An Investigation of the Performance of Competitive Incentive Schemes and Lottery Incentive Schemes vis-à-vis Fixed Fee Incentive Schemes in Improving Conjoint Analysis

Srikant Vadali
June 2009
July 2009
January 2010

Srikant Vadali is an Assistant Professor in the College of Management at The University of Massachusetts, Lowell, He can be reached via email at Srikant_Vadali@uml.edu, Tel: (978) 934 2810 and Fax: (978) 934 2064. The author thanks Seenu Srinivasan, Kenneth Train, Arvind Rangaswamy, Michel Clement and Krishnaraj Venkataraman for their helpful comments on a previous version of this manuscript and acknowledges financial support from the College of Management, The University of Massachusetts, Lowell.

An Investigation of the Performance of Competitive Incentive Schemes and Lottery Incentive Schemes vis-à-vis Fixed Fee Incentive Schemes in Improving Conjoint Analysis

Abstract

Paying a fixed amount of money to participants in choice based conjoint (CBC) studies is the industry standard. Recently, Ding (2007) has shown that a lottery incentive scheme outperformed a fixed fee incentive scheme when predicting out-of-sample choices.

We achieve two research goals in the current paper to extend our understanding of incentive schemes in the context of CBC studies. One, we investigate if a higher fixed fee (e.g., $50 instead of $10) helps improve out-of-sample predictions. Two, the lottery incentive scheme does not induce competition among CBC study participants. Therefore, we investigate the theoretical properties and empirical effectiveness of competitive incentive schemes relative to lottery and fixed incentive schemes.

Our key findings with respect to hit rates for out-of-sample predictions are: (a) offering higher amounts of money is ineffective, and (b) competitive incentive schemes outperform the lottery incentive scheme (Hit Rates of 41 % and 62% for the two proposed competitive schemes vs. 29% for the lottery incentive scheme).

Keywords: Conjoint Analysis, Incentive Schemes, Experiments, HB Estimation, Multinomial Logit

1. Introduction

Choice based conjoint (hereafter referred to as CBC) is one of the main quantitative market research techniques used by firms to identify promising new product designs, segment markets, decide prices etc. In a typical CBC study, consumers are asked several choice questions. Each choice question shows several potential product designs and they are asked to choose the design that they would purchase if they had to buy one at that point in time. The answers to these choice questions are then used to estimate each consumer's utility function which then serves as an input to decisions regarding segmentation, pricing, identifying promising product designs etc. Following the literature in conjoint analysis, we will henceforth refer to utility function estimates as partworths.

A CBC study can be divided into three stages each of which offer opportunities to improve our ability to accurately estimate partworths. The three stages are: (a) Design of the CBC study, (b) Data Collection from consumers and (c) Estimation of partworths. In the design stage, decisions are taken about how many attributes and how many levels for each attribute need to be included in the study, how questions are generated (e.g., fixed ahead of time or generated on the fly), number of questions to ask etc. After the design is finalized, data is actually collected from consumers and after data collection partworths are estimated. Considerable research exists that demonstrates different ways by which we can improve partworth recovery at the design, data collection and estimation stages. For example, for the design stage, there is extensive literature on question design that investigates the impact of alternative questions selection mechanisms on partworth recovery (Louviere et al. 2008;Toubia, Hauser, and Garcia 2007;Yu, Goos, and

Vandebroek 2009). At the data collection stage, research has examined alternative approaches to engage consumers in the CBC study so that consumers make accurate choices which would enhance data quality thereby resulting in accurate partworth estimation (Dahan, Soukhoroukova, and Spann 2007;Ding 2007; Ding, M., Park, YH, and E. Bradlow 2009; Park, YH, M. Ding and V. Rao 2008). Similarly, for the estimation stage, there is literature that investigates the impact of alternative estimation methods on partworth recovery (Allenby et al. 2005;Evgeniou, Pontil, and Toubia 2007;Liu, Otter, and Allenby 2007). See Netzer et al. (2008) and Rao (2008) for an overview of state-of-the-art in conjoint analysis and important areas of research in this area. In this paper, we investigate the effectiveness of competitive incentive schemes to improve the quality of data collected by examining their impact on our ability to recover partworths accurately. Thus, our paper falls in the stream of literature that attempts to improve partworth recovery by improving the quality of the data that is collected from consumers.

Offering a fixed amount of money (hereafter referred to as the Flat Fee Scheme) to compensate consumers for the time and effort it takes to answer choice questions is the typical incentive scheme used in CBC studies. Recently, Ding (2007), henceforth referred to as the Ding study, investigated the impact of a lottery incentive scheme (hereafter referred to as the Product Incentive Scheme) on partworth recovery in a CBC study involving the design of an iPod package. The product incentive scheme is a lottery incentive scheme as one randomly selected participant is offered additional compensation. The additional compensation is structured such that study respondents are motivated to answer choice questions truthfully. In light of the above, it is not surprising that the product incentive scheme is very effective in improving partworth recovery. Specifically, the iPod experiment in the Ding study showed that hit rates for the hold-out task improved from 17 percent for the flat fee incentive scheme to 36 percent for the product incentive scheme[1].

Answering choice questions consistently and truthfully for the entire duration of a CBC study requires effort. In the Ding study, participants were paid $10 which is a relatively low level of compensation. A natural question that springs to mind is: Do sufficiently powerful flat fee incentive schemes motivate consumers to exert effort and answer CBC questions accurately? For example, will study participants answer choice questions consistent with their true preferences if they are offered $50 instead of $10?

On the basis of utility maximization one would predict that flat fee compensation schemes will not work irrespective of the amount of fee paid to respondents. However, there is considerable research in experimental economics that indicates that consumers often take decisions motivated by a sense of fairness, justice etc and do not always follow the principle of utility maximization. For example, consider the stream of research involving the ultimatum game. In the ultimatum game, two players need to split a certain amount of money as follows. Player 1 decides what percentage of the available money should go to player 2. Player 2 then decides whether to reject or accept player 1's proposal. If player 2 rejects the proposal then neither player gets anything. On other hand, if player 2 accepts the proposal then the money is divided as proposed by player 1. Economic theory predicts that player 1 should offer the least

[1] The Ding study performed two experiments using different versions of the iPod: iPod Shuffle and iPod Nano. Due to space constraints, we focus only on experiment 1 in this paper.

amount of money to player 2 (say one cent) and player 2 should accept the proposal as that is better than receiving nothing. However, a meta-analysis of 75 ultimatum game experiments indicated that the player 1's average proposal was to offer 40 % of the pie and that 16 % of the offers are rejected (Oosterbeek, Sloof, and Kuilen 2004). 'A sense of fairness' in players is one reason that is commonly offered for the above deviation from the predictions of economic theory. In other words, proposing players do not offer one cent as they believe that such an offer is not a fair division of the available money and players reject proposals that are too low as they consider it as an unfair proposal. Along the same lines, intuition suggests that if consumers are informed that they will be paid $50 so that they are adequately compensated for the time and effort it takes to answer CBC questions accurately, a sense of fairness may motivate them to answer CBC questions according to their true preferences. In light of the above, an understanding of how consumers respond to powerful flat fee incentives in the context of CBC studies is important from a theoretical perspective. Therefore, our first research goal is to investigate the effectiveness of a Strong Flat Fee Incentive scheme (e.g., $50) vis-a-vis a Weak Flat Fee Incentive scheme (e.g., $15).

The product incentive scheme offers one approach to motivate consumers to exert the required effort to answer choice questions truthfully. An alternative way to motivate consumers to exert effort is to induce competition among them such that exerting effort increases the chances that they win an attractive prize. Therefore, out second research goal of this paper is to investigate the theoretical properties and empirical effectiveness of incentive schemes that induce competition among study participants in a CBC study. Specifically, we focus on two incentive schemes: the award incentive scheme and the hybrid incentive scheme. In the award incentive scheme, respondents answer a certain number of choice questions (e.g., 24). These choice questions are divided into two sets: an estimation set (e.g., 16 questions chosen at random) and a prediction set (e.g., the remaining 8 questions). Answers to the estimation set are used to estimate partworths for all respondents and these estimated partworths are used to predict their choices to the choice questions in the prediction set. The respondent whose choices in the prediction set we are able to predict the best wins a cash award. In contrast to the product incentive scheme, the award incentive scheme is a competitive scheme as the structure of the scheme induces competition among the study participants. In the hybrid incentive scheme, we incorporate features of both the product incentive scheme and the award incentive scheme. In other words, the scheme is structured such that study participants are motivated to answer choice questions truthfully and the reward they obtain at the end of the study is contingent on the extent to which other study participants provide consistent answers.

In order to understand the theoretical properties of the award incentive scheme and the hybrid incentive schemes we model the strategic behavior of participants using game theory and characterize the Bayesian Nash Equilibrium of the resulting game. Using digital cameras as a context, we conducted five between-subject experiments in order to evaluate the empirical effectiveness of the five incentive schemes (i.e., the Strong Flat Fee, the Product Incentive, the Award Incentive, the Hybrid Incentive and the Weak Flat Fee). In each experiment, participants answered 24 choice questions and subsequently answered a hold-out choice question. Each choice question showed four alternatives (three digital cameras and a ‘None of these’) whereas the holdout showed 17 alternatives (sixteen digital cameras and a 'None of these'). We assessed the relative effectiveness of the five incentive schemes by computing hit rates for the hold out task using partworth estimates from the answers to the 24 choice questions. Our theoretical and empirical findings are summarized below.

From a theoretical perspective, truth-telling is not the only Bayesian Nash equilibrium for the award incentive. Any strategy that requires study participants to be consistent with an arbitrary preference structure is an equilibrium strategy. Truth-telling is the only Bayesian Nash equilibrium for the hybrid incentive scheme. From an empirical perspective we find that:

1. The Strong Flat Fee does *not* outperform the Weak Flat Fee.
2. The Product Incentive outperforms the Weak Flat Fee by a 2 to 1 margin (Hit Rates: 29% vs. 14%) and
3. The Award Incentive Scheme outperforms the Product Incentive Scheme by a 2 to 1 margin (Hit Rates: 62% vs. 29%) thus outperforming the Weak Flat Fee by a 4 to 1 margin (Hit Rates: 62% vs. 14%).
4. The Hybrid Incentive scheme outperforms the Product Incentive Scheme by a 1.4 to 1 margin (Hit Rates: 41% vs. 29%) thus outperforming the Weak Flat Fee by a 2.9 to 1 margin (Hit Rates: 41% vs. 14%).

We make three observations about the above findings at this point and defer a detailed discussion of these findings when we discuss our results. One, the presence of multiple equilibria including truth-telling in the award incentive is not surprising as the award incentive has no mechanism to induce truth telling. Consequently, we designed our experiment such that respondents did not have sufficient time or knowledge to resort to any other strategy except truth-telling. We elaborate on this issue where we discuss the design decisions for the experiments we conducted. Two, since the hybrid incentive scheme incorporates features of the product incentive scheme, it is not surprising that truth telling is the unique equilibrium. Three, the lack of superiority of the strong flat fee is somewhat surprising. Contrary to our intuition, but consistent with economic theory, respondents in the strong flat fee experiment did not answer CBC questions any more accurately than respondents in the weak flat fee experiment. The inferiority of the strong flat fee suggests that even sufficiently powerful fixed fee schemes do not motivate respondents to answer CBC questions accurately. Four, our findings regarding the superior performance of competitive incentive schemes indicate that we must carefully design compensation schemes in order to motivate respondents to the maximum.

In summary, our paper contributes to the growing literature[2] on the use of incentive schemes to encourage truth-telling in conjoint studies by establishing the superiority of a competitive incentive schemes (i.e., the award incentive and the hybrid incentive) relative to a lottery incentive scheme (i.e., the product incentive). Our results also demonstrate the robustness of the lottery incentive scheme introduced in the Ding study for a different product category (i.e., iPods in the Ding study and Digital cameras in our paper). Finally, from a theoretical perspective we show that offering higher fixed fee compensation does not improve partworth recovery.

[2] See Dong, S., M. Ding and J. Huber 2009; Park, YH, M. Ding and V. Rao 2008; Ding, Min 2007; Ding, M., Grewal,R., and J. Liechty 2005 for recent research that looks at incentive schemes in the context of conjoint analysis.

These results extend our understanding of the effectiveness of different incentive structures (i.e., competitive, lottery and fixed fee) in the context of conjoint studies.

Before we discuss our results, we describe the five incentive schemes in Section 2 and establish the theoretical properties of the award and the hybrid incentive schemes. In Section 3, we describe the implementation details of the five CBC studies to establish experimental equivalence so that we can rule out alternative explanations for our findings. In Section 5, using our description of the CBC studies from Section 3 and the findings from Section 4, we identify issues that need further research and help provide direction to practitioners and academic researchers. We start by describing all five incentive schemes.

2. The Five Incentive Schemes

The Ding study investigated the effectiveness of the product incentive scheme in the context of the design of an iPod package which consisted of an iPod shuffle with some combination of associated accessories. A potential package comprised of an iPod shuffle with a specific storage capacity, a case holder, headphones, speakers, car audio integration kit, power kit and warranty terms all of which were bundled together at a specific price point. In the CBC study, participants were asked 24 choice questions. In each choice question participants were shown three different iPod packages and were asked to choose the package they would buy if they had to buy one or choose 'None of these' if they do not like any of the packages in the choice question. After they answered all the 24 choice questions they were shown the specific iPod package that was available for purchase. Subsequently, a final hold-out choice question was shown to the participants which required them to choose one iPod package from 16 different alternative package designs or perhaps 'None of these' if they did not like any of the design shown. Study participants were informed that a randomly selected participant would be given $250 which can be used to buy the iPod package that was revealed to them after they completed answering the first 24 choice questions. Whether they get to buy the package and at what price is determined by the procedure is shown in Figure 1.

[Insert Figure 1 here]

In the product incentive scheme study participants face unattractive outcomes if they do not answer CBC questions truthfully. If their choices in the CBC choice questions are inconsistent with their true preferences then the estimated willingness to pay for the package (i.e., WTP in Figure 1) can either be too low or too high. If it is too low then the chances of being able to buy the package at an attractive price decrease as the random number (i.e., x in Figure 1) that is drawn is likely to be higher than the estimated willingness to pay. In contrast, if the estimated willingness to pay is too high then they may end up paying more than the camera is worth to them.

In the award incentive scheme, study participants were told that, apart from a flat fee ($15), a monetary award ($300) can be won by one participant in the research study. They were told that they will answer 24 choice questions and a final hold-out choice question. We informed them that their answers to 16 randomly selected choice questions out of the 24 choice questions will be used to predict their choices for the remaining 8 choice questions. The participant whose choices we are able to predict the best will be the winner of $300.

The hybrid incentive scheme combines features of the product incentive and the award incentive scheme. In the hybrid incentive scheme, study participants followed a similar sequence of activities as the product incentive scheme. In other words, they answered 24 choice questions, saw a product that they can potentially buy and answered a final hold-out question. However, unlike the product incentive scheme, the participant who was given the opportunity to buy the product shown after 24 choice questions was not selected at random. Instead, we used the procedure outlined in the award incentive scheme to select the study participant. Thus, the hybrid incentive scheme combines the competitive nature of the award incentive scheme and the truth-telling component of the product incentive scheme. Figure 2 shows the compensation schemes we evaluated in this paper.

[Insert Figure 2 here]

The Ding study showed that truth-telling (i.e., answering choice questions consistent with that of their underlying preferences) is the unique Bayesian Nash Equilibrium in the context of the product incentive scheme. In the appendix, we show that truth-telling is not the only equilibrium strategy in the award incentives scheme. Any strategy that requires respondents to be consistent with some arbitrary preference structure is an equilibrium strategy. The presence of multiple equilibria in the award incentive is not surprising as, unlike the product incentive, the award incentive does not have any mechanism to induce truth-telling in study participants. Therefore, we decided to structure the experiment such that study participants did not have sufficient time or knowledge to think of an alternative to truth-telling. We elaborate on this issue in the context of experimental design in the next section. Since the hybrid incentive study incorporates the truth-telling component of the product incentive scheme, the unique Bayesian Nash Equilibrium in the hybrid incentive is truth-telling.

3. The Five CBC Studies

In this section, we describe the five CBC studies in order to delimit the scope and applicability of our findings. As mentioned in the introduction, a typical CBC study can be divided into three phases: Experimental Design, Data Collection and Estimation. We describe the decisions we took in each phase and the controls in place to establish experimental equivalence across all five experiments so that we can rule out alternative explanations for our findings.

3.1 Experimental Design

We imposed four overall constraints on our experimental design. One, we wanted to keep the power of incentives offered to be on par with that of the Ding study. Two, we wanted to choose a different but related product as the iPod so that we can extend our understanding of the effectiveness of the product incentive to another product category. Three, we wanted to keep the complexity of the CBC study as similar as possible to that of the Ding study. Fourth, we designed our experiments such that, to the extent possible, the only difference between all five experiments is the incentive scheme used in the experiment. These constraints ensure that we are making a 'apples-to-apples' comparison when we compare and contrast the hit rates found in the Ding study vis-à-vis the five experiments in our study. Finally, our research budget constraints prevented us from considering a more expensive product such as a laptop.

Choice of Product

The constraints on the experimental design mentioned above led us to choose digital cameras as the product in our CBC study. Specifically, digital cameras are as attractive as iPods to undergraduate students and the pricing of digital cameras and iPods tend to be in the same ballpark range which ensures that we keep the power of incentives similar to that of the Ding study. Digital cameras, as a context for ratings based conjoint, was also used by Netzer and Srinivasan (2009) successfully with student respondents.

Complexity of CBC Study

The amount of time and effort it takes to answer choice questions accurately depends on the complexity of the CBC study. The complexity of a CBC study increases as the number of features and the number of levels per feature increase. In light of the above, we decided to keep the number of features and the number of levels per feature to be as identical as possible to that of the Ding study. See Table 1 for a list of product features and corresponding levels used in the Ding study and in our five experiments. Our choice of features and levels are similar to that of Netzer and Srinivasan (2009) and were validated by an informal survey of brands available at Best Buy's website. A pilot study with undergraduate students was also used to validate our choices. Our choices resulted in a $2^1 \times 3^6 \times 4^1$ design which is comparable to the $2^2 \times 3^5 \times 4^1$ design used by the Ding study.

[Insert Table 1 here.]

Question Sequence.

Similar to the Ding study we generated a fixed efficient design using SAS and modified the design to eliminate dominated alternatives. Participants in all five studies answered 24 choice questions each one of which asked them to choose from one of three digital camera designs and a 'None of these' alternative in case they did not like any of the designs shown in that choice question. A final hold-out choice question was then presented which required them to choose one alternative from among 16 different digital camera designs and a 'None of these' alternative in case they did not like any of the designs shown in the hold-out choice question. In the Product Incentive and the Hybrid Incentive, the product that a participant could buy was revealed to them after they completed answering all the 24 choice questions but before they answered the holdout question.

Design of Holdout Profiles.

In the Ding study, the holdout task showed 16 different iPod packages and the selected participant was required to purchase the chosen holdout alternative if the outcome of a coin toss was tails (see Figure 1). Thus, the effectiveness of the product incentive scheme in the Ding study was evaluated using realistic product choices. However, designing a hold-out task such that all the alternatives in the hold-out are available for purchase is not always easy or even desirable. Note that, to the extent feasible, a holdout task should have the same 'type' of product profiles as were used in the estimation set. Conjoint studies usually include hypothetical levels to take product design decisions and construct unrealistic combination of levels into product profiles so that the resulting experimental design is efficient. In contrast, realistic products do not have unrealistic levels and do not have unrealistic combination of levels. Thus, constructing a realistic holdout task that accurately mirrors the product profile in the estimation set is not easy.

If the profiles in the holdout task are very different than the profiles in the estimation set then our ability to predict the holdout choices may be poor. In such a situation, it is not possible to ascertain if our inability to predict the holdout choices is because the holdout profiles are very different compared to the choice profiles or because the incentive scheme is ineffective. The Ding study's choice of iPod package was a product bundle offered by the author which made it easy to construct a holdout task with realistic profiles which were similar to the ones shown in the choice set. In our context of digital cameras, we found that constructing a realistic holdout set that is also similar to the profiles in the estimation set was not easy. Thus, we chose to use include hypothetical profiles in the holdout task instead of actual digital cameras.

3.2 Data Collection

We will now describe our data collection procedures with a specific focus on our recruitment email and knowledge of participants.

Recruitment of Study Participants

The participants for all studies were recruited via email using a list provided by the university. To preserve the integrity of all experiments we did not disclose the full compensation structure in the recruitment email. The recruitment email for the award, the weak flat fee and the strong flat fee mentioned that participants would earn $15 for participating in a study which would take about an hour of their time. In contrast, the recruitment email for the product incentive study mentioned that participants would earn $10 whereas the recruitment email for the hybrid incentive mentioned that participants would earn $20. We offered $10 instead of $15 to students in the product incentive study to maintain parity with the Ding study which also offered $10 to study participants[3]. The hybrid incentive scheme was the last study to be conducted and we increased the compensation to $20 with the hope that we would obtain a higher sample size.

Study participants in the product, award, hybrid and the strong flat fee were informed about the true compensation structure after all participants arrived at the campus lab to participate in the study but before they actually started answering the CBC questions. Our disclosure of the additional compensation after all participants showed up in the lab ensured that students in all experiments had similar motivation to attend the study and that there were no attempts made to figure out the survey questions in advance. In addition, we wanted to ensure that participants in the product incentive and the hybrid incentive studies were adequately motivated to answer CBC questions and wanted to maintain parity with the Ding study. Therefore, we told potential participants in the product and the award incentive studies that they should agree to participate if they had an interest in buying a digital camera over the next few months. The Ding study had a similar qualification criterion. Our recruitment emails generated the following sample sizes: Weak Flat Fee: 28, Strong Flat Fee: 19, Product Incentive: 24, Award Incentive: 21 and Hybrid Incentive: 22. These sample sizes are comparable to experiment 1 of the Ding study whose sample sizes were: Product Incentive: 25 and Weak Flat Fee: 24.

[3] Since the actual compensation structure was revealed to participants after they arrived at the campus lab, the differential amount of $5 only impacts their initial decision whether to participate in the study. Therefore, the differences in out-of-sample predictions between the product and the award incentive cannot be attributed to the difference of $5 between the two schemes.

Knowledge of Participants

The difficulty of choice questions and the ease with which a participant can make a choice depend on the degree of knowledge a participant has with respect to digital cameras. An expert in digital cameras who has a lot of experience in the area would probably find it very easy to make choices that reflect their true preferences whereas a novice would find it very hard. Therefore, to equalize the knowledge level across all experiments for all participants we provided a glossary of digital camera features that help in answering the choice questions (see Figure 3). It is possible to swing from one extreme, where we do not provide any information whatsoever, to the other extreme, where we discuss the fine details of the difference between 7 megapixels versus 8 megapixels. We chose a middle ground which we felt is a reasonable compromise between the two extremes. In summary, respondents knew which features they would see in the CBC questions and what the features mean but were *not informed* about the specific levels that will appear in the CBC question for each one of these attributes.

[Insert Figure 3 here]

Study Decisions that Encourage Truth-telling

As we mentioned earlier, truth-telling is not the only equilibrium strategy in the award incentive scheme. Therefore, several of the CBC study decisions we took are designed to encourage study participants to answer choice questions according to their true preferences. Specifically, note the following points:

1. While respondents did know the features they would see in the various CBC questions they did not know the corresponding levels for the features or the choice questions in the study.

2. Respondents were told not to use the browser's forward or backward buttons but to use the links provided within the web page to go to the next question. Each choice question had a link to go to the next question but not to the previous one.
3. Participants were told that even if they were to go back and change some of their answers their changes will not be stored by the server.
4. Each choice question was expired in cache so that a re-load required an explicit page reload from the web server. The web server would then skip all the choice questions that have been answered by the respondent and show the first as yet unanswered choice question.
5. A participant could take the survey just once after which they would not be able to log in to do the survey again.

The above decisions have the following consequences for respondents: (a) respondents have to answer a choice question without any knowledge of future choice questions, (b) they cannot go back to change their answers, (c) they cannot go back to see how they answered previous choice questions so that they can choose a consistent answer for a choice question. Therefore, answering choice questions truthfully is likely to be an easier task than trying to be consistent across all choice questions in some arbitrary manner. To reinforce truth-telling as a viable strategy, participants were explicitly told that answering the choice questions consistent with their true preferences would maximize their chances of winning $300[4]. Finally, respondents were told

[4] Respondents in the product and the hybrid incentive study were also told that answering the choice questions consistent with their true preferences would beneficial to them. Thus, this is not a point of difference between the studies.

about the award of $300 only after they arrived in the lab but before they started the study and hence we do not think that they had sufficient time come up with an alternative to truth-telling.

3.3 Estimation

Consistent with the Ding study and following the current practice in estimating choice models (see Allenby and Ginter (1995) and Allenby, Arora, and Ginter (1998) for similar models) we used a hierarchical Bayesian multinomial logit model to estimate individual partworths in all experiments. Specifically, we assumed that the utility of the i^{th} study participant for the a^{th} alternative in the q^{th} question is given by:

$$U_{iqa} = X_{qa}\beta_i + \epsilon_{iqa} \quad (1)$$

where,

X_{qa} is a $1\ x\ p$ design vector representing the a^{th} alternative in the q^{th} question,

β_i is a $p\ x\ 1$ vector that represents the i^{th} participant's partworth values,

ϵ_{iqa} is random error that is independently, identically distributed as extreme value and

p is the number of partworths we are estimating.

Thus, the probability that the i^{th} participant chooses the a^{th} alternative in the q^{th} choice question is given by:

$$P(y_{iq} = a) = \frac{\exp(X_{qa}\beta_i)}{\sum_a \exp\,(X_{qa}\beta_i)} \quad (2)$$

We further assumed that the individual partworth vectors, β_i, follow a normal distribution as given below:

$$\beta_i \sim N(\bar{\beta}, \Sigma) \quad (3)$$

We assumed a diffuse conjugate prior for $\bar{\beta}$ (i.e., a normal distribution centered at 0 and a 'large' variance) and a conjugate wishart distribution for Σ^{-1} whose mean is an identity matrix and whose degrees of freedom equal the number of parameters we are estimating in our model i.e., we set the degrees of freedom to equal p. Using the above assumptions, we estimated five HB multinomial logit models for each one of the five experiments and assessed convergence by running three parallel MCMC chains from dispersed starting points. We estimated the posterior means for partworths to compute hit rates only after trace plots of a few randomly selected parameters and the potential scale reduction factor indicated that that we had achieved convergence (Gelman et al. 2004, See pg 297 for a discussion of the role of potential scale reduction factor in assessing convergence of MCMC chains.).

Table 2 summarizes the decisions we took across each one of the three stages (i.e., Experimental Design, Data Collection and Estimation) of a typical CBC study. For the most part, the differences are relatively minor. Apart from the differences in the incentive structures per se, two major differences exist which can potentially offer an alternative explanation for some our results: One, all the experiments including the Ding study are between subject experiments. Thus, we cannot completely rule out differences in subject populations as a potential explanation for our findings. Two, the choice of product categories (digital cameras vs. iPods) can make a difference. However, this difference between the experiments can only explain any pattern of results between the five studies we ran and the experiments of the Ding study. In summary, the above discussion and Table 2 suggests that, for the most part, any differences in hit rates must be attributed to the differences in incentive structures between the experiments.

[Insert Table 2 here]

4. Results

In this section, we compare the relative performance of the product incentive, the award incentive, the hybrid incentive and the strong flat fee incentive vis-à-vis the weak flat fee incentive scheme using in-sample goodness of fit and out-of-sample predictions. For both in-sample goodness of fit and out-of-sample predictions, we used two metrics to assess performance which, for ease of exposition, we henceforth refer to as 'Top Hits' and 'Top Two Hits'. The 'top hits' metric is identical to the hit rate used in the literature to evaluate the performance of competing research methods in predicting answers to choice questions. In other words, for a particular choice question, we used the estimated partworths to compute the utilities of all the alternatives in that choice question. If the alternative with the maximum estimated utility is identical to the actual choice of a respondent then we count it as a 'hit'. We then computed the percentage number of hits for each respondent and report the average percentage across all respondents. Top two hits are computed in a similar way with one difference: When computing a top two hit, we count a prediction as a 'hit' if a respondent's answer to a choice question is identical to either the alternative with the highest estimated utility or the alternative with the second highest estimated utility. Where appropriate we also report hit rates from experiment 1 of the Ding study.

Table 3 reports the in-sample goodness-of-fit for the competing incentive schemes. We also report the relative improvement of the product, the award, the hybrid and the strong flat fee incentive schemes vis-à-vis the weak flat fee incentive scheme and corresponding p-values which were computed using bootstrap. Consistent, with the finding of the Ding study, the in-sample goodness of fit of the product and the weak flat fee incentive scheme are comparable to each other. In addition, the strong flat fee also shows comparable in-sample goodness-of-fit. However, the in-sample goodness of fit for the award and the hybrid incentive schemes is significantly better than that of the weak flat fee. These results suggest that respondents in the award and the hybrid incentive schemes were more consistent in their answers when responding to choice questions as compared to the respondents in the other three incentive schemes.

[Insert Table 3 here].

Table 4 reports the out-of-sample predictions for the competing incentive schemes. Four points are of interest. One, somewhat surprisingly, the strong flat fee is *not* better than the weak flat fee in out-of-sample predictions. Two, the hit rates for the product incentive scheme in our study and that of the Ding study are similar. Three, the product incentive scheme outperforms the weak flat fee incentive scheme in predicting out-of-sample choices by a 2 to 1 margin (Hit rates: 29% vs. 14%). Finally, both the award and the hybrid incentive schemes outperform the product incentive scheme in predicting out-of-sample choices by a 2 to 1 margin (Hit rates: 62% vs. 14%) and a 1.4 to 1 margin (Hit Rates: 41% vs. 29%) respectively. We elaborate on these findings over the next few paragraphs.

[Insert Table 4 here]

The performance of the product incentive scheme in our study is comparable to the performance of the product incentive scheme of the Ding study thus validating the effectiveness of the product incentive scheme when predicting hypothetical choices in a different product category and with a different respondent population. These results establish the robustness of the product incentive scheme in motivating respondents in CBC studies to answer choice questions consistent with their underlying preferences.

In order to better understand the reasons for the poor performance of the strong flat fee incentive scheme, we performed additional analysis as described next. We estimated the partworths of all respondents using their answers to the first 10 choice questions and computed hit rates with respect to their choices for the 11th choice question. Similarly, we used their answers to the first 23 choice questions and computed hit rates with respect to the 24th choice question. As originally pointed out in the Ding study, the above tests of out-of-sample performance are weak because we need to predict correctly only 1 out of 4 possible alternatives[5]. Table 5 reports the top hits and top two hits when predicting respondents choices for the 11th choice question and Table 6 reports the hit rates when predicting the choices for the 24th choice question. When we compare the 'top hit' results from Table 4, Table 5 and Table 6, we see the following pattern with respect to the performance of the four competing schemes (i.e., the Strong Flat Fee, the Product Incentive, the Award Incentive and the Hybrid Incentive) vis-à-vis the Weak Flat Fee:

(a) All four competing schemes outperform the Weak Flat Fee when predicting choices for the 11th choice question (See Table 5).

[5] This test is weak because a naïve method of picking an alternative at random as the predicted choice is guaranteed to have a hit rate of 25 % (1 out of 4) as compared to the hit rate of 6 % (1 out of 17) that would be obtained by the same method for the actual hold-out question used in our paper and in the Ding study.

(b) All four competing schemes are comparable to the Weak Flat Fee when predicting choices for the 24th choice question (See Table 6).

(c) The Strong Flat Fee is comparable to the Weak Flat Fee when predicting hold-out choices (See Table 4).

(d) In contrast, the Product Incentive, the Award Incentive and the Hybrid Incentive are better than the Weak Flat Fee in predicting hold-out choices (See Table 4).

[Insert Table 5 and Table 6 here]

The above pattern of results suggests that in the strong flat fee experiment respondents answer choice questions accurately at the beginning of the study but are unable to maintain their accuracy levels for the entire duration of the study. In contrast, respondents in the product, the award and the hybrid incentive are able to answer choice questions accurately for the entire duration of the study which eventually translates into superior hit rates for the hold-out task. In other words, the increase in compensation from $15 (as promised in the recruitment email) to $50 on the day of the study motivates respondents to answer CBC questions accurately but their motivation weakens towards the end of the study. We now discuss the superiority of the award and the hybrid incentive schemes.

There are two possible explanations for the superiority of the award incentive study. It is possible that respondents find it easier to be consistent instead of answering choice questions according to their true preferences. For example, a respondent can potentially select 'None of these' across all choice questions including the hold-out and be a potential winner. While we took steps to encourage truth-telling for the participants in the award incentive study we cannot completely rule out the above possibility. The second explanation relies on the differences in the structure of the award incentive and the product incentive. The product incentive scheme is a lottery scheme where a respondent obtains additional payoff (via $300 plus the opportunity to buy the digital camera) only if he/she is selected by the random draw. On the other hand, the award incentive scheme is a competitive scheme where a respondent obtains the additional payoff of $300 only if they win the competition by being the person whose choices we were able to predict the best. Thus, it is possible that respondents in the award incentive scheme are better motivated because of the competitive nature of the scheme to answer choice questions consistent with their true preferences. Unfortunately, in the award incentive study, the above two explanations are confounded and hence we cannot definitively identify which explanation is the primary reason for the superiority of the award incentive scheme.

In contrast, the structure of the hybrid incentive allows us to offer a cleaner explanation for its superiority. Unlike the award incentive scheme, the hybrid incentive scheme motivates respondents to truth-telling. Similar to the product incentive scheme, if a respondent were to answer choice questions untruthfully they face potentially unattractive outcomes if they are the winner of the competition. Answering choice questions inconsistent with their true preferences would result in either a low or a high estimate for their willingness to pay for the product that is available to purchase. If it is low then the chances of their being able to buy the product decreases whereas if it is too high then they may end up paying more than their true willingness to pay. Therefore, in the hybrid incentive scheme, respondents have an incentive to not only be consistent but to also be truth-telling. Thus, the superiority of the hybrid incentive vis-à-vis the product incentive scheme can be attributed to the competitive nature of the scheme.

In Table 7, we report the utility decrease for a $100 increase in price from the five experiments we conducted and from experiment 1 of the Ding study. As the table indicates, the Ding study found that mean price sensitivity in the product incentive was comparable to that of the flat fee. Our findings in the context of digital cameras indicate the opposite effect. In our experiments, participants in the product incentive were the most sensitive to price changes followed by the award incentive and the hybrid incentive. Participants in the weak flat fee and the strong flat fee had comparable price sensitivities. The extent of heterogeneity among participants on price sensitivity also show a different pattern when we compare our findings with that of the Ding study. Specifically, in the Ding study, participants were more heterogeneous in the flat fee whereas in our study participants were more heterogeneous in the product, award and hybrid incentives with participants in the hybrid incentive being the most heterogeneous. The Ding study suggested that the difference in price sensitivity patterns is probably due to the price intervals used in the experiments (i.e., the maximum and the minimum price differences between the alternatives) In our experiments, the minimum and the maximum possible price difference between the alternatives were similar to that of the Ding study ($30, $90 in the Ding study and $34.50, $70 in our study). Since, our findings are the opposite to those of the Ding study despite similar price intervals; we can discount price intervals as an explanation for the price sensitivity findings across all the experiments.

[Insert Table 7 here]

5. Discussion and Conclusion

In this paper, we investigated the performance of five types of incentive schemes (i.e., the hybrid incentive scheme, the award incentive scheme, the product incentive scheme, the weak flat fee and the strong flat fee) in motivating respondents to accurately answer choice questions in a CBC study. Our primary finding is the superiority of the award and the hybrid incentive schemes vis-à-vis the product incentive scheme which suggests that incentive schemes that induce competition among respondents are better than those that do not. In addition, the robustness of the product incentive in predicting hypothetical choices is reassuring. From a theoretical perspective, we find that while a higher fixed fee does motivate respondents to be accurate their accuracy levels drop towards the end of the CBC study.

The superiority the award incentive comes at a cost. Unlike the product incentive scheme, the award incentive scheme does not have any mechanism to motivate respondents to be truthful. Consequently, respondents may simply be consistent with some arbitrary preference structure instead of their true preferences. However, the hybrid incentive scheme does not have any such disadvantages as it builds on the strengths of the award and the product incentive scheme namely competitiveness and truth-telling. Therefore, the hybrid incentive scheme is a viable incentive scheme that can be used by practitioners and academics to motivate respondents to be truthful.

Our results also suggest three important areas for future research. One, our findings regarding price sensitivity is not consistent with the findings of the Ding study. Further research is needed to better understand how price sensitivity changes when consumers are offered different types of truth-telling versus fixed fee compensation schemes. Two, competitive incentive schemes have the potential to discourage some of the study participants to not fully engage in the study if they believe that they do not stand a chance of being a winner. Participants may feel discouraged if they do not have enough experience in the product category or if the number of competitors is too high (e.g., if one winner is being selected from 300 respondents). Therefore, it is important to investigate the impact of product knowledge and sample size on the effectiveness of competitive incentive schemes. Three, it is important to assess the effectiveness of the product and the award incentive schemes with higher sample sizes in order to test their robustness in motivating respondents to answer CBC questions truthfully. We hope to address some of these issues in future research.

References

Allenby, Greg et al. (2005), "Adjusting Choice Models to Better Predict Market Behavior," *Marketing Letters,* 16 (Dec), 197-208.

Allenby, Greg M., Neeraj Arora, and James L. Ginter (1998), “On the Heterogeneity of Demand,” *Journal of Marketing Research*, 35 (August), 384–89.

——— and James L. Ginter (1995), “Using Extremes to Design Products and Segment Markets,” *Journal of Marketing Research*, 32 (November), 392–403

Dong, S., M. Ding and J. Huber (2009), "A Simple Mechanism to Incentive Align Conjoint Experiments", *Intern. J. of Research in Marketing,* forthcoming

Dahan, Ely, A. Soukhoroukova, and M. Spann (2007), "Preference Markets: Organizing Securities Markets for Opinion Surveys with Infinite Scalability," (November 19, 2007). Available at SSRN: http://ssrn.com/abstract=1031754

Ding, Min (2007), "An Incentive-Aligned Mechanism for Conjoint Analysis," *Journal of Marketing Research,* 44 (May), 214-223.

Ding, M., Grewal,R., and J. Liechty. (2005), "Incentive-Aligned Conjoint Analysis," *Journal of Marketing Research*, 42(1), 67-82

Ding, M., Park, YH, and E. Bradlow (2009), "Barter Markets for Conjoint Analysis," *Management Science*, Vol. 55, No. 6, pp 1003-1017.

Evgeniou, Theodoros, Massimiliano Pontil, and Olivier Toubia (2007), "A Convex Optimization Approach to Modeling Consumer Heterogeneity in Conjoint Estimation," *Marketing Science,* 26 (Nov/Dec), 805-818.

Gelman, Andrew et al. (2004), *Bayesian Data Analysis.* Chapman & Hall/CRC.

Liu, Qing, Thomas Otter, and Greg M. Allenby (2007), "Investigating Endogeneity Bias in Marketing," *Marketing Science*, 26 (Sep/Oct), 642-650.

Louviere, Jordan J. et al. (2008), "Designing Discrete Choice Experiments: Do Optimal Designs Come at a Price?" *Journal of Consumer Research,* 35 (Aug), 360-375.

Netzer, Oded, V.“Seenu” Srinivasan (2010) "Adaptive Self-explication of Multi-attribute Preferences", *Journal of Marketing Research*, forthcoming

Netzer, Oded et al. (2008), "Beyond Conjoint Analysis: Advances in Preference Measurement," *Marketing Letters*, 19 (Dec), 337-354.

Oosterbeek, Hessel, Randolph Sloof, and Gus V. D. Kuilen (2004), "Cultural Differences in Ultimatum Game Experiment: Evidence from a Meta-Analysis," *Experimental Economics*, 7 171-188.

Park, YH, M. Ding and V. Rao (2008), "Eliciting Preference for Complex Products: A Web-Based Upgrading Method,".*Journal of Marketing Research*, 45, 562-574

Rao, Vithala (2008), "Developments in Conjoint Analysis" in *Marketing Decision Models*, Berend Wierenga, ed. Springer

Toubia, Olivier, John Hauser, and Rosanna Garcia (2007), "Probabilistic Polyhedral Methods for Adaptive Choice-Based Conjoint Analysis: Theory and Application," *Marketing Science*, 26 (Sep/Oct), 596-610.

Yu, Jie, Peter Goos, and Martina Vandebroek (2009), "Efficient Conjoint Choice Designs in the Presence of Respondent Heterogeneity," *Marketing Science*, 28 (Jan/Feb), 122-135.

Tables

iPod Package		
Features	No of Levels	Values
Storage	2	512 MB, 1 GB
Case holder	3	None, armband, sport case
Headphones	3	Apple standard, Nike Vapor, Nike Duro
Speakers	3	None, Monster, Creative
Car Audio	3	None, Cassette Adapter, FM Transmitter
Power	3	USB, USB + Battery, USB + Power Adapter
Warranty	2	Basic, Extended
Price	4	\$129, \$159, \$189, \$219

Digital Camera		
Features	No of Levels	Values
Movie Mode	2	Not Present, Present
Color	3	Black, Silver, Blue
Internal Storage	3	8 MB, 16 MB, 32 MB
LCD Screen Size	3	1 inch, 2 inch, 3 inch
Brand	3	Canon, Nikon, Sony
Optical Zoom	3	1x, 2x, 3x
Megapixels	4	7, 8, 9, 10
Price	3	\$149.99, \$185.49, \$219.99

Table 1: Product features and levels used in the Ding study and the five experiments performed in this paper

Stage of CBC Study	Decisions Regarding	Weak Flat Fee	Strong Flat Fee	Award	Product	Hybrid	Product (From the Ding Study)
Experimental Design	Conjoint Task	24 CBC questions from a fixed, efficient design.					
	Hold-Out Task	16 alternatives plus a 'None of these' alternative					
	Study Complexity	2^1 x 3^6 x 4^1					2^2 x 3^5 x 4^1
	Product	Digital Cameras					iPod Bundle
	Type of Study	Between Subjects					
	Qualification Criteria	Greater than 18 years of age.			Greater than 18 years of age and Must have an interest in buying a camera over the next few months.		
Data Collection	Profile of Respondents	Undergraduate Students					Undergraduate and Graduate Students
	Knowledge of Participants	Identical Controls (See Figure 3)					Not Known
	Study Location	Campus Lab					
	Compensation in Recruitment Email	$15 for 1 hour*; No mention of any extra compensation.					
	Informed of Additional Compensation	Not Applicable	After all participants arrived at the lab but before they started the study.				
	Technology and Non-Technology Controls	Identical Controls (See Section: "The Three CBC Studies")					
Partworth Estimation	Model	HB Multinomial Logit with Conjugate Priors					
	No of Parameters Estimated	16					14

* Study participants in the Product Incentive scheme were offered $10 instead of $15 and participants in the Hybrid Incentive were offered $20.

Table 2: Differences and Similarities between the five experiments and with Experiment 1 from the Ding Study

Incentive Scheme	Top hit	Difference vis-à-vis Weak Flat Fee	p-value	Top Two Hits	Difference vis-à-vis Weak Flat Fee	p-value
Weak Flat Fee	87%			96%		
Strong Flat Fee	84%	-3%	0.08	96%	0%	0.48
Product Incentive	88%	1%	0.31	97%	1%	0.20
Award Incentive	92%	5%	0.00	99%	3%	0.00
Hybrid Incentive	92%	5%	0.00	99%	3%	0.00
Weak Flat Fee (Ding Study)	78%			Not Known		
Product Incentive (Ding Study)	78%	0%	0.49			

Table 3: In-Sample Goodness of Fit for Competing Incentive Schemes.

Incentive Scheme	Top hit	Difference vis-à-vis Weak Flat Fee	p-value	Top Two Hits	Difference vis-à-vis Weak Flat Fee	p-value
Weak Flat Fee	14%			39%		
Strong Flat Fee	21%	7%	0.25	32%	-7%	0.29
Product Incentive	29%	15%	0.09	54%	15%	0.13
Award Incentive	62%	48%	0.00	76%	37%	0.00
Hybrid Incentive	41%	27%	0.02	64%	25%	0.04
Weak Flat Fee (Ding Study)	17%			38%		
Product Incentive (Ding Study)	36%	19%	0.09	64%	26%	0.04

Table 4: Out-of-sample Predictions for Competing Incentive Schemes when predicting the choices for the hold-out choice question

Incentive Scheme	Top hit	Difference vis-à-vis Weak Flat Fee	p-value	Top Two Hits	Difference vis-à-vis Weak Flat Fee	p-value
Weak Flat Fee	29%			54%		
Strong Flat Fee	58%	29%	0.02	84%	30%	0.01
Product Incentive	63%	34%	0.01	75%	21%	0.05
Award Incentive	67%	38%	0.00	86%	32%	0.01
Hybrid Incentive	86%	57%	0.00	95%	41%	0.00

Table 5: Out-of-sample Hit-Rates for the Competing Incentive Schemes when predicting the choices for the 11^{th} choice question

Incentive Scheme	Top hit	Difference vis-à-vis Weak Flat Fee	p-value	Top Two Hits	Difference vis-à-vis Weak Flat Fee	p-value
Weak Flat Fee	57%			86%		
Strong Flat Fee	58%	1%	0.47	95%	9%	0.14
Product Incentive	46%	-11%	0.20	75%	-11%	0.15
Award Incentive	67%	10%	0.24	95%	9%	0.12
Hybrid Incentive	73%	16%	0.12	95%	9%	0.10

Table 6: Out-of-sample Hit-Rates for the Competing Incentive Schemes when predicting the choices for the 24th choice question

Incentive Scheme	Valuation	Heterogenity
Weak Flat Fee	-3.21 (0.53)	4.14 (1.89)
Strong Flat Fee	-3.52 (0.74)	6.58 (3.86)
Product Incentive	-6.89 (0.85)	9.97 (5.30)
Award Incentive	-5.77 (1.09)	17.38 (9.38)
Hybrid Incentive	-5.14 (1.27)	24.99 (12.71)
Flat Fee (Ding Study)	-5.87 (0.66)	6.12 (2.83)
Product Incentive (Ding Study)	-5.43 (0.47)	1.96 (1.25)

Table 7: Utility Decrease per $100 increase in price from our study and from that of the Ding Study. Figures in brackets are standard deviations.

Figures

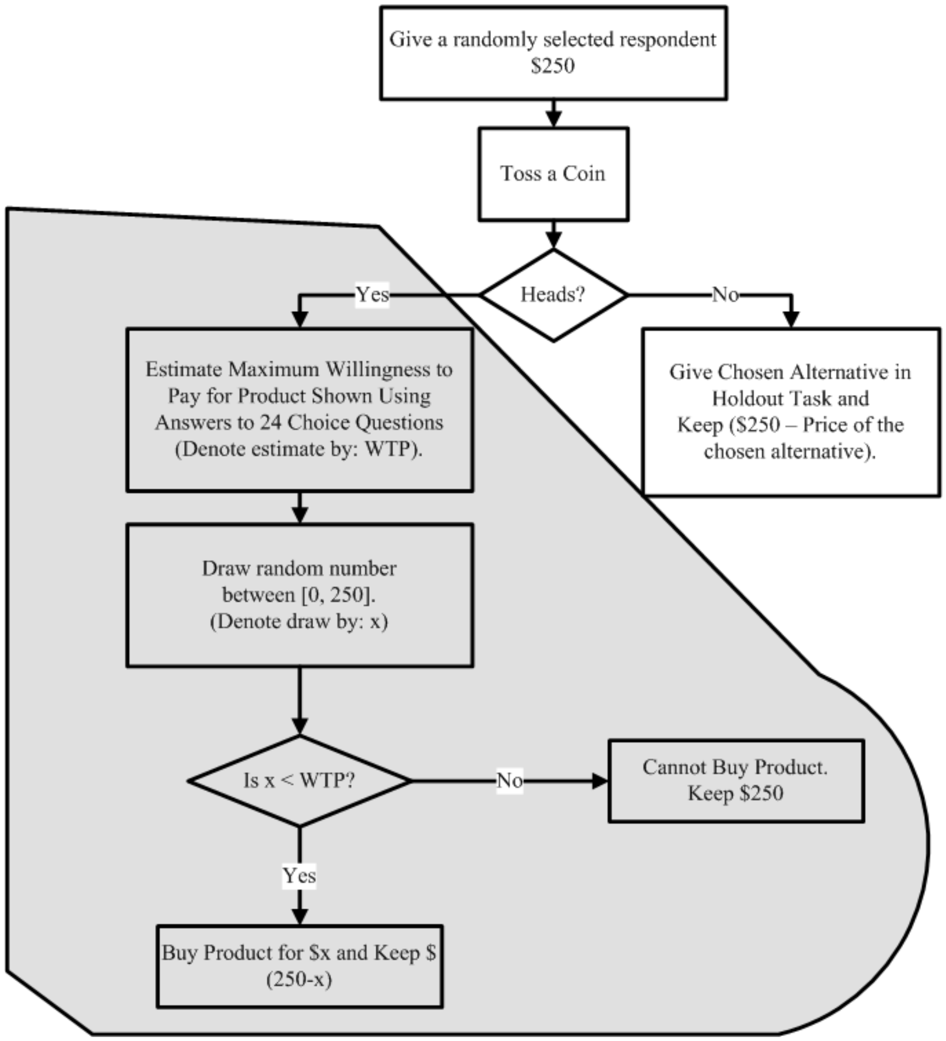

Figure 1: The Product Incentive Scheme. The grey shaded part of the scheme is the component that motivates respondents to answer choice questions consistent with their true preferences.

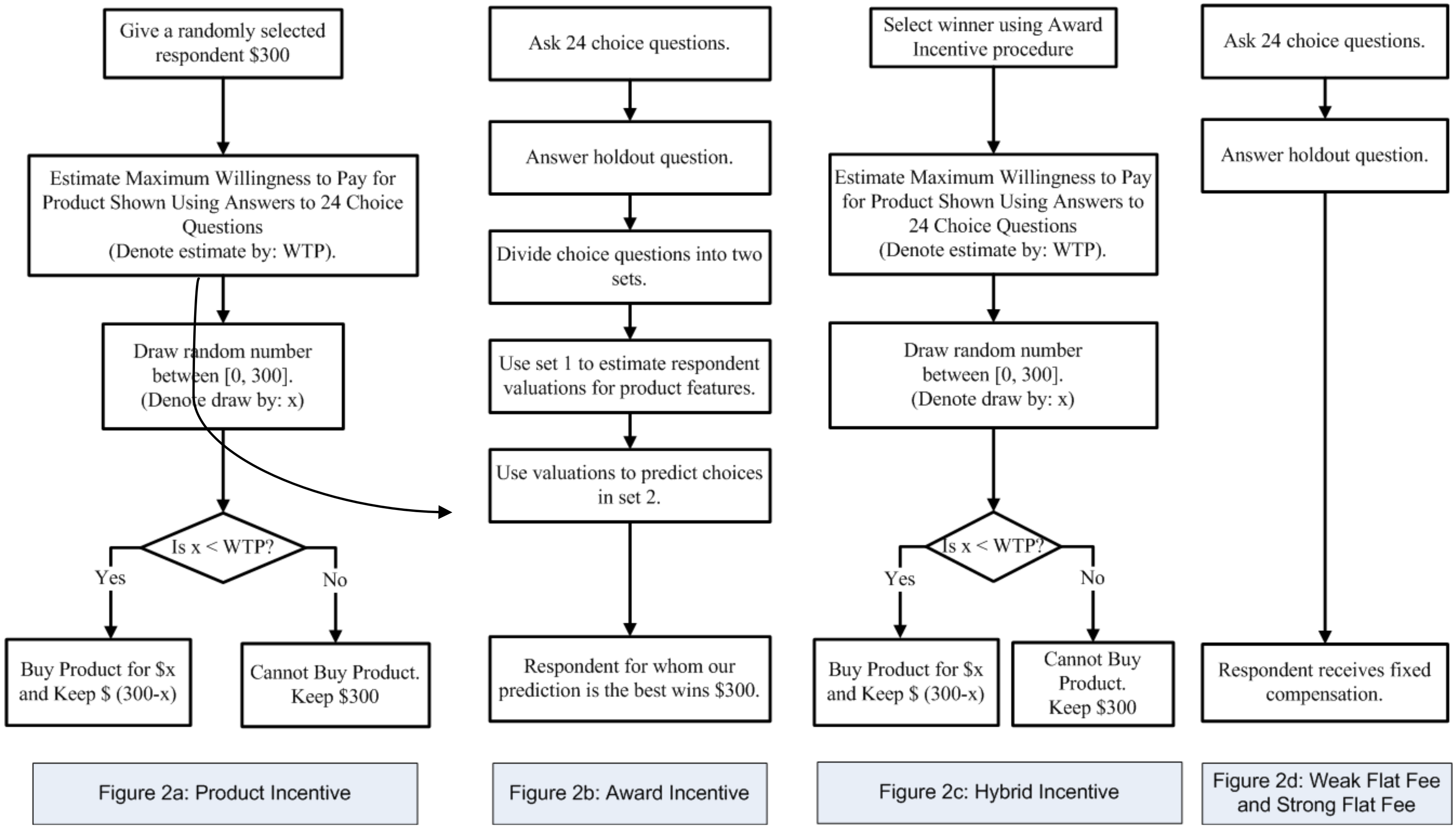

Figure 2: The five compensation schemes that we evaluate in this paper.

Glossary

We will ask you to consider different types of digital camera designs which vary on various features such as Brand Name, Color etc. The table below shows a list of features that will appear in the various designs alongwith a brief description of each feature.

Please go through the glossary carefully and consider carefully the type of digital camera that you would buy if you were to buy one today.

Features	Description
Brand Name	The brand names you will see in the study are well know brand names.
Color	The colors you will see in the study are some of the typical colors available for a digital camera.
Internal Memory	Internal memory refers to the amount of storage that is integrated with the camera. The more the memory the more pictures you can store on the camera.
LCD Size	LCD screens let you frame the shot, review the shots after they have been taken and display various menu settings. In general, bigger screen sizes offer more flexibility.
Movie Mode	Movie mode refers to the ability of camera to take short video clips.
Optical Zoom	Optical zoom magnifies images so that you can take close-up shots of faraway objects. The higher the zoom (say 3x as opposed to 2x) the better will be your ability to zoom in on a faraway object.
Megapixels	Megapixels refer to the clarity that an image has. The higher the megapixels the lower is the quality decrease when you enlarge an image.
Price	All prices are in US Dollars.

Continue

Figure 3: Glossary of Product Features Shown to Study Participants in all five Experiments

Appendix

Bayesian Nash Equilibrium for the Award Incentive Study

Conditions of CBC Study:

For simplicity, we impose some conditions on the CBC study as given below:

Condition C1: There are 2 respondents in the study.

Condition C1 is not restrictive from a game theoretic perspective. Increasing the number of respondents from 2 to an arbitrary number, say n, will increase our mathematical burden without lending any additional insight.

However note that robust estimation (e.g., using HB multinomial logit) of partworths is difficult if we have data from just 2 respondents. Since, the focus of this appendix is on the equilibrium behavior of respondents and not on statistical estimation per se, we simply assume that we can estimate partworths even when we have data from just 2 respondents.

Condition C2: Answers to the first q choice questions are used to predict the choices of respondents to the $q + 1^{\text{th}}$ question.

Condition C2 specifies that respondents know that we will use their answers to the first q choice questions to predict their answer to the last choice question.

Our implementation of the award incentive scheme differs from C2 in two respects: One, we predict the choices of respondents for 8 questions instead of just 1. Two, we select these 8 questions at random. Relaxing C2 to accommodate the above is not difficult but increases notational burden without lending any additional insight. Thus, we chose to work with C2.

Condition C3: A respondent is declared as a winner if we are able to predict correctly his/her choice for the $q + 1^{\text{th}}$ choice question.

Condition C4: If there are multiple winners or if there are no winners then a respondent selected at random is will receive $300.

Conditions C3 and C4 are not restrictive.

Assumptions:

A1: Each respondent has two strategies:
S1: 'Consistent'
S2: 'Inconsistent'

Under the consistent strategy, a respondent answers every choice question consistent with some arbitrary partworth vector. Under the inconsistent strategy, a respondent answers every choice question by picking an alternative at random.

Note that A1 implicitly assumes that a respondent decides whether to be 'consistent' or 'inconsistent' at the beginning of the study and sticks with that choice for all choice questions.

An enhanced strategy space can be specified by assuming that respondents choose between being consistent and inconsistent for each choice question separately. Accommodating the above enhanced strategy space would require us to suitably change our second assumption (see A2 below). Modifying A2 to accommodate the enhanced strategy space could be done along the following lines:

(a) Define degree of consistency for the i^{th} respondent, say C_i, as the "No of choice questions for which that respondent chooses a consistent strategy" By definition, C_i lies between 0 and $q+1$.
(b) Replace A2 by: "p_i increases as C_i increases" where p_i is the probability that we can correctly predict respondent i's choice to the $q+1$th question.

Our conclusion about equilibrium behavior does not change we use (a) and (b) instead of assumptions A1 and A2. Thus, we use the simpler assumption A1. Assumption A1, while restrictive, is not critical.

Let, p_c be the probability of a correct prediction to the $q+1^{\text{th}}$ question correctly when a respondent uses the 'consistent' strategy. Similarly, let, p_{ic} be the probability of a correct prediction to the $q+1^{\text{th}}$ question correctly when a respondent uses the 'inconsistent' strategy. Finally, let β_i be the true partworth vector for the ith respondent.

A2: $p_c > p_{ic}$
A3: p_c and p_{ic} do not depend on β_i,

Assumptions A2 and A3 are the critical assumptions.

A2 simply states that the probability of a correct prediction is greater when a respondent uses a consistent strategy. We believe A2 is a reasonable assumption as we follow standard statistical methods of estimation (i.e., HB estimation).

Consider A3. When a respondent uses the inconsistent strategy the probability of a correct prediction cannot be dependent on β_i because the answers to the choice questions are not dependent on the partworth vector.

In order to see why A3 is plausible when a respondent uses a consistent strategy, consider an analogous situation involving linear regression where we have $q + 1$ observations. We want to assess how closely we can predict the q+1$^{\text{th}}$ prediction using the first q observations just as we want to predict the $q + 1^{\text{th}}$ choice using the q^{th} answers in the CBC study. The OLS estimator is:

$$\hat{\beta} = (X'X)^{-1}X'Y \tag{E1}$$

Where

X and Y are of appropriate dimensions.

Thus, the predicted observation is:

$$\hat{y}_{q+1} = x_{q+1}\hat{\beta}. \tag{E2}$$

Whereas the observed value is:

$$y_{q+1} = x_{q+1}\beta \tag{E3}$$

Notice that, by analogy to the consistent strategy, there is no error in equation E3. Therefore, the probability that $\hat{y}_{q+1}$ is within $\pm\,\epsilon$ of y_{q+1} is given by:

$$P(-\epsilon \leq \hat{y}_{q+1} - y_{q+1} \leq \epsilon) \tag{E4}$$

The distribution of $d = \hat{y}_{q+1} - y_{q+1}$ is independent of the true β vector as the OLS estimate given by equation E1 is an unbiased estimate of the underlying β. Therefore, we observe that the probability given in equation E4 is not dependent on the underlying vector β.

A3 extends the above observation to the context of discrete choice models.

Analysis:

Let, p_i be the probability that the i^{th} respondent is declared a winner. Depending on the strategy used by the i^{th} respondent p_i is equal to either p_c or p_{ic}. The expected utility for the i^{th} respondent, u_i, can be calculated considering the following three scenarios:

Scenario 1: Only the i^{th} respondent is the winner or
Scenario 2: Both respondents are winners or
Scenario 3: Neither respondent is a winner.

Using the above logic for both respondents, we have

$$u_1 = p_1 (1 - p_2) U_1 + p_1 p_2 \frac{U_1}{2} + (1 - p_1)(1 - p_2) \frac{U_1}{2} \quad \text{(E5)}$$

$$u_2 = p_2 (1 - p_1) U_2 + p_1 p_2 \frac{U_2}{2} + (1 - p_1)(1 - p_2) \frac{U_2}{2} \quad \text{(E6)}$$

Where,

U_i: Utility of \$300 for respondent i.

Simplifying equations E5 and E6, we get:

$$u_1 = \left(\frac{1}{2} + \frac{p_1 - p_2}{2}\right) U_1 \quad \text{(E7)}$$

$$u_2 = \left(\frac{1}{2} + \frac{p_2 - p_1}{2}\right) U_2 \quad \text{(E8)}$$

Using the ‘Consistent’ strategy is the unique Bayesian Nash Equilibrium given the expected payoffs given by equations E7 and E8. This follows from the following three observations:

1. If both respondents use the inconsistent strategy or if both respondents use the consistent strategy then the payoffs are:

$$u_1 = \frac{U_1}{2} \quad \text{(E9)}$$

$$u_2 = \frac{U_2}{2} \quad \text{(E10)}$$

2. However, when both respondents are using the inconsistent strategy, there is an incentive for one of the respondents to deviate. For example, suppose that respondent 1 deviates by using the consistent strategy when respondent 2 uses the inconsistent strategy then:

$$\tilde{u}_1 = \left(\frac{1}{2} + \frac{p_c - p_{ic}}{2}\right) U_1 \quad \text{(E11)}$$

Since, $p_c > p_{ic}$, it follows that $\tilde{u}_1 > u_1$ and respondent 1 will deviate. But, then the expected pay off for respondent 2 is given by:

$$\tilde{u}_2 = \left(\frac{1}{2} + \frac{p_{ic} - p_c}{2}\right) U_2 \tag{E12}$$

It follows that $\tilde{u}_2 < u_2$ and hence respondent 2 also deviates from being inconsistent to consistent.

3. When both respondents use the consistent strategy then neither respondent has an incentive to deviate as deviating reduces their expected payoff.

Therefore, it follows that using a consistent strategy is the Bayesian Nash Equilibrium for both respondents.