Intern. J. of Research in Marketing 27 (2010) 25–32

Contents lists available at ScienceDirect

Intern. J. of Research in Marketing

journal homepage: www.elsevier.com/locate/ijresmar

A simple mechanism to incentive-align conjoint experiments

Songting Dong [a], Min Ding [b,*], Joel Huber [c]

[a] School of Economics and Management, Tsinghua University, Beijing 100084, China
[b] Smeal College of Business, Pennsylvania State University, University Park, PA 16802, United States
[c] Fuqua School of Business, Duke University, Durham, NC 27708, United States

ARTICLE INFO

Article history:
First received in 3, February 2009 and was under review for 5 months

Area editor: Peter Danaher

Keywords:
Experimental economics
Conjoint analysis
Product development
Incentive alignment
Choice experiments

ABSTRACT

Recent literature has established the importance of incentive-aligning research participants in conjoint analysis. Pertinent studies have also proposed and validated a fairly general incentive-aligning mechanism (willingness-to-pay, or WTP) that achieves incentive alignment by using respondents' data to determine their value for a reward product (Ding, 2007). This mechanism, however, requires an estimation of the value of money and is relatively difficult for the average respondent to understand. We propose an alternative mechanism based on inferred rank order for situations where conjoint practitioners have more than one version of real products. In an empirical test of choice-based conjoint, we show that the RankOrder mechanism leads to substantial improvement in predictive performance when compared to non-aligned hypothetical choices. A second test shows that both incentive-aligned mechanisms – RankOrder and WTP – produce very similar predictive performances. RankOrder, however, dominates the WTP mechanism in user preference, an outcome shown both by perceived understanding and by the incentive-aligned money that respondents are willing to pay to switch from one mechanism to the other.

1. Introduction

Conjoint analysis is designed to uncover individuals' preferences across a range of alternatives defined by specific attributes (Carroll & Green, 1995). Since its introduction in 1971 (Green & Rao, 1971; Green & Srinivasan, 1990), conjoint analysis has become one of the most widely adopted marketing methods by researchers and practitioners (e.g., Baumgartner & Steiner, 2007; Camm, Cochran, Curry, & Kannan, 2006; Cattin & Wittink, 1982; Chen & Hausman, 2000; Jedidi & Zhang, 2002; Lynch, Buzas, & Berg, 1994; Vermeulen, Goos, & Vandebroek, 2008; Wittink & Cattin, 1989; Wittink, Vriens, & Burhenne, 1994; Wuyts, Verhoef, & Prins, 2009). New estimation methods designed to improve conjoint analysis – such as the hybrid model (Green & Krieger, 1996), Bayesian estimation (Allenby & Ginter, 1995; Kim, Menzefricke, & Feinberg, 2007; Otter, Tuchler, & Fruhwirth-Schnatter, 2004), polyhedral methods (Toubia, Hauser, & Simester, 2004; Toubia, Simester, Hauser, & Dahan, 2003), and partial conjoint profiles (Bradlow, Hu, & Ho, 2004) – continue to be developed.

More recently, work by Ding, Grewal, and Liechty (2005) shows that the performance of conjoint analysis can be improved substantially if the data are collected in an incentive-aligned manner. Building on principles from experimental economics (Smith, 1976), the authors of this study achieve incentive alignment by probabilistically rewarding a participant with an alternative chosen by him or her from a choice set during the conjoint exercise. In this paper, we call this incentive a *direct-alignment mechanism*. For ease of discussion, we will denote this incentive-aligned choice-based conjoint as *Direct* and the hypothetical choice-based conjoint as *Hypothetical*. A key implementation limitation of the Direct mechanism is that it requires that all product profiles shown in a conjoint study be available, as any one of them can be potentially awarded to a participant.

Since the 2005 study, an indirect method of implementing incentive alignment has been proposed and validated – one that requires only one reward to incent the respondents (Ding, 2007). Under the willingness-to-pay mechanism (WTP), the reward product is delivered to the respondent only if the respondent's willingness to pay derived from the conjoint is greater than or equal to a randomly determined price. If so, then the respondent must take the product at the randomly generated price. If the derived price is below the randomly generated price, the respondent is not permitted to buy the product. This procedure follows the BDM approach (Becker, DeGroot, & Marschak, 1964), except that it replaces the stated WTP with the inferred WTP from conjoint responses. It is shown that the mechanism motivates respondents to truthfully respond to conjoint questions so that the inferred price will be as close as possible to the true value they attach to the product.

Despite its benefit of requiring only one prize, this mechanism has its limitations. First, it requires the ability to infer a participant's willingness to pay for the reward product. Thus, it can only be used where price is an attribute and when the conjoint choices include an

* Corresponding author. Tel.: +1 814 865 622; fax: +1 814 865 3015.
E-mail addresses: dongst@sem.tsinghua.edu.cn (S. Dong), minding@psu.edu (M. Ding), Joel.Huber@duke.edu (J. Huber).

0167-8116/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijresmar.2009.09.004

outside good or a ‘no purchase’ option. Second, the WTP mechanism is difficult for many participants to understand. Although BDM has been used widely in experimental economics and has proved useful in aligning incentives, it has been used largely with student subjects who are accustomed to following difficult procedures. In contrast, the average commercial participant may be unwilling or unable to understand the meaning of a randomized price or why it provides a reason to be truthful.

While the WTP mechanism appears to be the current state-of-the-art incentive-aligned conjoint measure when the experimenter has only one version of a real product, the question in this paper is whether researchers can benefit from an alternative method when several different versions of products are available. Thus, we propose *RankOrder* as an alternative indirect incentive mechanism. The RankOrder mechanism also uses values generated by the respondent's conjoint, but instead of predicting an individual's monetary value for one product, it predicts preferences for a list of reward products and gives or sells the top-rated one to the respondent. As a result, the respondent is incented to attend to the task and provide truthful responses. In an empirical test, we show that the RankOrder mechanism leads to substantial predictive improvements compared to the Hypothetical mechanism. In a second study, we replicate earlier work showing that various incentive-aligned mechanisms have similar predictive performance. However, we also show that RankOrder dominates WTP in user preference. In a novel incentive-aligned assessment, an overwhelming majority of respondents were willing to pay real money to be switched from the WTP to the RankOrder mechanism but not the other way around.

We expect RankOrder to do better than Hypothetical because past tests of incentive-aligned methods against hypothetical ones have demonstrated consistent predictive benefits. Table 1 gives the result of eight independent tests of Hypothetical against incentive-aligned methods. The incentive-aligned method is more effective than the hypothetical in seven cases and equal in one case. It generates an average percentage gain in hit rates from 25% to 34% (i.e., 48% improvement). We do, however, expect WTP to suffer relative to RankOrder with respect to the perceived understanding of the mechanism and the willingness to pay to switch across mechanisms.

Notice that while most of the studies shown involve the Direct mechanism, two, involving iPods, use the WTP mechanism. Surprisingly, these WTP studies appear to do somewhat better against the Hypothetical mechanism than the others, even though the sizes of holdout tasks vary across studies. Thus, prior research does not provide a clear expectation as to whether WTP will do better or worse than the new RankOrder mechanism.

The rest of the paper is organized as follows. First, we describe the new RankOrder mechanism and present the two empirical studies that contrast it with the Hypothetical and the WTP mechanisms. The contrast between RankOrder and WTP features an incentive-aligned assessment of how much respondents are willing to pay to avoid one mechanism over another. Finally, we synthesize the findings with extant literature and provide recommendations as to when each conjoint collection process is appropriate.

2. The RankOrder incentive mechanism

The RankOrder mechanism proceeds as follows. First, before the experiment, participants are told that there is a list of possible reward products. Their answers in the conjoint task will be used to infer their own preference (e.g., A>D>B>C) regarding the products on the list (e.g., {A, B, C, D}). The respondent randomly selected as a winner will receive the top-ranked product as determined by his or her conjoint responses. Thus, the respondent has a motive to answer the questions carefully and truthfully.

There are two important differences between the proposed mechanism and Ding's (2007) WTP mechanism. First, RankOrder replaces the one reward in the WTP mechanism with a list containing at least two alternatives. Second, instead of estimating the price the respondent would pay for the one item, the respondent gets the product of his or her preference model put as the most preferred on the reward list. Fig. 1 graphically depicts the proposed mechanism and contrasts it with Ding (2007).

As shown in Table 2, the mechanisms differ in terms of the contexts in which they are applicable. While Hypothetical does not require any real product, WTP requires one, RankOrder requires a few, and Direct requires the availability of all product profiles shown in the conjoint choices. WTP is the only method that always requires both price as a variable and the ability of the respondent to opt out of a choice set by indicating that none of the items is desirable.

3. Study 1: contrasting RankOrder against the Hypothetical mechanism

The goal of the first study was to test RankOrder against the Hypothetical mechanism in a context not involving money so that the WTP mechanism could not be used. We also sought a category where the conjoint choices were relevant to the potential participants so that they would care about the outcome. To find a category, we interviewed a small sample of potential participants and determined that an all-expense-paid weekend trip to a nearby big city provided an appropriate conjoint topic. As the university is located in a small college town in a rural area, such a trip was quite appealing to the students. We framed the study as an effort to design an attractive weekend excursion package that rewards one lucky participant with an all-expenses-paid trip.

3.1. Experiment

To determine the appropriate attributes and levels for a weekend vacation, we conducted research over the web, spoke with individuals familiar with the city, and conducted a focus group of potential participants. We identified seven attributes, each with three

Table 1
Tests of predictive accuracy of hypothetical vs. incentive-aligned conjoint.

Product category/mechanism/payoff lag	Reference	Number of alternatives in holdout question	Hit rate of incentive- aligned conjoint	Hit rate of hypothetical conjoint	Percentage improvement in hit rate
Chinese dinner/direct/immediate	Ding et al. (2005)	20+no-choice option	48%	26%	85%
Snack combo/direct/immediate	Ding et al. (2005)	30+no-choice option	18%	13%	38%
iPod nano package/WTP/immediate	Ding (2007)	16+no-choice option	36%	17%	112%
iPod shuffle package/ WTP/immediate	Ding (2007)	16+no-choice option	34%	21%	62%
Bahamas cruise/direct/immediate	Ding et al. (2009)	10	39%	30%	30%
Bahamas cruise/direct/two weeks after conjoint	Ding et al. (2009)	10	35%	24%	46%
Camcorder/direct/immediate	Ding et al. (2009)	10	33%	29%	14%
Camcorder/direct/two weeks after conjoint	Ding et al. (2009)	10	32%	32%	0%
Weighted average			34%	25%	48%

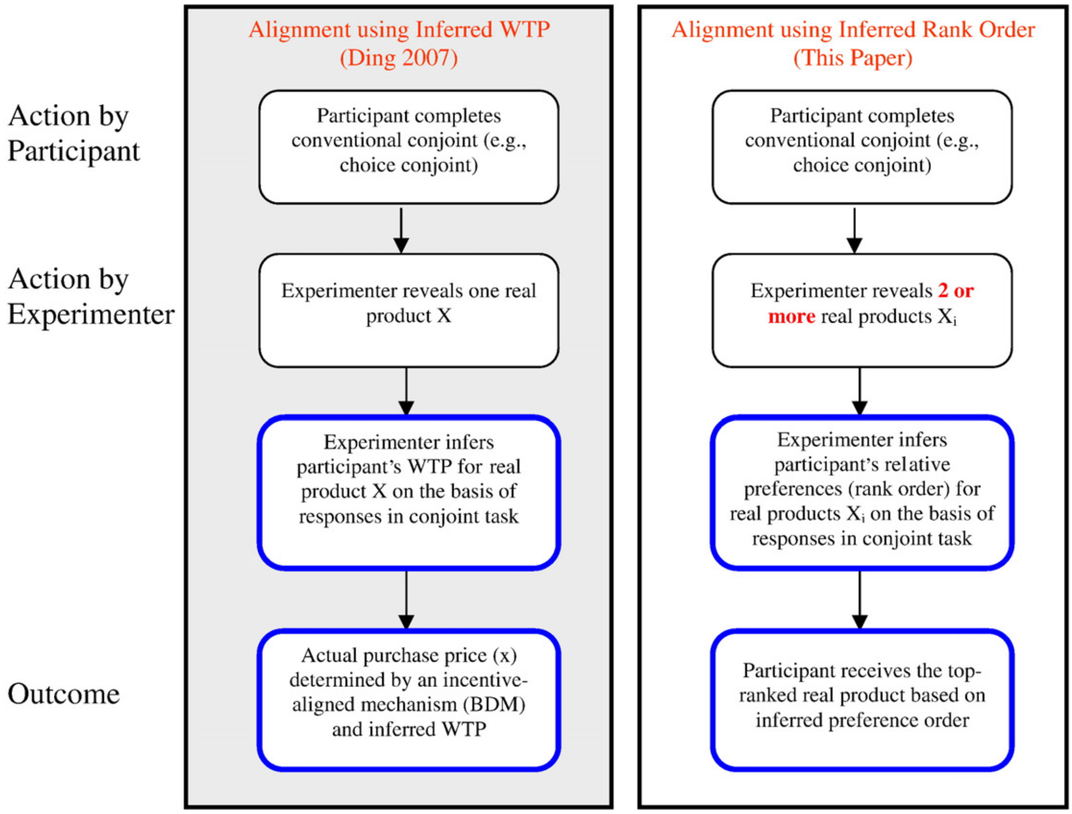

Fig. 1. Comparing rank order-based mechanism with WTP-based mechanism.

levels: a map system for navigating the city (paper map, Garmin eTrex GPS (small screen),and Magellan RoadMate GPS (large screen)); one ticket to see a Broadway show (3rd tier seat, 2nd tier seat, and orchestra seat); Saturday dinner (dinner only, dinner next to live jazz band, dinner next to live jazz band and free access to dance floor); Saturday night clubbing (free access to 1, 3, or 5 clubs of their choice); Saturday night accommodation at a 3-star hotel (room only, room and free access to health club, room and free access to health club and free in-room internet access); spa treatment (30, 60, or 90 min); and Sunday museum tours (ticket to 1, 3, or 5 museums of their choice). Throughout the study, participants had available to them detailed descriptions of these attributes, including pictures of the two GPS devices, the seating chart for the theatre, and a comprehensive list of all museums in the city.

SAS experimental design macros (Kuhfeld, 2007) guided the design of the actual profiles used in the empirical study. These 36 profiles produced by the macro were divided into 12 sets with 3 profiles for each conjoint choice set. We generated an additional 10 non-duplicate profiles, one of which was deemed dominant during pre-test and was eliminated. The remaining 9 profiles were used in the holdout task, which also served as the list for the incentive-aligning mechanism.

The experiment was a between-subjects design with two conditions, the Hypothetical and the RankOrder mechanisms. The holdout task was a Direct mechanism for all respondents, while the choice-based conjoint task was incentive-aligned only for the respondents in the RankOrder condition. In all cases, participants completed a choice-based conjoint with 12 triples, followed by the holdout task, where a participant selects his or her most preferred trip from a list of 9 different trips from which they might receive the item chosen and finally a brief survey of participants' demographics and attitude towards the study.

In addition to an $8 base compensation, participants were told that an actual trip would be awarded to one randomly selected participant from the entire pool of 85 participants. This Direct holdout task was real in the sense that all respondents had a chance to receive their choice. By contrast, those in the Hypothetical condition had a chance to receive the vacation package selected only in the holdout task. For those in the incentive condition, a coin flip determined whether their holdout choice or their conjoint task would determine the actual reward based on the RankOrder mechanism.

Table 2
Comparisons of Choice-Based Conjoint Data Collection Formats.

Mechanism	Requirements for conjoint estimation: Number of real product needed	Need 'none' option[a]	Must infer WTP
Hypothetical choice-based conjoint	None	No	No
Direct-incentive mechanism of choice-based conjoint	All profiles used in choices	No	No
WTP mechanism-based choice-based conjoint	One	Yes	Yes
RankOrder mechanism-based choice-based conjoint	A few	No	No

[a] Sometimes called "no-choice option" in the literature.

3.2. Estimation method

The estimation method that we employed is the standard approach used in conjoint analysis. Specifically, a hierarchical Bayesian multinomial logit model is used for estimation, similar to that specified by Allenby and Ginter (1995) and Allenby, Arora, and Ginter (1998). The probability that the ith participant chooses the jth alternative from the tth choice set is given by

$$\Pr(z_{it}=j)=\frac{\exp\left\{\beta_i^t d_{itj}\right\}}{\sum_{\ell}\exp\left\{\beta_i^t d_{it\ell}\right\}} \tag{1}$$

where z_{it} is the choice made by the ith participant in the tth choice set, $d_{it\ell}$ describes the ℓth option in the tth choice set evaluated by the ith participant, and β_t is a vector of partworths for the ith participant. We assume, a priori, $\beta_i \sim N(\bar{\beta}, \Lambda)$ and diffuse conjugate priors for $\bar{\beta}$ and Λ. Inferences were made after the convergence properties of the Markov chain Monte Carlo (MCMC) analysis were met. In addition, we tested a range of different prior values to ensure that the results were invariant to prior specifications.

3.3. Results

Table 3 provides the posterior mean and standard deviation of $\bar{\beta}$. If we consider a mean significant when zero is outside the 95% estimated confidence interval of the mean, we observe a greater number of significant attributes for those experiencing the Hypothetical conjoint. In that condition, all partworth means are greater than zero except for the room and health club and the access to 3 museums. Under RankOrder, three partworths that were significant under Hypothetical are no longer so. They are access to 3 or 5 clubs (over the baseline of 1 club) and access to 5 museums (over the baseline of 1 museum). The lack of significance of multiple clubs and museums has strong face validity, as it is unlikely that respondents will have time to visit multiple clubs and museums during this trip. The positive value for more choices in the Hypothetical condition is consistent with those respondents simply assuming that more is always better than less, whereas the RankOrder condition encourages respondents to think about the impact of these attributes on the trip.

The fact that RankOrder uncovers a somewhat different preference structure than that under the Hypothetical condition does not mean that this incentive-alignment mechanism is better. As Green and Srinivasan (1990) note, out-of-sample prediction provides an appropriate validation for conjoint methodology and thus is employed in this paper to judge whether an incentive-aligning mechanism adds value to conventional conjoint analysis.

The Hypothetical conjoint correctly predicted 10 (out of 41) participants' choices (24%) in the holdout task. On the other hand, the RankOrder incentive-aligned conjoint correctly predicted 18 (out of 44) participants' choices (41%).[1] Given that the literature reviewed earlier has shown incentive-aligned methods consistently outperforming the Hypothetical method, we conducted a one-tailed z-test of two independent sample proportions (following Fleiss, Levin, & Paik, 2003) to test whether RankOrder has greater predictive power than the Hypothetical mechanism. The p-value for the one-sided test is $p = 0.049$. Given the relatively small samples (41 and 44), this result provides evidence for the validity of the proposed incentive-aligning mechanism and its managerial relevance. Further, the average rank of choice predicted by RankOrder (2.295) is also smaller than that of Hypothetical (2.854), with a p-value of 0.052 that approximates the conventional 0.05 significance level. These results replicate support for incentive alignment in light of the consistent benefits of the incentive-aligned over hypothetical choices (Ding, 2007; Ding et al., 2005; Ding, Park, & Bradlow, 2009).

4. Study 2: Contrasting RankOrder against the WTP mechanism

Study 1 provided evidence that RankOrder is more accurate than the Hypothetical mechanism. In Study 2 we examine the relative performance of, and respondent reaction to, two incentive-aligned mechanisms, WTP and RankOrder.

4.1. Experimental design

The first design issue we addressed was the matter of what product categories we should use to compare these indirect incentive-aligned mechanisms. To test for robustness, we wanted both a low- and a high-priced product. Through interviews and focus groups similar to those conducted for Study 1, we identified a digital picture frame as an expensive and a T-shirt as an inexpensive but high-involvement category for students. The attributes used for the digital picture frame are based on the Best Buy website (bestbuy.com) and include six brands (Ality, Kodak, Philips, Westinghouse, Viewsonic, and Sony), six colors (white, gray, black, light brown, pink, and orange), two screen sizes (7-in. and 10-in.), two memory capacities (128 MB and 512 MB), three power sources (battery, AC, and both battery and AC), and multiple prices (from \$109 to \$189).[2] The custom-made T-shirt featured six different Chinese popular sayings (Perspective, Single, Beauty, BoyLoveGirl, Sweet Person, and Girl Power), six colors (white, black, pink, red, yellow, and gray), two fabrics (cotton and polyester/cotton blend), two necklines (jewel and V-neck), three sleeve options (short sleeve, sleeveless, and tank top), and three prices (\$3, \$4, and \$5). As in Study 1, SAS experimental design macros (Kuhfeld, 2007) generated the profiles used in the conjoint and holdout tasks.

The second issue we addressed was the actual structure of the study. The objectives of Study 2 were to contrast WTP against RankOrder on (1) their predictive performance and (2) user preferences. The first objective leads to a design similar to Study 1, which includes a specific version of a choice-based conjoint and one or more holdout tasks. Developing an incentive-aligned value to discriminate preferences between the mechanisms required a novel design.

Estimating an incentive-aligned value of the mechanisms required a two-wave study, where respondents were exposed to both mechanisms during the first wave that they could then modify in the second wave. At the end of the first wave, we asked each respondent to look ahead to a similar task a week later, when the mechanisms assigned for each category would be randomly determined. However, he or she could pay to get a different mechanism if dissatisfied with the one assigned. If the amount stated was greater than a randomly drawn but later revealed price, then we would switch the mechanism and the respondent's earnings would decrease by the revealed price. Each respondent indicated a willingness to pay to switch to RankOrder if assigned to WTP as well as a willingness to pay to switch to WTP if assigned to RankOrder. This mechanism – essentially the BDM procedure – is incentive-compatible, thus providing a rigorous measure of user preference in real dollar terms.

4.2. The experimental process

The first wave, completed in a campus computer lab, included (1) two pre-holdout tasks (one question for each product category, each with 19 options plus the option of no purchase); (2) two conjoint tasks (9 questions for each category, each with 4 options plus the option of no purchase); (3) two post-holdout tasks (one question for each product, each with 19 options plus the option of no purchase); and (4) two pay-to-switch tasks (one for each category) described above. Participants were randomly assigned to one of eight conditions ($2 \times 2 \times 2$): whether the pre-holdout occurs before or after the

[1] There are at least two approaches used to compute such predictive performance in the literature. One is to make a prediction at each iteration and then calculate the probability that the actual choice was predicted over all iterations for each individual. The second approach is to make a single prediction, e.g., using the posterior mean of partworth for each individual. The results from both approaches are usually consistent at the aggregate level across individuals. We report the predictive performance using the second approach here.

[2] We set three base price levels (\$109, \$129, and \$149) for the digital frames. In order to prevent dominant profiles, we added monetary value to two obviously more valuable features (i.e., the 10-in. screen is worth \$30 more than the 7-in. screen, and 512 MB of memory is worth \$10 more than 128 MB of memory) (this follows the practice in literature, e.g., Hauser, Toubia, Evgeniou, Silinskiai, & Befurt, 2009). The price differences (\$30 and \$10) were based on the differences of market price at the time of the experiment, given that other features remained the same. As a result, the study used prices ranging from \$109 to \$189.

Table 3
Mean and standard deviation of posterior distribution in Study 1.

Attribute		Without incentive alignment		With incentive alignment	
		Mean	Standard deviation	Mean	Standard deviation
Map	(baseline: paper)	0			
	Small GPS	**1.10**	0.47	**0.80**	0.37
	Large GPS	**0.91**	0.38	**1.44**	0.34
Broadway Show	(baseline: 3rd tier seat)	0			
	2nd tier seat	**0.77**	0.34	**0.96**	0.32
	Orchestra seat	**1.99**	0.33	**1.72**	0.28
Dinner	(baseline: food only)	0			
	Food and jazz	**1.52**	0.32	**0.68**	0.28
	Food, jazz and dance	**1.86**	0.31	**1.09**	0.30
Nightclub	(baseline: 1 club)	0			
	3 clubs	**1.01**	0.41	0.63	0.34
	5 clubs	**1.41**	0.34	0.18	0.29
Hotel	(baseline: room only)	0			
	Room and health club	0.31	0.44	0.49	0.28
	Room, health club and internet	**1.18**	0.36	**1.32**	0.35
Spa	(baseline: 30 min)	0			
	Spa 60min	**1.24**	0.39	**1.27**	0.29
	Spa 90min	**2.05**	0.41	**1.58**	0.40
Museum	(baseline: 1 museum)	0			
	3 museums	0.76	0.41	0.54	0.29
	5 museums	**1.04**	0.51	0.58	0.46

This table shows the posterior mean and standard deviation of the mean of the first-stage prior on the partworths.
A mean in bold indicates zero is outside its 95% confidence interval.
Features with dark background indicate that the mean under the incentive-alignment condition is outside the 95% confidence interval of the mean under the Hypothetical conjoint condition (without incentive alignment).

conjoint, the order between picture frame and T-Shirt in the conjoint task, and the assignment of mechanism to product category. Because these order conditions had no consistent impact on the accuracy of the methods, the results below pool across these manipulations. Note that every participant was exposed to both mechanisms and product classes. Following Wertenbroch and Skiera (2002), we informed respondents under WTP mechanism that the range used in BDM procedure covers all reasonable valuations, without actually revealing the range. In RankOrder mechanism, we informed the respondents that the prices of the products in the List are below or at the market price for the same product.

The second wave occurred one week after the first, completed away from the lab on any internet-linked computer. The second wave was simpler. It included (1) new conjoint tasks (18 questions for each product, each with 4 options plus the option of no purchase), and (2) new holdout tasks (2 questions for each category, each with 11 options plus the option of no purchase). Participants were paired with a different mechanism for each product based on the initial random assignment, except for the 28 respondents who were willing to pay enough money to switch mechanism.

All participants received a custom-made T-shirt and $18 from which was deducted the price of the T-shirt they received and, when appropriate, the price they paid for switching mechanisms in the second wave. In addition, two participants (1 in about 50) were randomly chosen to receive a reward in the form of a digital picture frame and the difference between $200 and the price of that picture frame. The specific T-shirt (for all) and digital picture frame (for the two winners) were determined as follows: (1) we first randomly picked a task from all the tasks they completed; (2) if the task was a holdout task, they would receive what they selected; (3) if the task was a conjoint task, they received the product and paid the price based on the specific mechanism described in the study (WTP or RankOrder). Specifically, under WTP, a random price is drawn from a distribution. If the randomly drawn price is smaller or equal to a respondent's inferred willingness to pay (based on the conjoint responses) for the product, the respondent will receive the product and pay the randomly generated price. Otherwise, the respondent keeps the money ($18 for the T-shirt, and $200 for the picture frame).

4.3. Results

A total of 94 participants provided complete and meaningful responses and were used in the analysis. We designed the pay-to-switch task as the last task in the first wave to provide a rigorous assessment of which incentive mechanism participants would like to use in real-life applications. For the digital picture frame, 61% of participants offered to pay us money to be switched from WTP to RankOrder, while only 18% offered to pay to be switched to the WTP mechanism, and 21% had no preference. Similar preference results were observed for the T-shirt. 58% of participants offered to pay us money to be switched from WTP to RankOrder, 14% sought to switch from RankOrder to WTP, and 28% had no preference.

We ran a logit analysis where the dependent variable is 1 if a respondent was willing to pay to switch to RankOrder and 0 if he or she was willing to switch to WTP. The independent variables were then the category and the order of the mechanism. The mean coefficient for RankOrder over WTP of 0.94 was strongly significant ($p<0.01$), while both category and order were non-significant ($p>0.20$). The means of the amounts of money respondents were willing to pay to switch from RankOrder to WTP or WTP to RankOrder were $2.97 (std = 2.47) and $3.18 (std = 2.54), respectively, for digital frames. For T-shirts, they were $1.71 (std = 2.00) and $1.83 (std = 1.51), respectively. None of the monetary differences between mechanisms was significant.

At the end of the first wave, we asked on 5-point scales anchored at *extremely easy* and *extremely difficult* on how easy it was to understand each of the mechanisms. Participants found RankOrder (mean = 1.85, std = 0.82) significantly ($p<0.01$) easier to understand than WTP (mean = 2.36, std = 0.75). They also felt that it was significantly ($p<0.01$) easier to understand why it was in their best interest to tell us their true preferences with RankOrder (mean = 1.74, std = 0.76) than with WTP (mean = 2.02, std = 0.80). These results show respondents' preference of RankOrder over WTP method is consistent with the fact that they think Rankorder is much easier to understand (both method itself as well as why they need to be truthful).

While we found strong differences in respondents' valuations and perceptions of the mechanisms, there were no differences in the accuracy of holdout prediction. We present the results below calibrated using conjoint responses from the first wave and predicting the choice in the two holdout tasks. In particular, the methods did not differ on the hit rates: RankOrder (choice hits) = 34%; WTP (choice hits) = 31%; and the rank of chosen alternative was approximately the same for RankOrder (avg. rank = 3.79, std = 3.67) and WTP (avg. rank = 3.88, std = 3.71).

To gain more confidence in these null results, for each accuracy criterion, we built a model of predictive accuracy as a function of the mechanism, product category, mechanism order, and holdout order. None of these was significant at the $p>0.20$ level. We also observed similar results when we incorporated predictive performance based on preferences inferred from the conjoint responses from the second wave. However, we downplay here predictive accuracy in the second wave because of a particular endogeneity problem. Since respondents had some control over the mechanism they might see in the second wave, or because they may have reacted to not receiving a change that they bid for, model accuracy might change in ways that are very hard to interpret.

In summary, despite a strong preference on the part of the respondents for RankOrder over WTP, there is no detectable difference in accuracy. As shown above, the attitude difference across mechanisms did not translate to difference in accuracy. To explore the relationship between accuracy and perceptions more deeply, we tested whether attitude towards a method affects accuracy by including individual terms of understanding and liking in the accuracy regressions above. The first regression included a measure of understanding, reflected in the average of two 5-point scales assessing the extent to which respondents understood the meaning of the incentive and if they understood why it motivates truthful behavior. This measure was non-significant for accuracy in predicting choice ($b = 0.06$, $p = 0.61$) for or the rank of the one alternative chosen ($b = 0.001$, $p = 0.97$). We also examined the positive dollar amount needed to switch methods, coded as negative if the person wishes to switch from one method and zero if the other way around. That measure of preference was also non-significant for accuracy, predicting accuracy of choice ($b = -0.01$, $p = 0.72$) and rank of the chosen alternative ($b = -0.01$, $p = 0.13$). Notice that this analysis implies that preference has no significant impact on accuracy either within or between mechanisms. In other words, the effects of the methods are robust even when respondents do not particularly like them or understand them.

While we had expected attitude towards the mechanism to be related to its accuracy, these results demonstrate that despite a strong preference on the part of the respondents for RankOrder over WTP, there is no detectable difference in accuracy.

We also assessed the time taken to answer the conjoint questions in each task and found them not significantly different at the $p<0.20$ level. However, based on this study, it is also clear that respondents overwhelmingly prefer the RankOrder mechanism over the WTP mechanism. There are several reasons why a respondent might prefer RankOrder over WTP. First, it is harder to intuit the randomization and price determination aspect of the BDM procedure. Second, RankOrder uses several versions in the reward list, whereas WTP employs only one. A respondent may feel a greater sense of excitement over the outcome under RankOrder, or at least his or her responses will help optimize the product he or she will receive, while under WTP, the product is already fixed. That inference may also lead them to believe that they will receive a product that is more valuable under RankOrder. Third, participants in the RankOrder condition who qualify for the prize always receive a product, while participants in the WTP condition may or may not receive an actual product. The perception that a participant may “fail” to win a product under WTP condition may have contributed to the preference for RankOrder.[3] Finally, respondents may perceive the expected payoff from RankOrder as more tangible and less likely to be disappointing since they receive the best in a set.[4] We suspect that these reasons have led to the dominant preference of the RankOrder over the WTP mechanism. However, the enjoyment of the mechanism and the preference for it interestingly had virtually no impact on its accuracy.

This insensitivity to preference for a task might be interpreted to suggest that a researcher can put a respondent through an unpopular task with a minimal cost in accuracy. However, we prefer to emphasize the conclusion that the RankOrder mechanism dominates in delivering greater respondent satisfaction at no cost in time taken or accuracy. This observation has important implications; a researcher can be expected to have an easier time recruiting and keeping participants with a more preferred mechanism and additionally may be able to pay less for their cooperation.

5. Conclusions and general discussion

Building upon current research on incentive-aligning mechanisms (Ding, 2007; Ding et al., 2005), this paper proposes an indirect-alignment mechanism based on inferred RankOrder that is more respondent-friendly and only requires that two or more different versions of the products be available. The first empirical study demonstrates that RankOrder is superior to the Hypothetical conjoint in a situation where the WTP mechanism cannot be used. The second study demonstrates that where both RankOrder and WTP mechanisms can be used, both mechanisms lead to similar predictive performance, but users overwhelmingly prefer the RankOrder mechanism over the WTP mechanism.

It is appropriate to speculate on why we found no difference in predictive accuracy between the two incentive-aligned mechanisms, despite respondents' strong preference for the RankOrder tasks. The main advantage of the incentive-aligned mechanisms is that they encourage the respondent to think about what it would be like to pay for and possess the items, be they T-shirts or digital frames. Both the WTP and the RankOrder mechanism encourage this attention shift from abstract liking to a focus on the price and use of the products. Accordingly, the value of the mechanism may depend less on a respondent's understanding why it works or liking how it works and more on a simple alignment of what happens in a conjoint exercise with what happens in the marketplace. We speculate that respondents are not sensitive to the magnitude of the incentive as long as it is within a reasonable range. For example, we expect that the impact on hit rates would be unchanged if the probability of getting the reward in our study shifted from 1 in 50 to 1 in 25. However, this hypothesis needs to be validated in future studies.

Combining the new results in this paper with the extant literature, Table 4 provides guidelines on when Hypothetical, Direct, WTP, or RankOrder mechanisms should be used in practice. The recommendations are organized based on the requirements for conjoint implementation (Table 1), and the recommended mechanism in each situation (i.e., each cell in Table 4) is determined based on (1) which mechanisms can satisfy the requirements for that cell, (2) which mechanism has the best predictive performance if there are multiple mechanisms satisfying the condition in a cell, and (3) which

[3] We thank an anonymous reviewer for this observation.

[4] We calculated the expected value of participating in the WTP method and the RankOrder method for each participant, treating their estimated preference as their true preferences and assuming they form their expectation, by randomly drawing large samples from various distributions and ignoring estimation errors. For the picture frame, we found that 14, 60, and 20 participants have higher, equal, and lower expected utility from participating in WTP compared to RankOrder, respectively. The corresponding numbers for the T-shirt are 18, 33, and 43. While this calculation used several strong assumptions, it is consistent with our conjecture that psychological reasons are likely driving the observed preference for the RankOrder method over the WTP method in this study.

Table 4
Recommended incentive-aligning mechanisms in different applications.

		Availability of different versions of real products			
		None	One	A few	All profiles used in a study
Whether WTP can be estimated	Yes		WTP	RankOrder	Direct
	No	Hypothetical			

mechanism is most preferred by participants given multiple mechanisms satisfying the conditions and each having equal predictive performance. Specifically:

- If no real products are available, only the Hypothetical mechanism can be used.
- If there is only one real product, the WTP mechanism should be used if price can be estimated from the preference measurement tasks and there is a 'none' alternative; otherwise, the Hypothetical mechanism is the only viable option.
- If two or more real versions of products under study are available, RankOrder should be used when WTP cannot be estimated. However, we still recommend RankOrder as the mechanism of choice even if WTP can be estimated, based on the results in Study 2.
- In the rare case where all product profiles under study are available and can be provided at reasonable cost (e.g., as the Chinese food in Ding et al., 2005, or when products are simple combinations of existing products, such as the iPOD gift package in Ding, 2007), we recommend that the direct-incentive mechanism (Direct) be used because it is the most straightforward incentive-alignment mechanism.

5.1. Implementation of the RankOrder mechanism

We now discuss caveats related to the implementation of the RankOrder mechanism and the implementation of incentive-aligning mechanisms in general. First, what products and how many products should be used on the reward list? In terms of motivation, it is important that the list contains prizes that reflect a broad set of desired and undesired products to motivate the respondent to care about the accuracy of the inferred conjoint model. Further, while it is important that the list not be revealed until after the conjoint, as the respondent may ignore attributes in the conjoint that do not differ on the list, from a motivational standpoint it may prove effective to give a sample list to emphasize possible regret at receiving an undesired product or missing out on a desirable product. Developing strategies for list generation provides a fruitful area for future research.

Second, it is important to maintain the credibility of the mechanism over the long term. Incentive alignment works so long as the respondents believe they are better off revealing their own true values. However, mechanisms are incentive-aligned only if the respondent cares about the outcome. For a single-stage game, incentive alignment holds regardless of the composition of the reward list, but consider what would happen if the reward products were always so undesirable that none of them would be liked or were so similar that there was no difference. Over time respondents would learn to ignore such motivational mechanisms. Thus, for incentive-aligned mechanisms to achieve a sustainable equilibrium in the field, it is important that the possible rewards include both desirable and undesirable alternatives. Thus, promising research could manipulate how the incentive-aligned mechanism is communicated to respondents and the extent to which a mechanism leads respondents to be more rather than less careful in subsequent surveys.

Third, incentive-aligned mechanisms make it more important to carefully select appropriate respondents. In particular it is important to select respondents interested in a category but to avoid those who have recently purchased a durable product. Such respondents who do not need the product may try to maximize their cash return or minimize their effort. Notice that this recommendation is different from the case of a hypothetical conjoint, where it is common to ask respondents how they would have acted had they been required to replace their current purchase.

Fourth, conjoint practitioners should take great effort to build a list to avoid the situation where the products in the list, given their associated prices, could all generate negative utilities. While this will not impact the data of that particular study, it will alienate these respondents and erode their trust in future conjoint studies using RankOrder mechanism.

Non-aligned methods can produce inaccurate inferences regardless of the sophistication of the analysis. Given the fact that almost all data collections done in the field are performed without incentive alignment, this domain promises to be a fruitful and rewarding area for researchers. Many questions remain to be answered; we focus below on two fruitful areas for future research.

First, incentive alignment is not restricted to conjoint analysis. All it requires is that the underlying data collection method be capable of inferring individual preferences, thus including self-explicated tasks and all versions of conjoint. There are situations, however, where these mechanisms cannot be used even if the data collection method does measure preference. For example, a really new product under development may not have any real products available as prizes. In other cases, participants in some situations (e.g., a buyer for a firm) have a different internal value than the buying center or may be prohibited from accepting any compensation from the seller. We believe that these contexts present important challenges that frame worthwhile research opportunities.

Second, this paper, like the extant literature on incentive alignment, uses lab-based out-of-sample predictions as its validation (Green & Srinivasan, 1990). It should be noted that actual market behavior is a much better measure of validity, although it is usually beyond the reach of academic researchers. Nevertheless, it will be valuable to validate various incentive-alignment mechanisms to forecast actual market behavior in the future or to backcast actual behavior in the past.

In conclusion, we have demonstrated that it is valuable and relatively easy for researchers to incorporate incentive alignment into their preference measurement methods. We hope the RankOrder aligning mechanism described in this paper will provide researchers with another tool to achieve this objective.

Acknowledgments

The authors acknowledge financial support from the Smeal College of Business, Pennsylvania State University; Fuqua School of Business; and, Duke University; and assistance from the Laboratory for Economic Management and Auctions (LEMA) at Pennsylvania State University.

References

Allenby, G. M., Arora, N., & Ginter, J. L. (1998). On the heterogeneity of demand. *Journal of Marketing Research*, *35*(3), 384–389.
Allenby, G. M., & Ginter, J. L. (1995). Using extremes to design products and segment markets. *Journal of Marketing Research*, *32*(4), 392–403.
Baumgartner, B., & Steiner, J. W. (2007). Are consumers heterogeneous in their preferences for odd and even prices? Findings from a choice-based conjoint study. *International Journal of Research in Marketing*, *24*(4), 312–323.
Becker, G. M., DeGroot, M. H., & Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral Science*, *9*(3), 226–232.
Bradlow, E. T., Hu, Y., & Ho, T. -H. (2004). A learning-based model for imputing missing levels in partial conjoint profiles. *Journal of Marketing Research*, *41*(4), 369–381.
Camm, J. D., Cochran, J. J., Curry, D. J., & Kannan, S. (2006). Conjoint optimization: An exact branch and bound algorithm for share-of-choice problem. *Management Science*, *52*(3), 435–447.
Carroll, J. D., & Green, P. E. (1995). Psychometric methods in marketing research: Part I, conjoint analysis. *Journal of Marketing Research*, *32*(4), 385–391.
Cattin, P., & Wittink, D. R. (1982). Commercial use of conjoint analysis: A survey. *Journal of Marketing*, *46*(3), 44–53.

Chen, K. D., & Hausman, W. H. (2000). Mathematical properties of the optimal product line selection problem using choice-based conjoint analysis. *Management Science*, *46*(2), 327–332.

Ding, M. (2007). An incentive-aligned mechanism for conjoint analysis. *Journal of Marketing Research*, *44*(2), 214–223.

Ding, M., Grewal, R., & Liechty, J. (2005). Incentive-aligned conjoint analysis. *Journal of Marketing Research*, *42*(1), 67–82.

Ding, M., Park, Y.H., & Bradlow, E. (2009). Barter Markets for Conjoint Analysis. *Management Science, 55*(6), 1003–1017.

Fleiss, J. L., Levin, B. A., & Paik, M. C. (2003). Statistical methods for rates and proportions, 3rd ed New York: John Wiley & Sons.

Green, P. E., & Krieger, A. M. (1996). Individualized hybrid models for conjoint analysis. *Management Science*, *42*(6), 850–867.

Green, P. E., & Srinivasan, V. (1990). Conjoint analysis in marketing: New developments with implications for research and practice. *Journal of Marketing*, *54*(4), 3–19.

Green, P. E., & Rao, V. R. (1971). Conjoint measurement for quantifying judgmental data. *Journal of Marketing Research*, *8*(3), 355–363.

Jedidi, K., & Zhang, Z. J. (2002). Augmenting conjoint analysis to estimate consumer reservation price. *Management Science*, *48*(10), 1350–1368.

Kim, J. G., Menzefricke, U., & Feinberg, F. M. (2007). Capturing flexible heterogeneous utility curves: A Bayesian spline approach. *Management Science*, *53*(2), 340–354.

Kuhfeld, W. (2007). Experimental design and choice macros Available at: support.sas.com/techsup/technote/ts722i.pdf.

Lynch, J. G., Jr., Buzas, T. E., & Berg, S. V. (1994). Regulatory measurement and evaluation of telephone service quality. *Management Science*, *40*(2), 169–194.

Otter, T., Tuchler, R., & Fruhwirth-Schnatter, S. (2004). Capturing consumer heterogeneity in metric conjoint analysis using Bayesian mixture models. *International Journal of Research in Marketing*, *21*(3), 285–297.

Smith, V. L. (1976). Experimental economics: Induced value theory. *American Economic Review*, *66*(2), 274–279.

Toubia, O., Hauser, J. R., & Simester, D. I. (2004). Polyhedral methods for adaptive choice-based conjoint analysis. *Journal of Marketing Research*, *41*(1), 116–131.

Toubia, O., Simester, D. I., Hauser, J. R., & Dahan, E. (2003). Fast polyhedral adaptive conjoint estimation. *Marketing Science*, *22*(3), 273–303.

Vermeulen, B., Goos, P., & Vandebroek, M. (2008). Models and optimal designs for conjoint choice experiments including a no-choice option. *International Journal of Research in Marketing*, *25*(2), 94–103.

Wertenbroch, K., & Skiera, B. (2002). Measuring consumers' willingness to pay at the point of purchase. *Journal of Marketing Research*, *39*(2), 228–241.

Wittink, D. R., & Cattin, P. (1989). Commercial use of conjoint analysis: An update. *Journal of Marketing*, *53*(3), 91–96.

Wittink, D. R., Vriens, M., & Burhenne, W. (1994). Commercial use of conjoint analysis in Europe. *International Journal of Research in Marketing*, *11*(1), 41–52.

Wuyts, S., Verhoef, P. C., & Prins, R. (2009). Partner selection in B2B information service markets. *International Journal of Research in Marketing*, *26*(1), 41–51.