i from playing hand h is equal to $V_h^{i,r} = P_{h|r}^w (\alpha_i + \max_{j \in h} \{x_j \beta_i\})$, where $P_{h|r}^w$ is the probability of winning round r by playing hand h , and α_i is an intercept that captures consumer i 's utility from winning a prize in the game. In a two-player game, $P_{h|r}^w$ is equal to the probability that the other player's best hand is strictly weaker than h plus half of the probability that the other player's best hand is exactly as good as h (ties are broken randomly). In our experiments, each consumer played against the computer, the computer's cards were drawn randomly without replacement from the set of all possible cards, and the computer always played the strongest hand in each round. The resulting winning probabilities $P_{h|r}^w$ are computed in closed form in Appendix A.

The above expression assumed risk neutrality in the hand selection stage; i.e., the utility of the preferred card in a hand was assumed to have a proportional influence on the player's evaluation of the hand. Risk-averse (respectively, risk-seeking) behavior is obtained when that the player's evaluation of the hand is a concave (respectively, convex) function of the utility of the preferred card in the hand. Making the standard assumption of constant relative risk aversion gives rise to the following expression:

$$V_h^{i,r} = P_{h|r}^w \left(\alpha_i + \max_{j \in h} \{x_j \beta_i\} \right)^{\sigma_i}, \quad (3)$$

where σ_i is the risk-aversion parameter for consumer i . Risk neutrality is obtained when $\sigma_i = 1$, $\sigma_i < 1$ gives rise to risk aversion in hand selection, and $\sigma_i > 1$ gives rise to risk-seeking behavior.

We then model consumer i 's choice in the hand selection stage as

$$\Pr(h) = \frac{\exp(\delta \cdot V_h^{i,r})}{\sum_{h' \in \mathcal{H}} \exp(\delta \cdot V_{h'}^{i,r})}, \quad (4)$$

where δ is a logit scale parameter that captures the possibility that the response error is different in the hand selection and card selection stages (e.g., because of misrepresentations of the probabilities of winning).³

2.4. Design Algorithm

When estimating partworths from conjoint analysis data, the statistical efficiency of the estimates, typically measured by their asymptotic covariance matrix, depends on the profiles shown to consumers. In the case of CBC, this asymptotic covariance matrix depends further on the partworth estimates themselves (Huber and Zwerina 1996). Under mild conditions the asymptotic covariance matrix is equal to the inverse of the information matrix (see, for example, McFadden 1974, Newey and McFadden 1994). A series of statistical efficiency measures have been proposed, typically based on various types of matrix norms (e.g., determinant, trace) applied to the information matrix, and various design algorithms have been proposed to select profiles that maximize efficiency. In CBC, a common efficiency measure is D-efficiency (based on the determinant norm), and a common approach for creating D-efficient designs is to obtain some prior information on the partworths and, given that information, to apply a set of operators to transform a nonefficient design into a D-efficient design (see, for example, Huber and Zwerina 1996; Sandor and Wedel 2001, 2002, 2005). Four properties characterize D-efficient CBC designs: level balance (the levels of an attribute occur with equal frequency), orthogonality (any two levels of different attributes appear in profiles with frequencies equal to the product of their marginal frequencies), minimal overlap (each attribute level repeats itself within each choice set with minimal probability), and utility balance (profiles in each choice are similarly attractive).

This design approach may be extended to the construction of efficient CP designs. The key difference is in the computation of the information matrix. Given our choice model above, the log-likelihood function is equal to

$$\begin{aligned} L(\text{data} | X, \beta, \alpha, \sigma, \delta) \\ = \sum_i \sum_r \log(\Pr(h_{i,r})) \\ + \log(\Pr(j_{i,r} | h_{i,r})) + \text{constant}. \end{aligned}$$

³ An alternative version of the game would be such that players still win some amount of money if they lose the round. The same modeling framework could be applied to a situation like this by simply replacing $V_{h_i}^{i,r} = P_{h_i|r}^w(\alpha_i + \max_{j \in h} \{x_j \cdot \beta_j\})^{\sigma_i}$ with $V_{h_i}^{i,r} = P_{h_i|r}^w(\alpha_i + \max_{j \in h} \{x_j \cdot \beta_j\})^{\sigma_i} + (1 - P_{h_i|r}^w) \cdot \delta_i$, where δ_i would capture the utility from winning the amount of money offered when the round is lost.

Taking the Hessian of the likelihood function with respect to β , we find that the expected value of the information matrix Ω is proportional to (see Appendix B for details):⁴

$$\Omega = \sum_r \left[\sum_{h_r \in r} \tilde{z}_{h_r}^r \Pr(h_r) \tilde{z}_{h_r}^{rT} + \Pr(h_r) \sum_{j \in h_r} z_{j|h_r} \Pr(j | h_r) z_{j|h_r}^T \right], \quad (5)$$

where $\Pr(h_r)$ is given by (4), $\Pr(j | h_r)$ is given by (2), and

$$\tilde{z}_{h_r}^r = \delta \cdot \nabla V_{h_r}^r - \sum_{h'_r} \delta \cdot \nabla V_{h'_r}^r \Pr(h'_r),$$

$$z_{j|h_r} = x_j^T - \sum_{j' \in h_r} x_{j'}^T \Pr(j' | h_r),$$

$$\nabla V_{h_r}^r = \sigma P_{h_r|r}^w(\alpha + x_{j_{h_r}^*} \beta)^{\sigma-1} x_{j_{h_r}^*}^T, \quad \text{where } j_{h_r}^* = \arg \max_{j \in h_r} (x_j \beta).$$

The D-efficiency of any CP design may be computed from the determinant of the corresponding information matrix using the same formulas as the ones used to compute the D-efficiency of a CBC design given its information matrix. Moreover, the same operators (e.g., swapping, relabeling) used to improve the D-efficiency of a CBC design may also be used to improve the D-efficiency of a CP design.

One important difference between CP and CBC is that efficient CP designs tend to have more level overlap compared with efficient CBC designs such as those considered here (i.e., standard logit designs).⁵ Minimal overlap occurs when attribute levels are repeated within a choice set as little as possible. For example, in our experiments we achieved no overlap in our CBC design; i.e., each level of each attribute appeared exactly once in each choice set, leading to maximal design efficiency. In contrast, some overlap is actually desirable in CP designs. Indeed, in a CP design with no level overlap, all hands in all rounds are equally strong (all hands are double straights, and $P_{h_i|r}^w$ is constant for all r and h), and the intercept α_i and risk-aversion parameter σ_i are not identified (all hands have the same probability of winning).

In our field study, we followed Huber and Zwerina (1996) and conducted a pretest from which we

⁴ The expression in Equation (5) assumes a homogeneous partworth vector β (Huber and Zwerina 1996). In Appendix B, we provide the information matrix for the mixed logit model. This information matrix involves integrals and does not have a closed form (Sandor and Wedel 2002). For computational simplicity, the designs used in our experiments assume a homogeneous partworth vector (i.e., standard logit model).

⁵ Sandor and Wedel (2002) showed that efficient mixed logit designs tend to have more level overlap compared with efficient standard logit designs.

obtained prior estimates (our pretest had 56 and 58 respondents in the CBC and CP conditions, respectively). We then applied the relabeling and swapping operators to improve the D-efficiency of a design that was selected as a starting point (see, for example, Huber and Zwerina 1996; Sandor and Wedel 2001, 2002, 2005). For CBC, we used a D_0 -efficient design with no overlap as a starting point (a D_0 -efficient is a D-efficient design assuming that all partworths are equal to 0). This D_0 -efficient design was obtained using the standard cyclic approach of Bunch et al. (1994): starting with an orthogonal design, a set of choice alternatives was constructed by adding cyclically generated alternatives to each set. For CP, we used a perturbed version of that D_0 -efficient design as a starting point, where the perturbation was performed in order to introduce some amount of level overlap.⁶

3. Experiment Details

We now report the results of three experiments that compared incentive-compatible CP to incentive-compatible CBC. Our first study was a between-subjects online experiment that enabled us to compare the partworth estimates across the two methods and provided indirect evidence that CP participants considered more of the profile-related information presented to them compared with CBC participants. Studies 2 and 3 are within-subjects eye-tracking studies that enabled us to measure more directly the amount of information considered by respondents under the two methods.

3.1. Setup

The setup of all three studies was similar. CBC and CP were implemented on the same online platform with the same user interface. See Figure 2 for screenshots of the CP interface and Figure 4 for screenshots of the CBC interface. We used mini laptops as our product category. Each of our product profiles was a different customized version of the Dell Inspiron Mini 11z laptop computer. The display (11.6" high-definition

WLED display), processor (1.3 GHz Intel® Celeron 743 processor), RAM (2 GB), battery (28Whr lithium-ion battery), operating system (Genuine Windows® Vista Home Edition), and webcam (integrated 1.3 M pixel webcam) were held constant. Six attributes were varied, with four levels each: design (four different color schemes: promise pink, obsidian black, jade green, or ice blue), warranty (1-year limited, 1-year limited with in-home service after remote diagnosis, 2-year limited with in-home service after remote diagnosis, or 3-year limited with in-home service after remote diagnosis), McAfee® SecurityCenter antivirus (30-day, 15-month, 24-month, or 36-month subscription), hard drive (120 GB, 160 GB, 250 GB, or 320 GB), accessory (Logitech® black cordless mouse, Logitech red cordless mouse, Linksys wireless router, or Creative Labs headphones), and price (\$500, \$550, \$600, or \$650). Given this price range, and given the fact that Dell laptops are customizable, we were able to offer any product profile as an incentive (like in Ding et al. 2005).⁷

In each study, the flow of the experiment was as follows, for each respondent:

1. *Instructions*: Detailed instructions (using picture illustrations) were displayed on the introductory page. Care was taken to make the instructions in both conditions as symmetric as possible. In addition, we created a slide show for each condition with a shorter version of the instructions, which we embedded at the top of the page (using <http://www.authorstream.com>—slide shows and instructions are available from the authors upon request). The slide shows in both conditions used similar language (e.g., in both conditions, profiles were referred to as “cards”). Throughout the experiment, a link to the instructions was available to participants.

2. *External validity task*: As an external validity task, participants were asked to select one card from a set of eight. See Figure 5 for a screenshot. The format of the external validity task and the set of eight profiles were identical across conditions and participants. These eight profiles were randomly selected subject to a level-balancing constraint (each level of each attribute appears exactly twice across profiles) and such that exactly two of the four ordered attributes (warranty, security software, hard drive, price) were at one of the two most attractive levels in each profile. This last constraint was added to avoid dominance.

3. *Main task*: Participants in the CBC condition were given 20 CBC questions, each with four alternatives. Participants in the CP condition were asked to play 20 CP rounds against the computer. (CP participants were given two practice rounds between Steps 2

⁶ The cyclic approach of Bunch et al. (1994) is such that if attribute a is at level l in the q th profile of the orthogonal design, then it is at level l in the first card in the q th question of the choice design, level $l+1$ in the second card, etc. In other words, starting from the orthogonal design, the level of each attribute is incremented by 1 for each new card in the choice set. In our perturbed version, it was incremented with probability $1-t$ instead of 1. Therefore if attribute a was at level l in the q th profile of the orthogonal design, it was at level l in the first card in the q th round of Conjoint Poker. The value in the second card was then equal to $l+1$ with probability $1-t$ and to l with probability t . The value in the third card was then equal to that in the second card plus 1 with probability $1-t$ and equal to that in the second card with probability t , etc. We used $t=0.35$.

⁷ Although appealing, this property is not required. See, for example, Ding (2007) and Dong et al. (2010).

Figure 4 One CBC Question

The screenshot shows a green background with a large, faint 'Conjoint Poker' watermark. Four white cards are arranged horizontally, each containing product attributes. A yellow 'SELECT CARD' button is in the bottom right corner.

Card 1	Card 2	Card 3	Card 4
Color: ice blue	Color: promise pink	Color: jade green	Color: obsidian black
Warranty: 1 year limited	Warranty: 1 year in-home	Warranty: 2 years in-home	Warranty: 3 years in-home
Security software: 30 days	Security software: 15 months	Security software: 36 months	Security software: 24 months
Hard drive: 250 GB	Hard drive: 320 GB	Hard drive: 160 GB	Hard drive: 120 GB
Accessory: black cordless mouse	Accessory: red cordless mouse	Accessory: wireless router	Accessory: headphones
Price: \$650	Price: \$550	Price: \$500	Price: \$600

and 3.) See Figures 2 and 4 for screenshots from each condition. The CBC and CP designs were selected to be D-efficient as explained above. The only exception is the second eye-tracking study (Study 3), in which

both designs were identical. Within each condition, the design was identical across respondents.

4. *Follow-up questionnaire:* Participants were administered a short follow-up survey that measured

Figure 5 External Validity Task

The screenshot shows a green background with a large, faint 'Conjoint Poker' watermark. Eight white cards are arranged in two rows of four. A yellow 'SELECT CARD' button is in the bottom right corner.

Card 1 (Top)	Card 2 (Top)	Card 3 (Top)	Card 4 (Top)	Card 1 (Bottom)	Card 2 (Bottom)	Card 3 (Bottom)	Card 4 (Bottom)
Color: promise pink	Color: jade green	Color: ice blue	Color: obsidian black	Color: promise pink	Color: jade green	Color: ice blue	Color: obsidian black
Warranty: 1 year in-home	Warranty: 1 year limited	Warranty: 2 years in-home	Warranty: 3 years in-home	Warranty: 3 years in-home	Warranty: 1 year limited	Warranty: 2 years in-home	Warranty: 1 year in-home
Security software: 36 months	Security software: 30 days	Security software: 36 months	Security software: 24 months	Security software: 24 months	Security software: 15 months	Security software: 15 months	Security software: 30 days
Hard drive: 320 GB	Hard drive: 250 GB	Hard drive: 160 GB	Hard drive: 160 GB	Hard drive: 120 GB	Hard drive: 320 GB	Hard drive: 120 GB	Hard drive: 250 GB
Accessory: red cordless mouse	Accessory: headphones	Accessory: wireless router	Accessory: wireless router	Accessory: headphones	Accessory: red cordless mouse	Accessory: black cordless mouse	Accessory: black cordless mouse
Price: \$600	Price: \$550	Price: \$650	Price: \$600	Price: \$650	Price: \$550	Price: \$500	Price: \$500

their feedback on the task, their knowledge of regular poker, and how much they would need to be paid in order to participate in a similar study in the future. (More details are provided in §3.2.3.)

3.2. Study 1

Our participants were recruited from a commercial online panel typical of online market research. Participants were *not* screened based on their knowledge of poker or any other criteria. Respondents accessed the experiment from the Internet. We used a between-subjects design with $N = 318$ in each condition. Respondents were given a flat nominal incentive to participate (in the form of points that may be redeemed for rewards), in addition to the following incentives. In both conditions participants were informed (in Step 1) that one participant would be selected randomly at the end of the study and that one of the tasks (external validity or CBC in the incentive-compatible CBC condition, or external validity or CP in the incentive-compatible CP condition) would be randomly selected for that participant. If the external validity task was to be selected, the winner would receive the product on his or her preferred card and the difference between \$650 and the price of that product. If the CBC task was to be selected, then one of the 20 CBC questions would be selected, and the winner would receive his or her preferred card from that question and the difference between \$650 and the price of that product. If the CP task was to be selected, then one of the rounds would be randomly selected. If the selected respondent was to have won this round, then he or she would receive the product on his or her preferred card in that hand and the difference between \$650 and the price of that product. If the selected respondent was to have lost this round, then he or she would receive nothing (with the exception of the flat incentive offered to all respondents). If the selected respondent was to have been tied with the computer in this round, then he or she would receive the prize with probability 0.5. Note that the incentives given to the participants in the CP condition were smaller in expectation compared to the CBC condition, making the comparisons conservative.

3.2.1. Parameter Estimates. We estimated the respondents' partworths in each condition using hierarchical Bayes (HB) (Rossi and Allenby 2003). Details are provided in Appendix C. In each condition, the first stage prior on β_i was normal: $\beta_i \sim N(\beta_0, D_\beta)$. In the CP condition, we also specified priors on σ_i (risk-aversion parameter) and α_i (intercept): $\sigma_i \sim \text{TN}(\sigma_0, D_\sigma)$ (truncated between 0 and 2 for numerical stability) and $\alpha_i \sim N(\alpha_0, D_\alpha)$. Table 1 reports the estimates of the population means of all partworths (i.e., posterior mean of the first-stage

prior parameters) as well as 95% credible intervals for these population means and the point estimate of the standard deviation of each parameter across consumers (i.e., the square root of the diagonal terms of the matrix D_β). We used effects coding such that the partworths always sum to 0 within an attribute. We see that the average partworth estimates obtained from the two methods are fairly different. In fact, the 95% credible intervals do not overlap for 9 out of the 24 attribute levels. The partworth estimates based on incentive-compatible CBC are also more heterogeneous than those based on incentive-compatible CP. The average standard deviation (across attribute levels) of the partworths is equal to 0.748 with incentive-compatible CBC and 0.414 with incentive-compatible CP. The partworths obtained from incentive-compatible CBC are also more varied at the individual level. In particular, the average across respondents of the ratio between the importances of the most and the least important attributes is 20.238 under incentive-compatible CBC compared with 11.803 under incentive-compatible CP.

Further analysis of the raw data provides an explanation for these differences. In particular, a substantial proportion of respondents in the incentive-compatible CBC condition made all of their CBC choices based on a very limited subset of attributes and levels. For example, 74 of the 318 participants in the incentive-compatible CBC condition selected profiles that all had one attribute in common (e.g., all the profiles selected across the 20 CBC questions had 320 GB hard drives). Recall that D-efficient CBC designs have minimal overlap; i.e., each attribute level repeats itself within each choice set with minimal probability. In our case, a design with no overlap was available; i.e., in each choice set, each attribute level was present exactly once. This implies that partworth estimates for these respondents are very uneven and are only identified through shrinkage. Consider, for example, a respondent who always selects laptops with 320 GB in all CBC questions. Because there is only one 320 GB laptop in each question, the choice data from this respondent may be fitted perfectly by assigning a very large positive partworth to 320 GB and smaller weights to all other attribute levels, giving rise to uneven partworth estimates. Moreover, the only information contained in this respondent's data is that 320 GB is greatly preferred to all other partworths, but no information is provided on how large the difference is or on the relative preferences for the other attribute levels.⁸ Therefore this respondent's

⁸ More precisely, as long as the partworth for 320 GB minus the partworth for the next level of this attribute is greater than the sum over all other attributes of the highest minus the lowest partworths, the respondent would always prefer the laptop with 320 GB, irrespective of the other attribute levels.

Table 1 Partworth Estimates

Parameter	Incentive-compatible CBC	CP
Color		
Promise pink	0.386, [0.180, 0.597], 1.449	−0.498, [−0.601, −0.400], 0.598
Obsidian black	−0.610, [−0.802, −0.434], 1.078	0.215, [0.113, 0.325], 0.375
Jade green	0.276, [0.126, 0.419], 0.873	0.122, [0.039, 0.202], 0.361
Ice blue	−0.052, [−0.581, 0.496], 0.999	0.162, [−0.127, 0.449], 0.360
Warranty		
1-year limited	−0.022, [−0.144, 0.106], 0.789	−0.364, [−0.437, −0.297], 0.280
1-year in-home	−0.111, [−0.212, −0.010], 0.581	−0.376, [−0.453, −0.299], 0.354
2-year in-home	0.018, [−0.058, 0.100], 0.343	0.246, [0.192, 0.313], 0.182
3-year in-home	0.115, [−0.196, 0.415], 1.032	0.493, [0.282, 0.698], 0.535
Security software		
30 days	−0.229, [−0.361, −0.105], 0.709	−0.340, [−0.416, −0.248], 0.294
15 months	−0.091, [−0.160, −0.013], 0.229	−0.208, [−0.291, −0.134], 0.172
24 months	−0.054, [−0.173, 0.059], 0.464	0.183, [0.122, 0.244], 0.197
36 months	0.375, [0.033, 0.694], 0.441	0.365, [0.138, 0.585], 0.321
Hard drive		
120 GB	−0.894, [−1.150, −0.613], 1.174	−1.337, [−1.499, −1.179], 1.089
160 GB	−0.555, [−0.702, −0.403], 0.905	−0.599, [−0.680, −0.514], 0.460
250 GB	0.533, [0.392, 0.678], 0.809	0.529, [0.439, 0.607], 0.370
320 GB	0.916, [0.339, 1.460], 1.727	1.407, [1.086, 1.741], 1.197
Accessory		
Black mouse	−0.142, [−0.248, −0.028], 0.572	0.047, [−0.013, 0.109], 0.273
Red mouse	−0.228, [−0.321, −0.146], 0.424	−0.103, [−0.166, −0.039], 0.251
Wireless router	0.473, [0.304, 0.638], 1.214	0.212, [0.127, 0.294], 0.462
Headphones	−0.104, [−0.464, 0.265], 0.402	−0.155, [−0.364, 0.052], 0.309
Price		
\$500	−0.042, [−0.162, 0.073], 0.475	−0.069, [−0.157, 0.024], 0.391
\$550	−0.064, [−0.162, 0.020], 0.387	−0.001, [−0.086, 0.076], 0.331
\$600	0.098, [0.004, 0.184], 0.254	−0.018, [−0.080, 0.038], 0.191
\$650	0.008, [−0.276, 0.320], 0.616	0.088, [−0.137, 0.323], 0.583

Notes. The population means estimate is reported first. The 95% credible interval of the population means estimate is reported second in brackets, and the point estimate of the standard deviation of the parameter across consumers is reported last.

partworths are identified only because of shrinkage; i.e., without the shrinkage performed by HB, the likelihood function corresponding to such a respondent would not have a finite maximum.⁹

Interestingly, the 74 incentive-compatible CBC respondents who selected profiles that were always similar on one attribute may actually be grouped based on the attribute level that is common across their preferred profiles. For example, 29 of these respondents always selected laptops with 320 GB hard drives, 13 always selected laptops with wireless routers, and 9 always selected pink laptops. Because there was no level overlap in the D-efficient CBC design, within each of these groups, all respondents gave the exact same answers to all CBC questions. In fact, 90 of the 318 incentive-compatible CBC respondents (28.30%) gave answers that were completely

identical to the answers of at least one other respondent in that condition. Such a phenomenon was not present in the CP data. None of the CP respondents gave answers that were all similar to the answers of another CP respondent, and none of the CP respondents selected profiles that all had one attribute in common.

Although the evidence so far is only indirect, these observations suggest that CBC respondents considered less of the profile-related information when making their choices compared with CP respondents. The next two studies confirm this phenomenon by providing direct evidence from eye-tracking data.

Finally, we discuss the estimates of the other parameters that are unique to CP. Table 2 reports the estimates of δ , α_0 , and σ_0 , and Figure 6 reports the distribution of the estimates of the risk-aversion parameter σ_i across CP participants. For each respondent, we construct a 95% credible interval for σ_i . For 3 of the 318 respondents, the upper bound of this credible interval is below 1, which corresponds to risk aversion in the hand selection stage. For 93 of

⁹ Johnson (2008, p. 3) also notes the limitations of questionnaire designs with minimal overlap when respondents “use simplification strategies, such as always choosing a preferred brand, or the lowest price, or the most extreme level of some other attribute.”

Table 2 Additional Conjoint Poker Parameters

Parameter	Posterior mean [95% credible interval]
δ	0.064, [0.027, 0.126]
α_0	43.741, [34.175, 51.752]
σ_0	1.121, [0.905, 1.304]

Note. The 95% credible intervals are reported in brackets.

the 318 respondents, this credible interval is above 1, which corresponds to risk seeking in the hand selection stage. Therefore, our results suggest that a significant proportion of CP participants are risk seeking in their selection of hands. Note that risk-seeking behavior is not uncommon in settings such as this one (see, for example, Thaler and Johnson 1990).

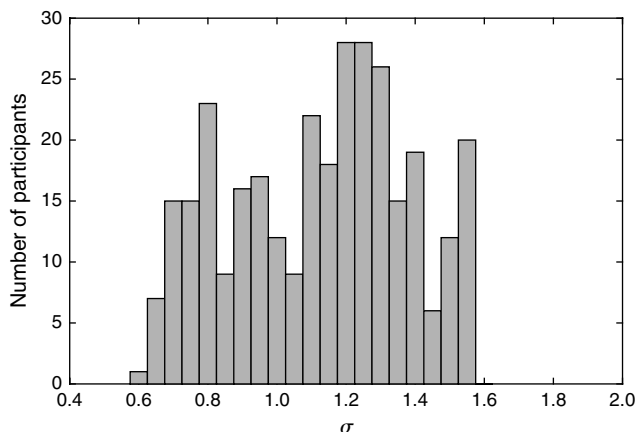
3.2.2. Performance Comparisons. As we show later, our eye-tracking studies suggest that respondents consider only a subset of the profile-related information contained in the external validity task. This raises questions about the ecological validity of such tasks in general and about their ability to detect a deeper level of processing in the calibration task in particular. Indeed, it is unclear how a deeper level of processing in the calibration task would impact predictive performance in the validation task, if the level of processing in the validation task itself is not as deep. With this caveat, in this section we compare the predictive performance of CBC versus that of CP for completeness.

We refer to the respective choice shares of the eight profiles in the external validity task based on the entire sample (i.e., both conditions combined—636 respondents) as the “observed out-of-sample aggregate choice shares.” For each method, we compute at each posterior draw in the Markov chain Monte Carlo (MCMC) the logistic probability that each respondent will choose each of the eight profiles

in the external validity task. Averaging these probabilities across posterior draws gives us point estimates of each respondent’s choice probabilities in the external validity task. We use these estimated probabilities to compute hit rates (where the prediction is that each respondent will choose the profile with the highest predicted choice probability). We also measure how well the aggregate predictions made by each method coincide with the observed out-of-sample aggregate choice shares. We compute an estimate of the out-of-sample aggregate choice shares at each posterior draw in the MCMC by averaging the choice probabilities across respondents. Point estimates are obtained by averaging these estimates across posterior draws. The accuracy of these predictions is measured using the mean absolute deviation (MAD), the correlation, and the root mean squared error (RMSE) between the predicted and observed out-of-sample aggregate choice shares. For all performance metrics, we also report 95% credible intervals across posterior draws. We note that the accuracy of out-of-sample choice share predictions is typically assumed to be highly correlated with the accuracy of out-of-sample individual-level predictions, and it is therefore often not reported in papers on CBC. However, we will see that this correlation may not be as high as is usually assumed.

The results are reported in Table 3. The hit rates achieved by incentive-compatible CBC and incentive-compatible CP are not significantly different (i.e., the 95% credible intervals overlap¹⁰). However, incentive-compatible CP produces out-of-sample aggregate choice share predictions that are significantly more accurate than those produced by incentive-compatible CBC. In fact, the aggregate choice share predictions based on incentive-compatible CBC are worse than chance.¹¹ Table 4 gives the actual and predicted choice shares for each profile in the external validity set. The fact that incentive-compatible CBC predicts out-of-sample individual choices well but predicts out-of-sample aggregate choice shares poorly is counterintuitive and worthy of subsequent investigation. In particular, future research may test the replicability and generalizability of this finding and identify its cause.

3.2.3. Follow-up Questionnaire. After completing the task, subjects in both conditions answered a follow-up questionnaire. The goal of this questionnaire was to further test the level of engagement of

Figure 6 Distribution Across Conjoint Poker Participants of the Point Estimate of the Risk-Aversion Parameter σ_i 

¹⁰ The difference is not significant either when comparing the two sets of hit rates using a *t*-test (*p*-value = 0.74).

¹¹ The difference between the accuracy of the out-of-sample choice share predictions made by the two methods is even more pronounced when the accuracy of each method is evaluated based on the shares observed in that same condition (details available from the authors upon request).

Table 3 Out-of-Sample Predictions

	Individual-level predictions: Hit rate	Choice share predictions: MAD	Choice share predictions: correlation	Choice share predictions: RMSE
Incentive-compatible CBC	0.330 [0.296,0.368]	0.0691 [0.0603,0.0757]	0.432 [0.354,0.522]	0.0814 [0.0715,0.0902]
CP	0.318 [0.236,0.315]	0.0267 [0.0208,0.0337]	0.890 [0.839,0.921]	0.0312 [0.0265,0.0387]

Notes. Based on external validity task. A null model predicting equal choice probabilities across eight profiles would achieve a hit rate of 0.125, a MAD of 0.0623, and a RMSE of 0.0667 (correlation is not defined when all share predictions are equal). The 95% credible intervals are reported in brackets.

Table 4 Detailed Choice Share Predictions

Profile description						Observed choice share (%)	Choice share predicted by incentive-compatible CBC (%)	Choice share predicted by CP (%)
Color	Warranty	Security software	Hard drive	Accessory	Price (\$)			
Pink	1 yr in-home	36 mos	320 GB	Red mouse	600	17.92	32.42	19.89
Green	1 yr limited	30 days	250 GB	Headphones	550	3.62	11.99	6.95
Blue	2 yrs in-home	36 mos	160 GB	Wireless router	650	19.65	14.34	16.65
Black	3 yrs in-home	24 mos	160 GB	Wireless router	600	16.35	14.21	17.47
Pink	3 yrs in-home	24 mos	120 GB	Headphones	650	3.14	7.91	5.06
Green	1 yr limited	15 mos	320 GB	Red mouse	550	19.34	10.26	21.70
Blue	2 yrs in-home	15 mos	120 GB	Black mouse	500	5.82	5.31	4.60
Black	1 yr in-home	30 days	250 GB	Black mouse	500	14.15	3.57	7.68

the respondents and to test their knowledge of regular poker.

Engagement. Our results so far (and the eye-tracking studies reported next) suggest that CP participants consider a greater proportion of the profile-related information compared with CBC respondents when making their choices. This increased effort is also reflected in the time spent by participants on the task. The average time used by CP participants to complete 20 rounds of CP was 1,335.44 seconds, compared to an average time of 625.76 seconds to complete 20 CBC questions in the other condition. To the extent that a longer response time is a sign of increased information processing, this is not necessarily a negative feature. However, from a managerial perspective, a longer response time may indicate that consumers would demand higher incentives to participate in a CP study compared to a CBC study (holding the number of rounds/questions constant). This would imply that for a fixed budget, more consumers could be surveyed using CBC compared to CP.

Our follow-up questionnaire allowed us to investigate these issues. Respondents were asked to indicate on a five-point scale from 1 (“strongly disagree”) to 5 (“strongly agree”) how much they agreed with the following statements:

- “Participation in this study was fun.”
- “The instructions were complex.”
- “Participation in this study required a lot of effort.”
- “Participation in this study required a lot of time”

In addition, in the spirit of Dong et al. (2010), we measured how much money respondents in each

condition would require in order to participate in a similar study in the future. Participants were given the following instructions:

We often run studies similar to this one, using the same methodology but on products different from laptop computers. Would you like to participate in such future studies? Assuming that your only compensation would be a fixed payment (i.e., you would not be entered into a lottery to win a product), how much would that compensation need to be for you to be willing to participate in a similar study of the same length? Please enter an amount in dollars below. We will invite you to a future study of the same length only if the payment for that study is at least as large as the amount that you write. Otherwise, we will not send you an invitation.

As expected, compared to participants in the incentive-compatible CBC condition, participants in the CP condition found the instructions significantly more complex (2.81 versus 1.74; Wald = 121.03, $p < 0.01$), found that the study required significantly more effort (2.70 versus 1.84; Wald = 79.28, $p < 0.01$), and found that it was significantly more time consuming (2.96 versus 2.06; Wald = 79.56, $p < 0.01$). However, they also thought that the study was significantly more “fun” (average scores of 4.12 versus 3.97; Wald = 6.90, $p < 0.01$), and the amounts of money that they requested in order to participate in a similar survey in the future were similar. In both conditions, the median was equal to \$5.00. (The average was significantly higher in the CBC condition because of the presence of outliers.) Moreover, a Kolmogorov–Smirnov ($K-S$) test could not reject the null hypothesis that the amounts provided in the two

conditions come from the same distribution ($K-S = 0.082$, $p = 0.23$).

To conclude this analysis, incentive-compatible CP is perceived as being more complex, more effortful, and more time consuming, and it is objectively more time consuming compared with incentive-compatible CBC. But it is also perceived to be more enjoyable, and consumers do not request higher incentives to participate in a CP versus a CBC study, holding the number of rounds/questions constant.

Knowledge of Regular Poker. Our follow-up questionnaire also tested the respondents' knowledge of traditional poker using a six-question quiz (see Appendix D—none of the answers to these questions could be inferred from the previous tasks). We created a score for each respondent equal to the number of correct answers. First, as expected, we found that there was no significant difference in the average score across the two conditions (4.48 versus 4.55; $t = 0.708$, $p\text{-value} = 0.48$). The proportion of respondents with perfect scores was also not significantly different across the two conditions (36.79% versus 34.59%; $z = 0.579$, $p = 0.72$). We tested whether CP affects the partworths estimates differently for poker experts versus nonexperts. Although we may expect poker experts to have preferences that systematically differ from those of other consumers (related to differences in gender, age, etc.), it is important that the CP interface does not bias or distort the preferences of poker experts relative to nonexperts. We reestimated the partworths in each condition with a first-stage prior of $\beta_i \sim N(\beta_0 + \beta_1 \text{Expert}_i, D_\beta)$, where Expert_i was a binary variable equal to 1 if respondent i received a perfect score on the poker quiz. The vector β_1 was estimated directly within the MCMC sampler (see, for example, Lenk et al. 1996). The 95% credible intervals for all 24 attribute levels captured by β_1 in the CBC condition overlap with those in the CP condition. This suggests that the CP task itself does not distort or biases the preferences of poker experts relative to nonexperts.

Our results so far are consistent with CBC respondents making their choices based on a smaller subset of the profile-related information presented to them compared with CP respondents. The following two eye-tracking studies provide direct evidence that CP participants consider more of the profile-related information presented to them compared with CBC participants. These studies also shed some light on the source of this difference, by separating the effect of differences in designs (i.e., sets of cards) from the effect of differences in the methods themselves.

3.3. Study 2

Eye-tracking has been used by several authors in marketing to model visual attention for brands, ads, and

in-store marketing (e.g., Chandon et al. 2009, Pieters and Wedel 2004, Wedel and Pieters 2000). In our setting, if CBC participants make their choices based on a smaller subset of attributes, they should consider a smaller proportion of the profile-related information available to them, resulting in decreased visual attention for a large proportion of attributes and levels.

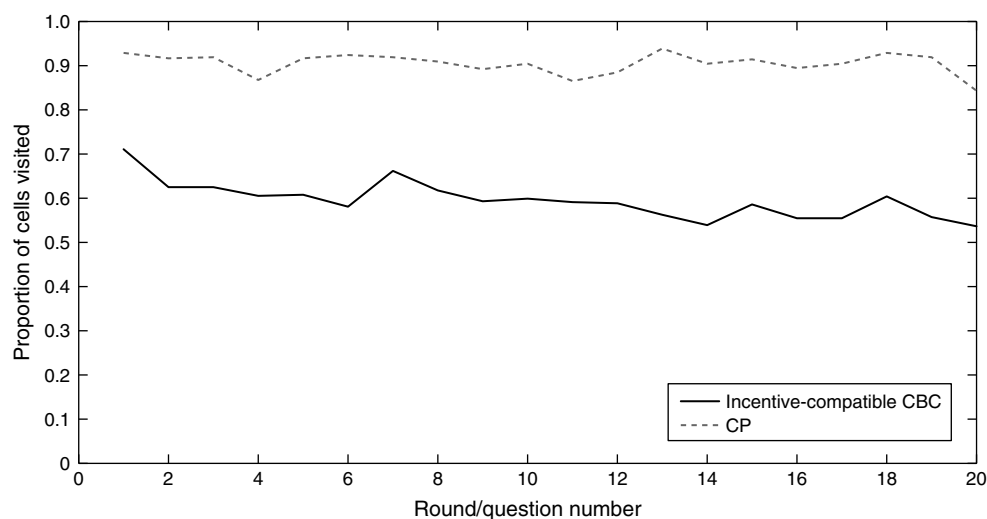
Participants in our eye-tracking studies were recruited at a large European university. The study was run in a special eye-tracking lab, with a free-standing, nonintrusive eyetracker. The eye-tracking apparatus was a Tobii® 2150 tracker, sampling infrared corneal reflections at 50 Hz with a 0.35° spatial resolution and an accuracy of 0.5°. Stimuli were presented on the 21-inch LCD monitor of the eye tracker, controlled by a PC with a display resolution of 1,600 × 1,200 pixels. The position of the participant's left eye and right eye was recorded separately (Van der Lans et al. 2011). After calibration of the eye tracker, respondents participated in the online study on the eye-tracking screen in exactly the same manner as in the main experiment. Each participant completed both CBC and CP in a counterbalanced order (in both Studies 2 and 3, similar results are obtained when using data from the first task only, thereby creating a between-subjects design).

Eye movements primarily consist of fixations and saccades (Wedel and Pieters 2000). Fixations are discrete periods of time (about 200–500 milliseconds) where the eye hardly moves. During this phase, information is extracted. Saccades are quick jumps (20–40 milliseconds) between fixation locations. To identify eye-fixations from the recordings of the point of regard, we used a recently developed velocity-based algorithm (Van der Lans et al. 2011). Each CBC question and each CP round involved four cards with six attributes each. We identified the coordinates of the $6 \times 4 = 24$ cells on the screen that correspond to each of these 24 areas, and we computed the number of fixations in each of these 24 cells.

In this study, the respective designs (i.e., sets of cards) used for CBC and CP were identical to those used in Study 1, and the sample size was $N = 17$. The results confirm that CP participants consider more of the available information compared with incentive-compatible CBC participants. In particular, CP participants considered (i.e., had at least one fixation in the hand selection stage or card selection stage) on average 21.71 of the 24 cells containing the description of the cards (where the average is taken across the 20 rounds and across all participants), whereas incentive-compatible CBC participants considered on average 14.30 of the 24 cells.¹² The difference in

¹² In the hand selection stage of Conjoint Poker only, participants considered on average 20.35 of the 24 cells, and in the card selection

Figure 7 Study 2—Average Proportion of Cells Considered by Participants in Each Condition



Note. Each cell contains the level of one attribute for one alternative.

the average number of cells across respondents is statistically significant (paired t -test, p -value < 0.01), and all respondents considered more cells on average under CP versus CBC. The eye-tracking data also suggest that the amount of information considered by respondents tends to decrease as the study progresses; see Figure 7. A binomial regression with a logistic link of the proportion of cells visited in each question on the question number reveals a significant negative trend for CBC respondents (p -value < 0.01) and a marginally significant negative trend for CP respondents (p -value ≈ 0.08).

We also report the amount of profile-related information considered in the external validity task. With eight profiles, there are $6 \times 8 = 48$ cells. We compute the average number of cells considered by participants in the external validity task the first time they were exposed to the task (irrespective of whether they completed CBC or CP first; the external validity task came before the main task). Participants considered on average 33.24 of the 48 cells. The fact that more than 30% of the cells were not considered on average raises questions on the ability of such external validity tasks to capture real-world decisions. Future research may explore whether our finding replicates in other settings and may possibly develop alternative measures to assess how well various preference

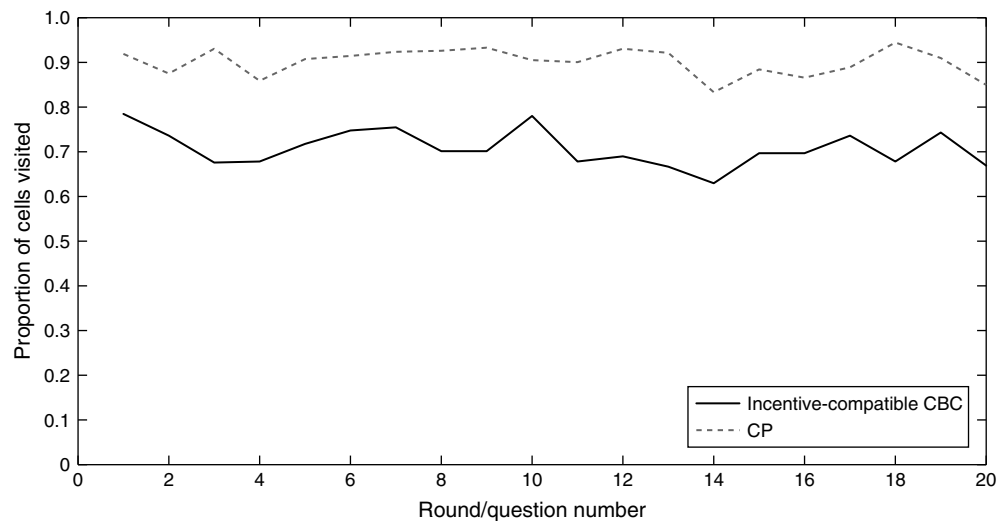
measurement methods explain and predict real-world decisions.

3.4. Study 3

Study 2 provided direct evidence that with the experimental design of Study 1, respondents consider more of the profile-relevant information presented to them under CP versus CBC. However, this finding may also be driven by differences in the designs (i.e., set of cards) used in the two conditions. In particular, the minimum-overlap D-efficient design used in the CBC condition is such that it is very easy for a respondent to choose profiles based on one attribute level (or a very small set of attribute levels). Study 3 addresses this alternative explanation by holding the set of cards constant across the two conditions. In this study, both conditions used the CP design used in Studies 1 and 2. Whereas this design is *statistically* efficient for CP, there is no apparent reason why it should increase the amount of information considered by respondents differently for one method versus the other. Unfortunately, it was not possible to run a study in which both methods used the CBC design from Studies 1 and 2, because this design would lead to all hands in all rounds in CP having the same strength (minimal overlap implies that all hands are double straights). Study 3 was otherwise identical to Study 2, with $N = 18$.

The results confirm that CP participants consider more of the profile-related information compared with incentive-compatible CBC participants, even holding the set of cards constant. In particular, CP participants considered (i.e., had at least one fixation in the hand selection stage or card selection stage) on average 21.63 of the 24 cells containing the description of the cards (where the average is taken across

stage only, they considered on average 11.79 of the 18 cells on the screen (in this stage only three cards remain on the screen). Note that the proportion of cells considered in CBC should not be compared directly to the proportion of cells considered in the card selection stage of Conjoint Poker, because it is likely that respondents remember at least part of the product-related information acquired during the hand selection stage in the card selection stage. In Study 3, Conjoint Poker participants considered on average 20.72 of the 24 cells in the hand selection stage and 11.77 of 18 in the card selection stage.

Figure 8 Study 3—Average Proportion of Cells Visited by Participants in Each Condition

Note. Each cell contains the level of one attribute for one alternative.

the 20 rounds and across all participants), whereas incentive-compatible CBC participants considered on average 16.99 of the 24 cells. The difference in the average number of cells across respondents is statistically significant (paired *t*-test, *p*-value < 0.01), and 17 of the 18 respondents considered more cells on average under CP versus CBC. The downward trend in the amount of information considered by respondents as the questionnaire progresses is also confirmed and is now statistically significant both for CBC (*p*-value < 0.01) and CP (*p*-value < 0.04); see Figure 8. The average number of cells considered by respondents in the external validity task was slightly lower than in Study 2 and equal to 28.56 (out of 48).

In conclusion, Study 3 suggests that the differences in the amount of information considered by respondents, which were observed indirectly in Study 1 and directly in Study 2, were at least partly driven by inherent differences between the methods and not only by differences in designs (i.e., sets of cards) across conditions. The increase in the amount of profile-related information processed by CP participants is at least partly a result of the structure of the task; i.e., CP induces respondents to attend to profile-related information that they may otherwise ignore in a repeated choice task.

4. Conclusions

Marketing researchers have become increasingly interested in improving the quality of preference measurement data. For instance, a stream of research in the survey research literature deals with the identification of various effects that can influence the accuracy of people's responses (see, for example, Baumgartner and Steenkamp 2001, Tourangeau et al. 2000) and with

mechanisms that may improve the accuracy of these responses (see, for example, De Jong et al. 2010, who developed a mechanism for the reporting of sensitive information).

Similar concerns have been voiced in conjoint analysis research. One major recent contribution to the preference measurement literature has been the development of incentive-compatible mechanisms that motivate respondents to reveal their preferences truthfully. Here, we build on that literature to propose an incentive-compatible preference measurement mechanism that provides formal incentives in an engaging environment. Based on an online study and two eye-tracking studies, we provide evidence that this approach increases the amount of profile-related information considered by respondents.

Several areas for future research may be identified. First and as noted above, future research may explore the generalizability and replicability of our finding that incentive-compatible CBC predicts out-of-sample individual choices well but predicts out-of-sample aggregate choice shares poorly. In our study, this result seems to have been linked to the existence of groups of CBC respondents who all gave the exact same answers to all CBC questions, where these answers relied on a small subset of the attributes.¹³

¹³ Consider, for example, the group of 29 CBC respondents who always selected laptops with 320 GB hard drives in all calibration questions. Both external validity profiles with 320 GB hard drives are attractive to this group of consumers. Because all of these respondents gave the exact same set of answers to all calibration questions, the set of predicted out-of-sample choice probabilities is identical for all these respondents. Moreover, our ability to predict which of these two 320 GB laptops will be chosen by these respondents in the external validity task relies only on the use of shrinkage. In other words, the out-of-sample predictions within the

Second, additional eye-tracking studies may be run to explore whether alternative conjoint analysis formats, such as rating-based conjoint analysis, are able to increase information consideration. Third, future research may identify situations in which deeper information processing in preference measurement tasks is actually *not* desirable, e.g., because it induces respondents to pay more attention to information that they would ignore when making real-life decisions. A fourth area for future research is the development and testing of multiplayer versions of Conjoint Poker. This would raise several issues, including (a) players may be able to learn each other's preferences; (b) players may be able to learn each other's playing style and, in particular, the propensity to play the strongest hand; (c) the assumptions made in Appendix A that the opponent's cards are drawn randomly and that the opponent always plays the strongest available hand would not hold anymore, and the computation of the probabilities of winning with each hand would be more complicated; and (d) the efficiency of each player's design would be influenced by each other player's design (through the probabilities of winning), and therefore designs should be optimized jointly across players.

Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at <http://mktsci.journal.informs.org/>.

Acknowledgments

The authors thank Ralf van der Lans for his help with the analysis of eye-tracking data. M. G. de Jong thanks the Niels Stensen Foundation and the Netherlands Organization for Scientific Research for financial support.

Appendix A. Hand Probabilities

A.1. Hand Strengths

We show here how to compute the probabilities that each hand is achievable from a random set of four cards drawn uniformly without replacement from the set of all possible cards in the case of $A = 6$ attributes with four levels each. The same approach may be used for any other number of attributes and levels. These probabilities capture the probability that each hand is achievable by the computer in our studies (against which our respondents played) and will be used next to compute the probability of winning with each hand when playing against the computer.

small set of profiles with 320 GB are “random guesses” driven only by population-level shrinkage. Because there are only two profiles with 320 GB, hit rates among these respondents are high (~50%). However, aggregate choice share predictions among these respondents are less accurate if the relative shares of these two profiles are not well predicted.

Let b equal to the probability that four random cards have at least three different levels on a given attribute: $b = \binom{4}{2} \cdot 4 \cdot 3 \cdot 2 / 4^4 + 4! / 4^4 = 42/64$.

Let c be the probability that at least three of four random cards have the same level on a given attribute: $c = (4 \cdot 4 \cdot 3) / 4^4 + 4 / 4^4 = 13/64$.

Let $d = 1 - b - c$ be the probability that four random cards are such that two have one level and two have another level on a given attribute: $d = (4 \cdot 3 \cdot 3) / 4^4 = 9/64$.

Let e the probability that exactly three of four random cards have the same level on a given attribute given that they have at least three: $e = ((4 \cdot 4 \cdot 3) / 4^4) / (13/64) = 12/13$.

Let f be the probability that exactly three of four random cards have different levels on a given attribute given that at least three do: $f = ((\binom{4}{2} \cdot 4 \cdot 3 \cdot 2) / 4^4) / (42/64) = 36/42$.

We have the following:

Pr(Double Flush)

$$= \sum_{k_2=2}^A \binom{A}{k_2} (1-c)^{A-k_2} c^{k_2} \left(1(k_2 > 4) + 1(k_2 \leq 4) \cdot \left(1 - e^{k_2} \frac{3!}{(4-k_2)!} \left(\frac{1}{4} \right)^{k_2-1} \right) \right).$$

Pr(Straight Flush)

$$= \sum_{k_2=1}^{A-1} \sum_{k_1=1}^{A-k_2} \binom{A}{k_1} \cdot \binom{A-k_1}{k_2} d^{A-k_1-k_2} b^{k_1} c^{k_2} \cdot \left[1 - f^{k_1} e^{k_2} \left(\left(\frac{1}{2} \right)^{k_1} \left(\frac{1}{4} \right)^{k_2-1} + \left(\frac{4}{2} \right) \left(\frac{1}{2} \right)^{k_2} \cdot \left(1 - \left(\frac{1}{2} \right)^{k_2-1} \right) \left(\frac{1}{6} \right)^{k_1} \right] \right].$$

$$\text{Pr(Flush)} = 1 - (1-c)^A.$$

Pr(Double Straight)

$$= \sum_{k_1=2}^A \binom{A}{k_1} \cdot (1-b)^{A-k_1} b^{k_1} \cdot \left[1(k_1 \geq 3) + 1(k_1 \leq 2) \cdot \left(1 - f^{k_1} \left(\frac{1}{6} \right)^{k_1-1} \right) \right].$$

$$\text{Pr(Straight)} = 1 - (1-b)^A.$$

$$\text{Pr(Pair)} = 1 - (b(1-f))^A.$$

A.2. Probabilities of Winning Against the Computer

For each round and for each possible hand, we develop closed-form expressions for the probability of winning if playing that hand against a player with random cards who always plays the best hand. We compute Pr(best hand is Double Flush), Pr(best hand is Straight Flush), Pr(best hand is Flush), Pr(best hand is Double Straight), Pr(best hand is Straight), and Pr(best hand is Pair). (It is easy to show that Pr(not getting anything) = 0.) Then the probability of

winning with a given hand is the sum of probabilities that the opponent will get a weaker hand plus half the probability that the opponent will get a hand of similar strength.

We have the following:

Pr(best hand is Double Flush)

$$= \sum_{k_2=2}^A \binom{A}{k_2} (1-c)^{A-k_2} c^{k_2} \left(1(k_2 > 4) + 1(k_2 \leq 4) \cdot \left(1 - e^{k_2} \frac{3!}{(4-k_2)!} \left(\frac{1}{4} \right)^{k_2-1} \right) \right).$$

Pr(best hand is Straight Flush)

= Pr(Straight Flush, no Double Flush)

$$= \sum_{k_2=1}^4 \sum_{k_1=1}^{A-k_2} \binom{A}{k_1} \binom{A-k_1}{k_2} d^{A-k_1-k_2} b^{k_1} c^{k_2} \cdot \left[1(k_2 = 1) + 1(k_2 > 1) e^{k_2} \frac{3!}{(4-k_2)!} \left(\frac{1}{4} \right)^{k_2-1} \right] \cdot \left[1(k_2 > 2) + 1(k_2 \leq 2) \left(1 - f^{k_1} \left(\frac{\binom{4-k_2}{2}}{6} \right)^{k_1} \right) \right].$$

Pr(best hand is Flush)

= Pr(Flush, no Straight Flush, no Double Flush)

$$= \sum_{k_2=1}^4 \sum_{k_1=0}^{A-k_2} \binom{A}{k_1} \binom{A-k_1}{k_2} d^{A-k_1-k_2} b^{k_1} c^{k_2} \cdot \left[1(k_2 = 1) + 1(k_2 > 1) e^{k_2} \frac{3!}{(4-k_2)!} \left(\frac{1}{4} \right)^{k_2-1} \right] \cdot \left[1(k_1 = 0) + 1(k_1 > 0) \cdot 1(k_2 \leq 2) \cdot f^{k_1} \left(\frac{\binom{4-k_2}{2}}{6} \right)^{k_1} \right].$$

Pr(best hand is Double Straight)

= Pr(Double Straight, no Flush,
no Straight Flush, no Double Flush)

$$= \sum_{k_1=2}^A \binom{A}{k_1} d^{A-k_1} b^{k_1} \left[1(k_1 \geq 3) + 1(k_1 = 2) \cdot \frac{5}{6} \right].$$

Pr(best hand is Straight)

= Pr(Straight, no Double Straight, no Flush,
no Straight Flush, no Double Flush)

$$= \sum_{k_1=1}^2 \binom{6}{k_1} d^{6-k_1} b^{k_1} \left[1(k_1 = 1) + 1(k_1 = 2) \cdot \frac{1}{6} \right].$$

Pr(best hand is Pair)

= Pr(Pair, no Straight, no Double Straight,
no Flush, no Straight Flush, no Double Flush)

$$= d^A.$$

Appendix B. Computation of the Information Matrix

B.1. Homogeneous Case

The log of the likelihood corresponding to a given round r in which the player played h and chose profile j is

$$L_r = \delta V_h^r - \log \left(\sum_{h'} \exp(\delta V_{h'}^r) \right) + x_j \beta - \log \left(\sum_{j' \in h} \exp(x_{j'} \beta) \right),$$

$$\nabla L_r = \delta \nabla V_h^r - \frac{\sum_{h'} \exp(\delta V_{h'}^r) \delta \nabla V_{h'}^r}{\sum_{h'} \exp(\delta V_{h'}^r)} + x_j^T - \frac{\sum_{j' \in h} \exp(x_{j'} \beta) x_{j'}^T}{\sum_{j' \in h} \exp(x_{j'} \beta)}$$

$$= \delta \nabla V_h^r - \sum_{h'} \Pr(h') \delta \nabla V_{h'}^r + x_j^T - \sum_{j' \in h} \Pr(j') x_{j'}^T.$$

The information matrix for round r is given by

$$\Omega_r = E_{h_r, j} ((\nabla L_r)(\nabla L_r)^T)$$

$$= \sum_{h_r} \sum_{j \in h_r} \left[\delta \nabla V_{h_r}^r - \sum_{h'_r} \Pr(h'_r) \delta \nabla V_{h'_r}^r + x_j^T - \sum_{j' \in h_r} \Pr(j') x_{j'}^T \right] \cdot \left[\delta \nabla V_{h_r}^r - \sum_{h'_r} \Pr(h'_r) \delta \nabla V_{h'_r}^r + x_j^T - \sum_{j' \in h_r} \Pr(j') x_{j'}^T \right]^T \cdot \Pr(h_r) \Pr(j | h_r)$$

$$= \sum_{h_r} \left[\delta \nabla V_{h_r}^r - \sum_{h'_r} \Pr(h'_r) \delta \nabla V_{h'_r}^r \right] \cdot \Pr(h_r) \left[\delta \nabla V_{h_r}^r - \sum_{h'_r} \Pr(h'_r) \delta \nabla V_{h'_r}^r \right]^T \sum_{j \in h_r} \Pr(j | h_r)$$

$$+ \sum_{h_r} \Pr(h_r) \sum_{j \in h_r} \left[x_j^T - \sum_{j' \in h_r} \Pr(j' | h_r) x_{j'}^T \right] \cdot \Pr(j | h_r) \left[x_j^T - \sum_{j' \in h_r} \Pr(j' | h_r) x_{j'}^T \right]^T$$

$$+ \sum_{h_r} \left[\delta \nabla V_{h_r}^r - \sum_{h'_r} \Pr(h'_r) \delta \nabla V_{h'_r}^r \right] \cdot \Pr(h_r) \sum_{j \in h_r} \left[x_j^T - \sum_{j' \in h_r} \Pr(j' | h_r) x_{j'}^T \right]^T \Pr(j | h_r)$$

$$= \sum_{h_r \in r} (\tilde{z}_{h_r}^r \Pr(h_r) \tilde{z}_{h_r}^{rT} + \Pr(h_r) \sum_{j \in h_r} z_{j|h_r} \Pr(j | h_r) z_{j|h_r}^T),$$

where

$$\tilde{z}_{h_r}^r = \delta \nabla V_{h_r}^r - \sum_{h'_r} \delta \nabla V_{h'_r}^r \Pr(h'_r) \quad \text{and}$$

$$z_{j|h_r} = x_j^T - \sum_{j' \in h_r} x_{j'}^T \Pr(j' | h_r).$$

B.2. Mixed Logit Case

In the case of a mixed logit model, following Sandor and Wedel (2002), we assume that $\beta_i = \mu + U_i \sigma$, where σ is a diagonal matrix and $U_i = \text{diag}(u_i)$, where $u_i \sim N(0, I)$. Let

$$\pi_{j|h} = \int_u \Pr(j | h, u) \Phi(u) du \quad \text{and}$$

$$\pi_h = \int_u \Pr(h|u) \Phi(u) du,$$

where Φ is the probability density function of u .

The log of the likelihood corresponding to a given round r in which the player chose hand h and profile j may be written as

$$L_r = \log(\pi_h) + \log(\pi_{j|h}),$$

$$\begin{aligned} \frac{\partial L_r}{\partial \mu} &= \frac{1}{\pi_h} \frac{\partial \pi_h}{\partial \mu} + \frac{1}{\pi_{j|h}} \frac{\partial \pi_{j|h}}{\partial \mu} \\ &= \frac{1}{\pi_h} \int_u \frac{\partial \Pr(h|u)}{\partial \mu} \Phi(u) du \\ &\quad + \frac{1}{\pi_{j|h}} \int_u \frac{\partial \Pr(j|h, u)}{\partial \mu} \Phi(u) du \\ &= \frac{1}{\pi_h} \int_u \left[(\delta \nabla V_h^r \Pr(h|u) \right. \\ &\quad \left. - \sum_{h'} \Pr(h'|u) \Pr(h|u) \delta \nabla V_{h'}^r \right] \Phi(u) du \\ &\quad + \frac{1}{\pi_{j|h}} \int_u U \left[(x_j^T \Pr(j|h, u) \right. \\ &\quad \left. - \sum_{j' \in h} \Pr(j'|h, u) x_{j'}^T \right] \Phi(u) du, \\ \frac{\partial L_r}{\partial \sigma} &= \frac{1}{\pi_h} \frac{\partial \pi_h}{\partial \sigma} + \frac{1}{\pi_{j|h}} \frac{\partial \pi_{j|h}}{\partial \sigma} \\ &= \frac{1}{\pi_h} \int_u \frac{\partial \Pr(h|u)}{\partial \sigma} \Phi(u) du \\ &\quad + \frac{1}{\pi_{j|h}} \int_u \frac{\partial \Pr(j|h, u)}{\partial \sigma} \Phi(u) du \\ &= \frac{1}{\pi_h} \int_u U \left[(\delta \nabla V_h^r \Pr(h|u) \right. \\ &\quad \left. - \sum_{h'} \Pr(h'|u) \Pr(h|u) \delta \nabla V_{h'}^r \right] \Phi(u) du \\ &\quad + \frac{1}{\pi_{j|h}} \int_u U \left[(x_j^T \Pr(j|h, u) - \sum_{j' \in h} \Pr(j'|h, u) \right. \\ &\quad \left. \cdot \Pr(j'|h, u) x_{j'}^T \right] \Phi(u) du, \end{aligned}$$

$$\Omega_r = \begin{bmatrix} E_{h_r, j} \left(\left(\frac{\partial L_r}{\partial \mu} \right) \left(\frac{\partial L_r}{\partial \mu} \right)^T \right) & E_{h_r, j} \left(\left(\frac{\partial L_r}{\partial \mu} \right) \left(\frac{\partial L_r}{\partial \sigma} \right)^T \right) \\ E_{h_r, j} \left(\left(\frac{\partial L_r}{\partial \sigma} \right) \left(\frac{\partial L_r}{\partial \mu} \right)^T \right) & E_{h_r, j} \left(\left(\frac{\partial L_r}{\partial \sigma} \right) \left(\frac{\partial L_r}{\partial \sigma} \right)^T \right) \end{bmatrix},$$

where

$$\begin{aligned} E_{h_r, j} \left(\left(\frac{\partial L_r}{\partial \mu} \right) \left(\frac{\partial L_r}{\partial \mu} \right)^T \right) &= \sum_{h_r} \sum_{j \in h_r} \left(\frac{\partial L_r}{\partial \mu} \right) \left(\frac{\partial L_r}{\partial \mu} \right)^T \pi_{h_r} \pi_{j|h_r}, \\ E_{h_r, j} \left(\left(\frac{\partial L_r}{\partial \sigma} \right) \left(\frac{\partial L_r}{\partial \mu} \right)^T \right) &= \sum_{h_r} \sum_{j \in h_r} \left(\frac{\partial L_r}{\partial \sigma} \right) \left(\frac{\partial L_r}{\partial \mu} \right)^T \pi_{h_r} \pi_{j|h_r}, \\ E_{h_r, j} \left(\left(\frac{\partial L_r}{\partial \sigma} \right) \left(\frac{\partial L_r}{\partial \sigma} \right)^T \right) &= \sum_{h_r} \sum_{j \in h_r} \left(\frac{\partial L_r}{\partial \sigma} \right) \left(\frac{\partial L_r}{\partial \sigma} \right)^T \pi_{h_r} \pi_{j|h_r}. \end{aligned}$$

Appendix C. Hierarchical Bayes Estimation

The likelihood function for Conjoint Poker is given by Equations (2) and (4). The likelihood function for CBC is given simply by

$$\Pr(j) = \frac{\exp(x_j \beta_i)}{\sum_{j' \in r} \exp(x_{j'} \beta_i)}.$$

Let p be equal to the number of partworths estimated ($p = 18$ in our case), and let I_p denote the identity matrix of size p . The same prior specification was used in both conditions:¹⁴

$$\beta_i \sim N(\beta_0, D_\beta),$$

$$D_\beta^{-1} \sim \text{Wishart}(0.01 I_p, p + 15),$$

diffuse improper prior on β_0 .

In addition, we used the following priors on the additional parameters in Conjoint Poker:

$$\sigma_i \sim \text{TN}(\sigma_0, D_\sigma) \text{ (truncated between 0 and 2),}$$

$$\alpha_i \sim N(\alpha_0, D_\alpha),$$

$$D_\sigma^{-1} \sim \text{Gamma}(1, 1),$$

$$D_\alpha^{-1} \sim \text{Gamma}(1, 1),$$

diffuse improper prior for σ_0 and α_0 ,

diffuse improper prior on \mathbb{R}^+ for δ .

We used MCMC with 50,000 iterations, using the first 10,000 as burn-in and saving 1 in every 10 iterations thereafter. Convergence was assessed visually using plots of the parameters. We used rejection sampling on α_i to ensure that $\alpha_i + \max_{j \in h} \{x_j \beta_i\} \geq 0$ for all i and h .

Appendix D. Poker Quiz

Please answer the following questions regarding real poker.

The *pot* is

- a set of cards placed in the middle of the table.
- an amount of money for which players compete.
- an amount of money that players may borrow during the game.

To *call* is

- to discard one's hand and forfeit interest in the current pot.

- to increase the size of the bet required to stay in the pot.

- to match a bet or match a raise.

To *fold* is

- to discard one's hand and forfeit interest in the current pot.

- to increase the size of the bet required to stay in the pot.

- to match a bet or match a raise.

An *ante* is

- a forced bet in which all players put an equal amount of money or chips into the pot before the deal begins.

¹⁴Of course, each condition was estimated separately and independently.

- the final community card.
- the first three community cards.

A *flop* is

- a forced bet in which all players put an equal amount of money or chips into the pot before the deal begins.

- the final community card.
- the first three community cards.

A *river* is

- a forced bet in which all players put an equal amount of money or chips into the pot before the deal begins.
- the final community card.
- the first three community cards.

References

- Baumgartner, H., J.-B. E. M. Steenkamp. 2001. Response styles in marketing research: A cross-national investigation. *J. Marketing Res.* **38**(2) 143–156.
- Bradlow, E. T. 2005. Current issues and a wishlist for conjoint analysis. *Appl. Stochastic Models Bus. Indust.* **21**(4–5) 319–323.
- Bunch, D. S., J. J. Louviere, D. Anderson. 1994. A comparison of experimental design strategies for multinomial logit models: The case of generic attributes. Working paper, Graduate School of Management, University of California, Davis, Davis.
- Chandon, P., J. W. Hutchinson, E. T. Bradlow, S. H. Young. 2009. Does in-store marketing work? Effects of the number and position of shelf facings on brand attention and evaluation at the point of purchase. *J. Marketing* **73**(November) 1–17.
- De Jong, M. G., R. Pieters, J.-P. Fox. 2010. Reducing social desirability bias via item randomized response: An application to measure underreported desires. *J. Marketing Res.* **47**(February) 14–27.
- Ding, M. 2007. An incentive-aligned mechanism for conjoint analysis. *J. Marketing Res.* **44**(1) 214–223.
- Ding, M., R. Grewal, J. Liechty. 2005. Incentive-aligned conjoint analysis. *J. Marketing Res.* **42**(1) 67–82.
- Ding, M., Y.-H. Park, E. T. Bradlow. 2009. Barter markets for conjoint analysis. *Management Sci.* **55**(6) 1003–1017.
- Dong, S., M. Ding, J. Huber. 2010. A simple mechanism to incentive align conjoint experiments. *Internat. J. Res. Marketing* **27**(1) 25–32.
- Green, P. E., V. Srinivasan. 1990. Conjoint analysis in marketing: New developments with implications for research and practice. *J. Marketing* **54**(4) 3–19.
- Huber, J., K. Zwerina. 1996. The importance of utility balance in efficient choice designs. *J. Marketing Res.* **33**(3) 307–317.
- Johnson, R. 2008. A perspective on adaptive CBC (what can we expect from respondents?). Research paper, Sawtooth Software, Sequim, WA. <http://www.sawtoothsoftware.com/download/techpap/acbcperspective.pdf>.
- Kohli, R., R. Sukumar. 1990. Heuristics for product-line design using conjoint analysis. *Management Sci.* **36**(12) 1464–1478.
- Lenk, P. J., W. S. DeSarbo, P. E. Green, M. R. Young. 1996. Hierarchical bayes conjoint analysis: Recovery of partworth heterogeneity from reduced experimental designs. *Marketing Sci.* **15**(1) 173–191.
- Liechty, J. C., D. K. H. Fong, W. DeSarbo. 2005. Dynamic models with individual level heterogeneity: Applied to evolution during conjoint studies. *Marketing Sci.* **24**(1) 285–293.
- Mahajan, V., P. E. Green, S. M. Goldberg. 1982. A conjoint model for measuring self- and cross-price/demand relationships. *J. Marketing Res.* **19**(3) 334–342.
- McFadden, D. 1974. Conditional logit analysis of qualitative choice behavior. P. Zarembka, ed. *Frontiers in Econometrics*. Academic Press, New York, 105–142.
- Netzer, O., O. Toubia, E. T. Bradlow, E. Dahan, T. Evgeniou, F. M. Feinberg, E. M. Feit et al. 2008. Beyond conjoint analysis: Advances in preference measurement. *Marketing Lett.* **19**(3–4) 337–354.
- Newey, W. K., D. McFadden. 1994. Large sample estimation and hypothesis testing. R. F. Engle, D. L. McFadden, eds. *Handbook of Econometrics*, Vol. 4. Elsevier Science, Amsterdam, 2113–2247.
- Park, Y.-H., M. Ding, V. R. Rao. 2008. Eliciting preference for complex products: A Web-based upgrading method. *J. Marketing Res.* **45**(5) 562–574.
- Pieters, R., M. Wedel. 2004. Attention capture and transfer in advertising: Brand, pictorial, and text-size effects. *J. Marketing* **68**(April) 36–50.
- Rossi, P. E., G. M. Allenby. 2003. Bayesian statistics and marketing. *Marketing Sci.* **22**(3) 304–328.
- Sandor, Z., M. Wedel. 2001. Designing conjoint choice experiments using managers' prior beliefs. *J. Marketing Res.* **38**(4) 430–444.
- Sandor, Z., M. Wedel. 2002. Profile construction in experimental choice designs for mixed logit models. *Marketing Sci.* **21**(4) 455–475.
- Sandor, Z., M. Wedel. 2005. Heterogeneous conjoint choice designs. *J. Marketing Res.* **42**(1) 210–218.
- Starmer, C., R. Sugden. 1991. Does the random-lottery incentive system elicit true preferences? An experimental investigation. *Amer. Econom. Rev.* **81**(4) 971–978.
- Thaler, R. H., E. J. Johnson. 1990. Gambling with the house money and trying to break even: The effects of prior outcomes on risky choices. *Management Sci.* **36**(6) 643–660.
- Tourangeau, R., L. J. Rips, K. Rasinski. 2000. *The Psychology of Survey Response*. Cambridge University Press, Cambridge, UK.
- Van der Lans, R., M. Wedel, R. Pieters. 2011. Defining eye-fixation sequences across individuals and tasks: The binocular-individual threshold (BIT) algorithm. *Behav. Res. Methods* **43**(1) 239–257.
- Wedel, M., R. Pieters. 2000. Eye fixations on advertisements and memory for brands: A model and findings. *Marketing Sci.* **19**(4) 297–312.
- Wittink, D. R., P. Cattin. 1989. Commercial use of conjoint analysis: An update. *J. Marketing Res.* **53**(3) 91–96.