An Article Submitted to

Review of Marketing Science

Manuscript 1141

Reference Dependence and Conjoint Analysis

Brennan Davis* Imran S. Currim† Rakesh K. Sarin‡

*Baylor University, brennan_davis@baylor.edu
†University of California - Irvine, iscurrim@uci.edu
‡University of California, Los Angeles, rakesh.sarin@anderson.ucla.edu

Reference Dependence and Conjoint Analysis

Brennan Davis, Imran S. Currim, and Rakesh K. Sarin

Abstract

Although there is enormous evidence that reference levels influence preferences, conjoint models, one of the most successful marketing research tools, assume that preferences depend on the absolute levels of attributes. In this paper we investigate the relevance of reference effects in two settings, compositional or self-explicated models in experimental studies 1 and 2, and decompositional or choice-based models in experimental study 3. In particular, we introduce a simple modification of the traditional self-explicated conjoint model which permits dependence of preference on reference levels. By eliciting gains and losses from expectations the model is adaptable to changes in respondents' reference points, which the traditional model is incapable of. Reference options are found to clearly affect subject choices in studies 1 and 2. In addition, the reference dependent self-explicated model is found to offer useful predictions when reference points are manipulated in study 1, and improve on predictions of its traditional counterpart when reference points are measured in study 2. In contrast, in study 3, the choice-based model's diagnostics and predictions are found to be robust to reference point manipulations. Taken together, these results suggest that the self-explicated model is more suited than the choice-based model to understanding and predicting how respondents make judgments relative to reference points because reference points and gains and losses from reference levels are more salient in the self-explicated model. We discuss implications for managers constructing conjoint models in product-market settings wherein reference points are changing due to new product introductions or marketing efforts.

KEYWORDS: Reference Dependence; Conjoint Analysis

1. Introduction

Conjoint models of consumer preference and choice continue to be one of the most successful tools of marketing research, providing useful inputs for product design, pricing, advertising, distribution, customer segmentation and targeting decisions. One fundamental assumption of the conjoint model is that preferences depend on the absolute levels of attributes. There is, however, enormous empirical and experimental evidence that reference levels influence preferences (e.g., Briesch, Krishnamurthi, Mazumdar and Raj 1997; Dholakia and Simonson 2005; Hardie, Johnson and Fader 1993; Kahneman and Tversky 1979; Kalyanaram and Winer 1995; Meyer and Johnson 1995). The key finding of this behavioral research is that the same attribute level has a greater impact on preference when it is perceived as a loss rather than as a gain.

The primary purpose of this paper is to show that a simple modification of the standard conjoint model captures that reference dependence[1]. Specifically, we modify the self-explicated approach for determining individual level part-worth utilities so that the elicited attribute ratings are for gains and losses from a reference level. A unique element of the resulting model is that it is useful in a variety of frequently and infrequently purchased consumer packaged goods product categories wherein managers are employing new marketing efforts to sway consumers away from their current reference levels particularly at the point of the purchase decision. The model is also useful for technology-based product categories (e.g., computers, flat panel televisions) in which price and feature reference points can change overnight due to new product introductions (Bridges, Yim and Briesch 1995). The traditional conjoint model and other current models that explicitly capture reference effects assume that the reference is fixed at the time the model is calibrated and consequently are unable to accommodate such subsequent changes in reference point. For example, the traditional self-explicated approach may be able to capture reference dependence through a kink in the value function at the reference point at the time of data collection. However, because the traditional method cannot separate out reference effects from non-reference based non-linearities such as threshold effects, it is difficult to judge whether the

[1] Traditional conjoint measurement focused on the estimation of partworth utilities for attributes based on preferences values for the bundle of attributes, which we label the decompositional approach (Green and Srinivasan 1977, Andrews, Ansari, and Currim 2002). A parallel literature developed on self-explicated utilities, generally based on the work by Fishbein and Ajzen (1975) on the theory of attitude measurement, and more specifically the work by Srinivasan and his colleagues (e.g., Srinivasan and Park 1997) on self-explicated conjoint measurement showing that consumers can provide relative values of attribute importance weights and desirability ratings, which we label the compositional approach.

kinks are reference-based. Most importantly, if the reference point changes at the point of the purchase decision it is not possible to judge how the consumer will react to features of a product given the change in reference point subsequent to calibration of the model.

Reference levels can be task related; for example, an intermediate option in a choice set may serve as a reference that influences the attractiveness of the extreme options. Non-task-related reference levels are either intrinsically held by individual consumers or suggested by marketers through advertising, promotions, shelf displays and placements, and other tactics. We propose a latent reference point model that accommodates both task related and non-task related reference points.

For frequently purchased product categories sold in supermarkets and drug stores, private label or store brands are often placed next to national brands and, in certain cases (e.g., over the counter medications), explicitly suggest a comparison to a specific national brand on the label of the store brand. For example, in Figure 1 the Rite Aid store brand label explicitly suggests a comparison to national brands such as Aquaphor and Eucerin. Rite Aid managers are aware that there will be several consumers who come to their store to purchase national brands such as Aquaphor or Eucerin for a variety of reasons. For example, some consumers may have purchased the national brand on several past occasions, so that it currently serves as their reference brand. Or their physician may have recently prescribed the national brand so that they enter the store with a predisposition to purchase it. By employing these shelf placement and labeling strategies managers are attempting to get such consumers to consider the store brand by focusing their attention on the price savings (or gains) relative to their reference national brand. A store manager who wishes to get a customer to switch from a national brand to the private label brand will need to understand how to compensate for the potential loss associated with moving from the national brand to the private label brand by a gain associated with price savings. Knowledge of the reference brand, and the relative gains and losses associated with switching products, will serve as important inputs into pricing, advertising and placement decisions.

For infrequently purchased product categories such as when buyers are customizing a Dell computer at the point of purchase, Dell indicates a base level which is included in the base price and highlights another potential reference level for each feature (e.g., speed, storage, warranty) that is presented as a most popular customer upgrade (see Figure 2). Higher (lower) levels are available at higher (lower) prices. Knowledge of the reference level and how customers react to levels above and below the reference level are important inputs into product presentation and pricing decisions. In addition, new product introductions lead to overnight changes in reference points so that a particular feature (e.g., price for a

certain screen size of flat panel television) that may have been viewed as a reference point a few months ago may now be viewed as a loss, resulting in changes in the value of options above and below the reference option.

In our first experimental study, study 1, we manipulate reference levels through *Consumer Reports* "Best Buy" recommendations, as a surrogate to how managers may promote their product in order to have it considered as a reference product. Thus, all subjects within a group are endowed with the same reference level. The manipulated reference levels, however, differ across two groups. We are interested in determining (a) if reference levels impact choices and (b) whether a conjoint model that explicitly incorporates reference levels is able to offer useful predictions. This is important because the traditional conjoint model cannot deal with changes in reference points subsequent to calibration of the model.

Reference levels can also be intrinsically held by individual consumers; for example, salary expectations of MBAs depend on the functional area desired by the individual (e.g., Finance, Marketing, Human Resources) and his/her past work experience. Our second experimental study, study 2, will use such specific individual reference levels for which we measure, rather than manipulate reference points. Our main interest is to investigate (a) if measured reference levels impact choices, and (b) whether such measured reference points are useful in predicting choices relative to traditional measurements in the self-explicated model which are not relative to a reference point. Such a comparison allows us to judge the usefulness of the modification to the data collection for the self-explicated model.

The use of experiments and the contrast between models and measurements is relevant for two reasons. First, identification of the reference point under the traditional approach can be challenging because it is not possible to distinguish reference effects from non reference-based non-linearities. Consequently, in experimental study 1 we manipulate the reference point and in experimental study 2 we directly elicit the reference point from the respondent so that identification is not required. Second, experimental contrasts between models and measurements control for omitted variables. It is important that the reference level is within the range of the attribute levels just like it is important that the attribute levels in the traditional conjoint model capture the range of availability in the marketplace. A key feature of our reference dependent conjoint model is that it is individual specific. This permits, for example, varying loss aversion across individuals as well as across attributes for a given individual. In other words all measurements of gains and losses are at the individual level so that there is no confound between heterogeneity and loss aversion (Bell and Lattin 2000).

While studies 1 and 2 focus on compositional or self-explicated conjoint models in which part-worth utilities are calculated based on self-assessments of

the importance of attributes and the desirability or value of attribute levels, study 3 focuses on decompositional or choice-based conjoint models in which part-worth utilities are estimated using statistical procedures. While we do not recommend a new reference-based version of the choice-based conjoint model, in experimental study 3 we are mainly interested if manipulations of reference points (as in study 1) impact part-worth utilities and holdout predictions of the choice-based model. This is important because it is informative about whether the choice-based model, which is not able to deal with reference point changes subsequent to model calibration, is at least able to diagnose changes in reference points prior to calibration. Taken together, studies 1, 2, and 3 have the potential to suggest whether the compositional self-explicated model or the decompositional choice-based conjoint model is more suited to understand and predict how respondents make judgments relative to reference levels.

2. Model

The reference dependent self-explicated conjoint model combines the simplicity of the self-explicated conjoint model with the behavioral observation that the preferences of an individual are influenced by reference levels. The part-worth of an attribute level is a constant number in the traditional self-explicated conjoint model and is given by:

[1] $$P_{ijk}^{s} = w_{ij}\, R_{ijk}\ (x_{jk})$$

in which, P_{ijk}^{s} = Individual i's part-worth for attribute j's k^{th} level in the self-explicated model, w_{ij} = Individual i's importance weight of attribute j, R_{ijk} = Individual i's desirability rating for attribute j's k^{th} level, and x_{jk} = Attribute j's k^{th} level.

In the reference dependent model, the part-worth of an attribute level depends on the reference level for that attribute. The carrier of desirability rating is not the absolute level of the attribute but its relative level compared to the reference level. Thus, a 32" flat screen television will have a positive desirability rating if the reference level for the screen size attribute is 26" or a negative desirability rating if the reference level for the screen size attribute is 38". Thus, the desirability rating, R_{ijk}, is assessed for gains or losses from the reference level.

Suppose r_j is the reference level for attribute j. Then, the part-worth calculation of Equation [1] is modified by eliciting the desirability rating

associated with a $(x_{jk} - r_j)$ gain if $x_{jk} > r_j$, or a $(r_j - x_{jk})$ loss if $r_j > x_{jk}$. In our example, the desirability rating for a 32" screen is no longer meaningful. Instead, we elicit the desirability rating of a gain by moving from a 26" screen to a 32" screen for a reference level of 26" or the desirability rating of a loss by moving from a 38" screen to a 32" screen for a reference level of 32". The desirability rating is set to the neutral level of zero when $x_{jk} = r_j$. We now write the part-worth of an attribute level for the reference dependent model.

[2] $$P_{ijk}^{r} = w_{ij} R_{ijk} (x_{jk} - r_{ij}),$$

in which r_j is individual i's reference level for the jth attribute. We note that the importance weights for attributes such as price and quality do not depend on the reference levels. The impact of reference levels on preference is captured by associating desirability ratings with gains and losses relative to the specified reference levels. This simplifying assumption is consistent with Tversky and Kahneman (1991).

In equation [2], we define the desirability ratings on deviations from reference levels $(x_{jk} - r_j)$. Thus, one can elicit desirability ratings of the price attribute for a $10 gain or loss from a reference price. For qualitative variables, such as compact, midsize, or large rental cars, one can elicit the desirability rating from receiving a large automobile rather than midsize at the same price or an undesirability rating of receiving a compact rather than midsize at the same price. In such cases, the minus sign reflects an exchange of r_j with x_{jk} and not a subtraction. We propose that categorical attribute levels be ordered by the participant from most to least preferred. If the reference point is manipulated, participants rate gains and losses given that they expect the manipulated level as in study 1. If reference points are not manipulated, participants indicate their expectation and rate gains and losses from the reference point as in study 2. The key difference between [1] and [2] is in how the desirability rating, R_{ijk}, is elicited. The desirability rating in the reference dependent model [2] will have the following theoretical properties.

2.1. Loss aversion

The desirability rating of a loss is theoretically expected to be more negative than a corresponding level of gain. Suppose the reference level for an attribute such as screen size is 32”. Loss aversion implies that if the desirability rating of a gain of 6" in screen size is 10 points, then the desirability rating of a loss of 6" in screen size may be -20 points. This will result in a concave utility

function for screen size since the desirability of moving from a 26” screen size to a 32” is greater than the desirability of moving from a 32” to 38”. Concave utility functions are often observed as a result of preference, in this example for a 32”, which influences the creation of a reference point for judging the desirability of other product features or screen sizes. In our applied measurement settings in both experimental studies 1 and 2 we do not require loss aversion as the desirability ratings for all gain and loss levels are directly assessed.

2.2. Diminishing sensitivity in gains

The theory suggests that increasing levels of gains generally have a diminishing impact on desirability ratings. Thus, if a gain of 6" in screen size is assigned a desirability rating of 10 points, then a further increase of 6" is likely to have an increase in the desirability rating of less than 10 points. Concavity of desirability ratings (value function) will produce diminishing sensitivity. Again, in our applied measurement settings in both experimental studies 1 and 2 we do not require diminishing sensitivity in gains as the desirability ratings for all gain levels are directly assessed.

2.3. Diminishing sensitivity in losses

Theoretically, the initial level of loss in reference level has the largest impact on desirability ratings. Further losses are likely to have a diminishing impact on desirability ratings. Convexity of desirability ratings in losses will produce diminishing sensitivity. Again, in our applied measurement settings in both experimental studies 1 and 2 we do not require diminishing sensitivity in losses as the desirability ratings for all loss levels are directly assessed.

2.4. Impact of Reference Levels on Preferences

We now illustrate the above behavioral properties and the impact of reference levels on preferences. Consider three alternatives A, B and C for flat screen televisions, with the following levels of two attributes of interest.

Alternative	Attribute 1 [Price]	Attribute 2 [Screen Size]
A	$1,399	38"
B	$1,099	32"
C	$799	26"

In [1], the traditional self-explicated conjoint model, assume that $w_1 = w_2 = 0.5$ and that the desirability rating is linear with the worst level of an attribute receiving 0 points and the best level receiving 10 points. Using model [1], the part-worths for attributes 1 and 2 for alternatives A, B and C are $(0,5)$, $(2.5,2.5)$ and $(5,0)$, respectively. Thus, A ~ B ~ C would each receive a total utility of 5 units.

In [2], the reference-dependent model, we again assume $w_1 = w_2 = 0.5$ (i.e., both attributes are equally important). Now, however, the desirability rating for an attribute depends on the reference level. For simplicity, assume that the desirability ratings for gains (preferred deviation from the reference level) as well as for losses (disfavored deviation from the reference level) are linear with a loss aversion of 2. Thus, if a gain of Δ units of an attribute receives a desirability rating of 10, then a loss of Δ units of the same attribute receives an (un)desirability rating of -20.

In our example, the desirability ratings for each attribute of the three alternatives vary depending on whether A, B, or C is regarded as the reference point. The desirability ratings for each of the three scenarios are shown below:

Reference Point	Alternative A		Alternative B		Alternative C	
	Price	Size	Price	Size	Price	Size
A	0	0	+10	-20	+20	-40
B	-20	+10	0	0	+10	-20
C	-40	+20	-20	+10	0	0

Thus, in contrast to model [1] that predicts A ~B ~ C; model [2] predicts:

$A \succ B \succ C$	Reference Point A
$B \succ C \sim A$	Reference Point B
$C \succ B \succ A$	Reference Point C.

2.5. Latent Reference Dependent Model

Simonson (1989) followed by a variety of other publications have demonstrated that the intermediate option (e.g., B in our example) is often deemed more attractive than an extreme option (e.g., A or C in our example). We propose a latent reference point model that accommodates both the task related reference point (i.e., the intermediate option) and the non-task related reference point (i.e., the level manipulated or levels intrinsically held by individual consumers)[2]. The underlying assumption is that the subject does not completely assimilate the manipulated or intrinsic reference level. The subject uses a *latent* reference level so that preference depends on both task related and non-task related preference levels.

Let m_j be the intermediate level of attribute j. Now, the implicit reference dependent part-worth is given by:

[3] $$P_{ijk}^{c} = w_{ij} R_{ijk} (x_{jk} - m_j).$$

In our example, regardless of whether option A or option C is manipulated to be the reference point, model [3] predicts that B $\succ$ A ~ C. In the latent reference dependent model, the part-worth of attribute j's k^{th} level is a weighted sum of the intermediate level part-worth [3] and the non-task related part-worth [2]. Thus,

[4] $$P_{ijk}^{\text{latent}} = \lambda_j P_{ijk}^{c} + (1-\lambda_j) P_{ijk}^{r},\ 0 \le \lambda_j \le 1.$$

Clearly, if $\lambda_j = 1$, then the level m_j is regarded as a reference level for attribute j. On the other hand, if $\lambda_j = 0$, then the manipulated or measured reference level determines the part-worth. The weights, λ_j's, are likely to vary across attributes.

In [4], the latent reference dependent model, the weights, λ_j's, do not vary across individuals. It is easy to modify [4] to permit variation of λ_j's across individuals. The assessment burden will, however, increase and subjects may have to provide rankings of options across several triplets. In the simpler model proposed above, the best fitting λ_j's are obtained by maximizing correct predictions across all subjects. Thus, each subject could provide their preference

[2] This proposition is conceptually similar to Mazumdar and Papatla (2000) who consider the relative role of internal and external reference prices in brand choice decisions calibrated using scanner panel data. We thank a reviewer for bringing this paper to our attention.

(ranking or the most preferred choice) for just one triplet. It is an empirical question whether the additional flexibility of permitting λ_j's to vary across individuals provides more accurate predictions. The basic idea of the latent reference dependent model is that the attractiveness of an option is influenced both by the intermediate (m_j) and the manipulated or measured (r_j) reference levels. This latent reference dependent model is empirically tested in Experiments 1 and 2.

3. Data collection

The data collection for the latent reference dependent model closely follows that of the traditional self-explicated method of Srinivasan and Park (1997) and the weighting and rating scheme of Winterfeldt and Edwards (1996).

Step 1: Attribute levels. For each attribute, identify the least and most preferred levels. Attribute levels that are considered unacceptable are eliminated from consideration. An option with an unacceptable level for any of the attributes will not be chosen regardless of how well the option scores on the remaining attributes.

Step 2: Attribute importance weights. These are determined using the swing weight method. The attribute that the individual most prefers to exchange the worst level with the best level is given a weight of 100. The weights for all other attributes are elicited relative to the most important attribute with weight 100. A visual scale for which the respondent marks the relative importance of each attribute may be helpful. The meaning of say 50 points for the weight of an attribute means that the individual regards the preference difference between the best and the worst levels of this attribute as half the preference difference between the best and the worst levels of the most important attribute. Therefore the normalized weight for the first attribute is 2/3 (100/150) and for the second attribute it is 1/3 (50/150). The swing weight method (based on the self-explicated approach), is conceptually similar to how weights (that sum to 1 or 100) are computed from traditional individual level conjoint analysis (based on the tradeoff or full profile approach) for each attribute by taking the swing in utilities for the attribute and dividing it by the sum of the swings in utilities over all attributes. Winterfeldt and Edwards (1986) discuss several successful applications of this method and demonstrate the robustness of the self-explicated weights compared to weights obtained using the tradeoff method. The weights are normalized between 0 and 1 without loss of generality. Like traditional individual level conjoint measurement (based on the tradeoff or full profile approach) computed utilities for options are invariant to scaling.

Step 3: Desirability ratings. The previous two steps of data collection are identical to those in Srinivasan and Park (1997). Desirability ratings for various levels of an attribute, however, require a slight modification in assessment. The carrier of desirability is the deviation from the reference level. A gain of Δ units (e.g., a $300 reduction in price) is assigned a rating of 10. The individual then provides the (un)desirability rating for a loss of Δ units (e.g., a $300 increase in price). Suppose the elicited rating for a loss of $300 is -20; one can then elicit the rating for a gain of 2Δ (e.g., a $600 reduction in price). Suppose the rating for a gain of 2Δ is 18 points. Similarly, one can elicit the (un)desirability rating for a loss of 2Δ (e.g., a $600 increase in price). Suppose the rating for a loss of 2Δ is -30 points. Now the ratings vary from -30 to +18 for a loss of 2Δ to a gain of 2Δ. These ratings are normalized between 0 and 1, so that the normalized values for ratings of -30, -20, 0, 10 and 18 will be 0, 10/48, 30/48, 40/48 and 1 respectively. Thus, the largest negative deviation from the reference level receives a rating of 0 and the largest positive deviation from the reference level receives a rating of 1. Computed utilities for options are invariant to normalization.

The main purpose of this measurement procedure is to elicit gains and losses from reference levels so that the model is capable of adapting to changes in reference levels or expectations which the traditional self-explicated or decompositional conjoint models are incapable of. The assessment is simplified if reference levels are intrinsically held by each consumer (our study 2). In this case, the subject first specifies the expectations or reference levels. Desirability ratings are then elicited for deviations from these reference levels. Another simplification is to explore whether a linear form with a constant loss aversion is justified for some attributes. Such a simplification may be possible for quantitative attributes such as price.

Step 3 follows the principles outlined in Kahneman and Tversky (1979), the certainty method of Currim and Sarin (1989), and the rating scheme of Winterfeldt and Edwards (1996). The succinct, theory-based description of measurement under step 3 above is followed by more practical and detailed descriptions under the measurement sub-sections for experimental studies 1 and 2. In addition, the web-based questionnaires for both studies are available for readers interested in measurement details.

In experimental studies 1 and 2 all measurements of steps 1, 2, and 3 are at the individual level. When holdout choices are predicted using the latent reference dependent model, we need to first determine the reference level for each individual so that the measurements from step 3 can be used to compute gains and losses from the appropriate reference level. The reference level is computed using equation (4) which is a linear combination of the intermediate or compromise level c of each attribute and the reference point r. In experiment 1 the reference level r is the one chosen for the 'best-buy' manipulation for all subjects while in

experiment 2 it is directly elicited from the respondent. In other words, the elicitation of desirability ratings in step 3 in both experimental studies 1 (in which reference points are manipulated) and 2 (in which reference points are directly elicited) relate to the second term on the right hand side of equation 4, so that if $\lambda_j = 0$, then the manipulated or measured reference level determines the part-worth.

4. Experimental Study 1: Effect of changes in reference points on choices

The purpose of study 1 is twofold. First, we investigate whether changing reference points affect observed choices. Second, we explore whether these observed choices can be predicted by a conjoint model that unambiguously incorporates reference dependence. The second question is important because a traditional conjoint model that does not explicitly incorporate reference dependence is unable to predict a change in preference when reference points shift. That is, if a manager is interested in predicting the likely effect of changing consumers' reference points on part-worth utilities, then it is important that reference dependence be directly incorporated into the model.

We ask these two questions because reference points are often evolving in markets. For example, in high technology markets (e.g., computers and flat panel televisions), new product introductions continuously modify consumer reference points on price and quality so that what was previously thought to be better than the norm may now be judged to be the norm or less than the norm. And in low technology markets, managers often employ advertising or other communications to make their products (e.g., Bayer's Aspirin) and services (e.g., General Motors' Goodwrench auto repair) the standard, or reference, so that deviations from such products or services on quality are viewed as much less desirable. Consequently, we are interested in investigating whether a traditional self-explicated conjoint model that explicitly incorporates reference dependence and has been calibrated prior to a change in reference points will be able to predict individual choices made subsequent to a change in reference points.

4.1. Subjects, products, attributes and levels

The subjects were 124 MBA students at a public West Coast university. The survey was administered online via a Web-based questionnaire. Students completed the survey anonymously and were offered course credit for their participation after receiving a code to be redeemed upon completion. Students were randomly assigned online to one of two groups. The questionnaires used to

elicit the value function and the importance weights for both groups were identical; therefore, part-worths were determined identically for both groups[3].

We selected two products in the consumer electronics category as choice stimuli: speakers and flat-screen televisions. For speakers, we considered price and power as attributes, each with three levels. Price levels were $100, $160 and $220, while levels for power were 50, 100 and 150 watts. For flat-screen televisions, we considered price and screen size as attributes, each with three levels as well. Price levels were $799, $1099 and $1399, while screen size levels were 26, 32 and 38 inches.

4.2. Measurement

We use a modified version of the traditional self-explicated method (Srinivasan and Park 1997) that allows us to compute reference dependent part-worths. In the traditional approach, two types of data are collected for each attribute: (i) the value function and (ii) the importance weight. These two data are used to calculate each individual subject's part-worth utilities for attribute levels. In addition, we collect holdout validation sample choices that are not used to calculate the part-worth utilities.

The value function is determined by collecting the desirability rating of each attribute level. The traditional method of data collection under the self-explicated approach is to ask the subject to rate their most-preferred attribute level at 10 and their least-preferred attribute level at 0 (after eliminating unacceptable levels) and subsequently to rate each remaining level (between 0 and 10). In order to explicitly incorporate expectations in the value function, this traditional method is modified as follows.

We ask the subject to imagine that they are considering buying products in a particular product category (e.g., speakers and LCDs) which vary on two attributes (e.g., product quality and price, the corresponding attributes and levels are presented in Figure 3). We further ask them to imagine that they have researched the particular product category prior to purchase and decided to make a purchase of a product with certain attribute levels. The subject is free to pick any attribute level in the ranges presented in Figure 3. For example, a particular subject might have conducted research on CNET.com and decided to purchase speakers rated at 100 watts of power for $160.

At the time of purchase, we ask them to suppose that they are able to get *exactly the same* product they expected. This gives them 0 points of satisfaction on a scale from 0 to 100. Zero points means that their expectations have been met. We then ask them to suppose that they discover at the time of purchase that they

[3] All questionnaires employed in the experiments reported in the paper are web-based and available from the authors on request.

are able to get a quality (or price) attribute level that is *X percent better* or *better by X* or *the next level better than expected* for the same price (or quality) for example speakers had 50 watts more power than expected for the same price, or the speaker system was $60 less in price than expected for the same amount of power.[4] This gives 10 satisfaction points on a scale from 0 to 100. Finally, we ask them to suppose that they discover at the time of purchase that they are only able to get a quality (or price) attribute level that is *X percent worse* or *worse by X* or *the next level worse than expected* for the same price (or quality), for example the speakers had 50 watts less power than expected for the same price, or were $60 more in price than expected for the same amount of power.[5] We then ask them to rate their dissatisfaction on a scale from 0 to 100. Hence, we directly elicit the value function, or the satisfaction or dissatisfaction, associated with gains and losses relative to a reference point or expectation in the spirit of the self-explicated approach. The levels of the attribute used were the same as those listed in the preceding section.

To measure importance weights, we employ the traditional self-explicated method by identifying the attribute for which improving from the lowest to the highest level is most important to the subject. That attribute is considered the *critical attribute* and is given 100 importance points. The subject subsequently rates the remaining attributes from 0 to 100 relative to the importance of the critical attribute.

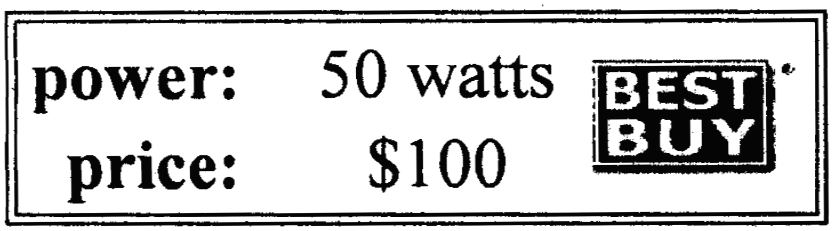
4.3. Manipulation and validation choices

After the calibration of the model but prior to the collection of holdout validation sample choices, we manipulated reference levels using a *Consumer Reports* Best Buy recommendation which read and was visually presented as follows: A recent Consumer Reports article identified the following product configuration as their BEST BUY recommendation.

---

[4] This often happens when new products are introduced between the time of the last review (e.g., CNET) and the shopping trip. For some attributes such as brand or color, levels are not monotonically increasing or decreasing in preference. In such cases, we can simply ask the subject to suppose only that they discover at the time of purchase that they are able to get the next brand or color (level) at the same price.

[5] This often happens because of stock-outs.



There were three options in the validation sample (Figure 3). That is, the subjects in Group 1 were told that *Consumer Reports*' Best Buy recommendation for speakers is option A. For LCD televisions, they were presented with the reference point as the opposite extreme so that option C was the Best Buy recommendation. In other words, for Group 1 option A served as a reference point for speakers and option C for LCD televisions. The subjects were aware of these three options and therefore knew throughout the validation sample part of the questionnaire that B is an intermediate option. For Group 2, we presented the opposite: option C served as a reference point for speakers and option A for LCD televisions. All subjects, regardless of the group they belonged to, specified their preferred option from the triplet A, B, C. Subjects also chose the most preferred option from each of the three pairs (AC, AB and BC).

4.4. Results

Loss aversion was observed for each attribute for both speakers and LCD televisions. The average loss aversion was approximately 3. For example, the induced satisfaction rating of +10 was associated with a price that is $60 less than expected for the same power and an average dissatisfaction rating of -30.5 was associated with a price that is $60 more than expected for the same power (average loss aversion = 3.05).

We did not observe concavity of value function for gains in our data. The jump in average satisfaction rating from 50 watts more than expected to 100 watts more than expected is larger than the jump in average rating from expected to 50 watts more than expected. These results could be due to a "peanut" effect as a gain of 50 watts (for speakers) or 4 inches (for LCD televisions) may not be regarded as significant by the subjects. The average value function for losses was either linear or convex and, for a great majority of subjects, it was steeper in losses than in gains.

The observed choices of each group for each product reveal a pattern that reflects reference dependence. In a standard conjoint model, the choice between options A and C is invariant whether option A or C is manipulated to be a reference option. The data reveal that the reference option is chosen

approximately two-thirds of the time. Thus, when option A is specified as the reference option, then a great majority chooses option A over C; and, when option C is specified as the reference option, then a great majority chooses option C over A. The table below shows the percentage of subjects who chose the reference option (indicated in brackets) for each case.

	Group 1	**Group 2**
Speakers	63% (A)	75% (C)
LCD TVs	65% (C)	63% (A)

Our data also indicates that the middle option may serve as the implicit reference point. Among the triplet A, B, C, the reference option, A or C, was most preferred in 44% of responses and the middle option, B, was most preferred in 41% of responses. The option that was neither the reference option nor the middle option was preferred in only 15% of responses. For the pairs in which the reference option is not included, AB if C is the reference option or BC if A is the reference option, loss aversion will predict a choice of option B. Indeed for such pairs, option B was preferred in 78% of responses.

Two conclusions can be drawn from the observed choices in the validation sample. One, the reference option has a clear influence on subjects' choices. Two, the intermediate option has an attraction of its own and may serve as an implicit reference point. It is possible that the reference option has a clear influence on subjects' choices because of a promotional-based effect, and not necessarily a reference-based effect. For this potential reason we will test whether a latent reference dependent model calibrated before the manipulation is presented to subjects is capable of predicting choices observed subsequent to the manipulation.

To test whether a latent reference dependent model can predict observed choices, we used the parts-worth based on gains and losses from the reference level to predict which option had the highest utility. The predicted choice was compared to the actual choice to determine the percentage of choices that were correctly predicted, compared with correct prediction that would be made by the simple chance criterion. The results are summarized in Table 1. The latent reference dependent model, model [4], is a useful predictor, better than chance in all cases ($p<0.01$).

Non-task related reference effects are of two types (a) changes in reference points resulting from managerial efforts to have their product considered as a reference; and (b) reference points held intrinsically by individual consumers. Study 1 focused on changes in reference points that may occur as a result of managerial efforts to have their product considered as a reference. We found that reference points clearly influence subjects' choices and a latent

reference dependent self-explicated model improves on chance predictions, admittedly not a strong criterion. Nonetheless, study 1 is important because (a) traditional conjoint models including the traditional self-explicated model are incapable of dealing with changes in reference points subsequent to the calibration of the model and (b) demonstration that a latent reference dependent model calibrated prior to the reference point manipulation is capable of offering useful predictions subsequent to the reference point manipulation shows that the explicit incorporation of reference effects in the model is useful in predicting choices resulting from subsequent managerial efforts to promote their products as reference products. In study 2, we focus on reference points intrinsically held by individual consumers and compare the predictions of a latent reference dependent self-explicated choice model to its traditional counterpart, a stronger criterion.

5. Experimental Study 2: Elicitation of reference points in multi-attribute settings

In study 2, we elicit intrinsically held reference points from each individual subject in a more complex multi-attribute setting. In order to do this, we asked each subject to state the value they expected for each attribute. We compared the results of the latent reference dependent model to the traditional self-explicated conjoint model. To enable the comparison without overloading the respondent, we designed treatment and control groups, which are described below.

5.1. Measurement and experimental design

The only difference between the measurements in the treatment and control groups was the elicitation of the value function. The treatment group's value function for each attribute level was assessed as in study 1 to unambiguously allow for reference dependence[6]. The control group's value function for each attribute level was collected via the desirability ratings of the traditional self-explicated conjoint model (Srinivasan and Park 1997). Importance weights were assessed as in Experiment 1. Holdout validation sample choices were also collected as in Experiment 1, but only for choices between triplets. Holdout questions 1, 2 and 3 required choices from triplets representing sets {A,B,C}, {B,C,D} and {C,D,E}, where option C is always the elicited reference point.

[6] For details the reader is referred to the questionnaire available on request from the authors.

5.2. Subjects, context, attributes and levels

The subjects were 129 MBA students at a public West Coast university. The survey was administered online via a Web-based questionnaire. Students completed the survey anonymously and were offered course credit for their participation after receiving a code to be redeemed upon completion.

The MBA students were randomly assigned to one of two groups: treatment and control. As noted above, the value function was based on gains and losses from the reference level for the treatment group (n=71). In the control group (n=58), the value function was based on desirability ratings as in a traditional self-explicated conjoint analysis. No gain or loss information was gathered from the control group.

In order to measure expectations that would be salient to the subjects, we presented the stimulus of job choice with eight attributes and multiple levels per attribute. As in Srinivasan and Park (1997), in order to maintain comparability to previous studies on MBA job choice (e.g., Wittink and Montgomery 1979), we used the same eight attributes and multiple levels per attribute, changing only two attributes slightly.[7]

We instructed each subject to consider a sealed envelope that had just arrived in the mail with a job offer. In the treatment group, each person was asked to state his/her expectation for each attribute before evaluating each level for all of the attributes. The eight attributes were salary, advancement opportunity, commute time, company growth, region, people, travel and functional activity (see attribute levels in Figure 4). We asked them to check the level of each attribute which best represented their expectations.

5.3. Results

The data on observed choices show that the explicit reference option, C, is chosen by a large percentage of subjects in each of the three choice sets. The summary of the percentages of correctly predicted choices using the latent reference dependent model versus the traditional self-explicated conjoint model is presented in Table 2. The latent reference dependent model performs better than the traditional self-explicated model ($p<0.05$).

The results from studies 1 and 2 are twofold. First, reference options have clear influence on subjects' choices in both studies. Study 1 subjects received the reference option as a manipulation similar in spirit to what may be attempted by a

[7] "Region of the U.S." was changed to "Region of California" to make the survey more relevant to the MBA students. "Local Environment" was changed to "Commute Time" to add another quantitative attribute.

manager seeking to promote his/her product as a reference product. In contrast, in study 2 reference points intrinsically held by each individual subject for each attribute of the job offering were elicited. This first result indicates that managerial efforts to promote their products as reference points and reference points held intrinsically by individual consumers can have a clear influence on consumer choices. Second, while the latent reference dependent model is found in study 1 to offer useful predictions relative to chance, admittedly not a strong criterion, in study 2 it is found to improve predictions relative to the traditional self-explicated model, a stronger criterion.

When the results of studies 1 and 2 were assessed one question that came up was whether changes in reference points would be reflected in the diagnostics (e.g., part-worth utilities) and predictions of a non-reference dependent choice-based conjoint model, one that did not explicitly incorporate reference points. In contrast to the compositional self-explicated model, which was the focus in studies 1 and 2, and computes part-worth utilities based on self-assessed desirability values and importance weights, the decompositional choice-based conjoint model estimates part-worth utilities from choices based on statistical procedures. The question that came up is important because if changes in reference points would be reflected in changes in diagnostics (e.g., part-worth utilities) and predictions of a non-reference dependent choice-based conjoint model then we could conclude that the choice-based model, which is incapable of adjusting to future reference point changes, at least reflects the effects of past efforts in changing reference points, presumably reducing the need to explicitly incorporate reference effects in conjoint models. Consequently, we conducted a third experimental study to address this question.

6. Experimental Study 3: Effect of manipulation of reference points on a choice-based conjoint model

We investigated whether a choice-based conjoint model that employs a Hierarchical Bayes (HB) estimation approach can capture the effects of reference point manipulations. Our goal was to determine whether there would be any predictive (e.g., holdout choices) or diagnostic (e.g., part-worth utilities) differences between choice-based conjoint models estimated in conditions with and without reference point manipulations[8].

[8] Identification of the reference point in decompositional conjoint measurement is difficult because one cannot separate reference effects from non-linearities such as threshold effects. In addition, decompositional partworths are combinations of attribute importance and desirability ratings. Consequently, it is inappropriate to validate decompositional partworths using desirability ratings. Rather than compare decompositional partworths to desirability ratings, we decided to compare

6.1. Measurement and experimental design

We used Sawtooth software to design our questions and estimate the choice-based models. We chose a credit card setting with 4 attributes, each with 2 or 3 levels: annual fee (none, $30, $60); interest rate (12%, 24%); frequent flyer program (yes, no); and credit line ($2000; $8000). We employed 2 reference point manipulations (1) high fees and high benefits ($60, 12%, yes, $8000) and (2) low fees and low benefits ($0, 24%, no, $2000). That is subjects were asked to assume that they currently owned and used a card that was either (a) high fees and benefits or (b) low fees and benefits, which they were considering replacing.

6.2. Subjects, context, attributes and levels

The subjects were 187 MBA students at a public West Coast university. The subjects were randomized into two groups A and B. In group A (B) the treatment group subjects (n=48 (47)) received the high (low) reference point manipulation while the control group (n=38 (54)) received no such manipulation but were given the same holdout validation questions.

6.3. Results

We found no statistically significant differences in part-worth utilities, attribute importance weights, or predictions (holdout hit rates or % of choices correctly predicted) between the treatment and the control groups[9]. As a result, for the purposes of our paper, the HB model is found to be invariant or robust across treatment and control conditions characterized by the presence and absence of reference point manipulations. In the choice-based procedure, a subject selects an option from an offered set of two alternatives. The pairs of options are generated by a computer program and are presented to the subject in a relatively rapid succession. The choices from a series of pairs of options may be less sensitive to the reference point because the reference point lurks only in the background and the feeling of gain or loss is not made vivid. Tversky, Sattath, and Slovic (1988) argue that in choice people focus on the more prominent attribute and do not make quantitative comparisons between gains and losses.

In the self-explicated approach, the subject is asked to evaluate the desirability of a specified gain or loss from a reference point and therefore the

---

decompositional partworths estimated under the manipulation condition to those estimated under the no manipulation condition.

[9] Details are available on request from the authors.

reference point is always salient. It is no surprise that the value functions in the choice-based procedure are relatively linear and do not exhibit a sharp kink at the reference level that separates gain and loss regions. For example, for annual fee, for which the utility of $60 (none) per year is set to 0 (1), the average utility for $30 per year is 0.46 (0.45) in the high (low) benefit high (low) price condition, while it is 0.46 in the control group condition.

The self-explicated method and the choice-based procedure are different elicitation procedures that highlight different aspects of options and suggest alternative heuristics that subject uses to construct his/her response. In consumer choice environments where a consumer is presented with a few options (which credit card should I choose from the options offered), a choice-based procedure provides a good representation of preferences. In contrast, consumer choice situations that demand a quantitative comparison of gains and losses from a reference point (should I pay a higher annual fee than what I am currently paying to obtain additional benefits), the reference dependent self-explicated model is likely to provide a more realistic representation of preferences.

7. Conclusions

Gains and losses from a reference level are important determinants of a consumer's preferences and therefore need to be modeled explicitly. We have also shown that a simple extension of a traditional conjoint model that incorporates dependence of preferences on reference levels improves predictive accuracy. Specifically, a latent reference dependent model in which a latent reference point that depends on task related reference points (i.e., an intermediate option) and non-task related reference points (i.e., changes in reference points due to managerial efforts to get their product consider as a reference, and intrinsically held reference points of individual consumers) is found to improve the predictive ability of the traditional self-explicated conjoint model. Our approach is particularly useful in frequently (e.g., Figure 1) and infrequently (e.g., Figure 2) consumer packaged goods product categories where marketers use new strategies to create new reference points for non-buyers in a product category (e.g., Figure 1) or manipulate reference points of product category buyers to enable brand-switching or product-upgrading within the brand (e.g., Figure 2), because traditional conjoint models including models that explicitly incorporate reference points cannot handle such managerial attempts to change reference points subsequent to the calibration of the model. Our approach is also useful in technology-based product categories in which innovation driven new product introductions can change reference points overnight so that what was considered as a reference product a few months ago may now be viewed as a loss.

The reference dependent self-explicated conjoint model places minimal demands on subjects. Only 1 of 3 data collection steps in the traditional self-explicated model, elicitation of desirability ratings, is modified. In the traditional model the least and most desired levels are set to 0 and 10 respectively, followed by a subject rating intermediate levels. The reference dependent version could require the respondent to identify the level they expect, which would be set to zero, the value of the next best level would be set at 10 points, followed by the subject rating remaining levels on a gain-loss scale. Such ratings may arguably be more realistic and valuable since the consumer has to think explicitly about and identify what they expect, and how they assess outcomes, which exceed or do not meet their expectations. This can relieve the analyst of the burden of predicting what the reference point of a consumer might be.

One limitation is that we did not collect any process measures to validate that our measurement of desirability gets at the reference point phenomenon. However, we validate the measurement approach for desirability by (a) presenting results on observed choices for each of two products (speakers, LCD TVs) in experimental study 1 (which *manipulated* the reference point) and the job-choice setting in experimental study 2 (which *measured* the reference point) both of which reveal a pattern reflective of reference dependence; (b) showing that the results of observed choices in experiment 1 also reveal a pattern which indicates that the middle option serves as the implicit reference point (compromise effect); and (c) demonstrating holdout sample results from models based on the desirability measurement approach that are found to improve predictions over a model that uses the conventional measurement approach in experiment 2 and over chance predictions in experiment 1. As a result of observing loss aversion and (a), (b), and (c), we believe that the desirability measurement approach is getting at the reference point phenomenon. We recommend future research to collect process measures for further validation of the desirability measurement approach.

Another potential limitation is that the measurement task preceded the holdout preference task in all three studies. Measurement can impact preference. In addition, it may have been useful to have a “filler task” between the measurement and holdout preference tasks. However our ordering of the measurement and holdout preference tasks and the absence of a filler task is consistent with most if not almost all published studies in the conjoint measurement literature. In fact, holdout preference tasks can impact measurement as well; nonetheless, future research in experimental settings could employ both orderings and filler tasks. We employed holdout choices for predictive validation, which were separate from the measurement data used to calibrate the conjoint models in all our studies, in a way that is consistent with the validation reported in almost all the published literature on conjoint measurement. Remarkably however, Srinivasan and Park (1997) went beyond the preference validation task

reported in almost all such studies and predicted actual job choices for MBAs, which we do not do, so that real world validation is left to future research.

The weighting parameter λ is estimated by maximizing the predictive ability of the choices. Whether lambda is estimated across subjects or for each subject depends on the number of choices available. In the paper, lambda is estimated across subjects because few choices are collected. However, it is indeed possible to estimate lambda for each subject by collecting data on a larger number of choices, though that would increase the assessment burden on subjects. We also recommend that future research assess the potential tradeoff between the assessment burden and improvement in predictive ability.

Some additional questions for future research are as follows. Is it easier to change the reference point for some consumers over others? What are the characteristics of such consumers, for example, are they loyal or switchers, heavy or light users, frequent or infrequent purchasers, etc? What are the effects of different appeals that can be employed to change the reference point, for example, appeals based on price or value vs. quality or brand name? Another question for future research is that if reference effects are context specific and depend for example on a past purchase or available offerings, will predictions to new environments for example a low or high assortment environment be affected? We hope that future research will build on our effort in such directions.

References

Andrews, R., Ansari, A., & Currim, I. (2002), "Hierarchical Bayes versus finite mixture conjoint analysis models: A comparison of fit, prediction, and partworth recovery," *Journal of Marketing Research*, 39, 87-98.

Bell, D. & Lattin, J. (2000), "Looking for loss aversion in scanner panel data: The confounding effect of price response heterogeneity," *Marketing Science*, 19 (2), 185-200.

Bridges, E., Yim, C. K., & Briesch, R. (1995), "A high-tech market share model with customer expectations," *Marketing Science*, 14 (1), 61-81.

Briesch, R., Krishnamurthi, L., Mazumdar, T., & Raj, S. (1997), "A comparative analysis of Reference Price Models," *Journal of Consumer Research*, 24 (2), 202-214.

Currim, I. S., & Sarin, R. K. (1989), "Prospect versus Utility," *Management Science*, 35 (1), 22-41.

Dholakia, U., & Simonson, I. (2005), "The effect of explicit reference points on consumer choice and online bidding behavior," *Marketing Science*, 24 (2), 206-217.

Fishbein, M., & Ajzen, I. (1975), Belief, Attitude, Intention and Behaviour: An Introduction to Theory and Research, Addison-Wesley, Reading, MA.

Green, P., & Srinivasan, V. (1977), "Conjoint analysis in consumer research: issues and outlook," *Journal of Consumer Research*, 5 (2) 103-123.

Hardie, B., Johnson, E., & Fader, P. (1993). "Modeling loss aversion and reference dependence effects on brand choice," *Marketing Science*, 12 (4), 378-.

Kahneman, D., & Tversky, A. (1979), "Prospect theory: an analysis of decision under risk," *Econometrica*, 47 (2), 263-91.

Kalyanaram, G., & Winer, R. (1995), "Empirical generalizations from reference price and asymmetric price response research," *Marketing Science*, 14 (3, Pt. 2), G161-69.

Mazumdar, T. & Papatla, P. (2000) "An investigation of reference price segments," *Journal of Marketing Research*, 37 (2), 246-258.

Meyer, R., & Johnson, E. (1995), "Empirical generalizations in the modeling of consumer choice," *Marketing Science*, 14 (3), 180-89.

Simonson, I. (1989), "Choice based on reasons: The case of attraction and compromise effects," *Journal of Consumer Research*, 16 (September), 158-174.

Srinivasan, V. (Seenu), & Park, C. (1997), "Surprising robustness of the self-explicated approach to customer preference structure measurement," *Journal of Marketing Research*, 34 (May), 286-91.

Tversky, A., & Kahneman, D. (1991), "Loss aversion in riskless choice: a reference-dependent model," *The Quarterly Journal of Economics*, 106 (4), 1039-61.

Tversky, A., Sattath, S., & Slovic, P. (1988), "Contingent weighting in judgment and choice," *Psychological Review*, 45 (July), 371-384.

von Winterfeldt, D., & Edwards, W. (1986), Decision analysis and behavioral research, Cambridge University Press, Cambridge, England.

Wittink, D., & Montgomery, D. (1979), "Predictive validity of trade-off analysis for alternative segmentation schemes," 1979 Educators' Conference Proceeding, eds. Neil Beckwith, Michael Houston, Robert Mittelstaedt, Kent B. Monroe, and Scott Ward, Chicago: American Marketing Association, 68-73.

Figure 1

Rite Aid Example

Figure 2

Dell Example

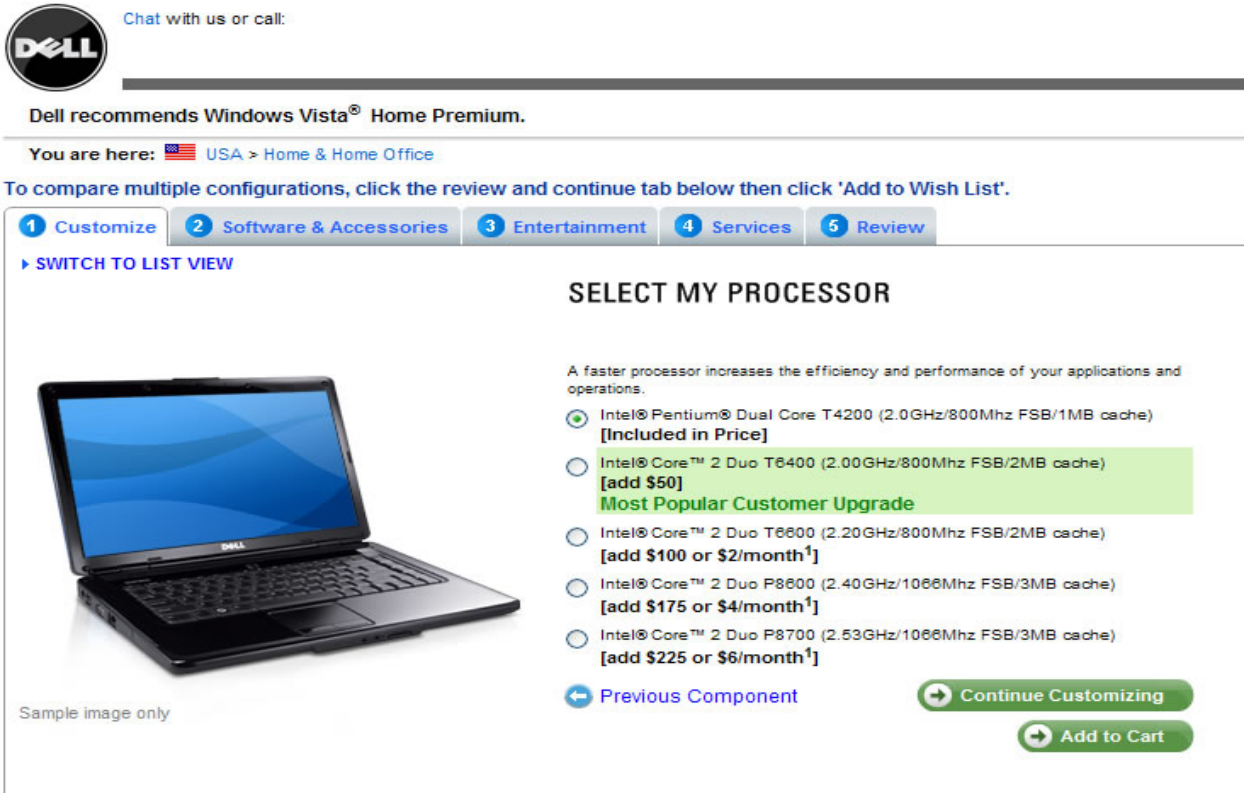


Figure 3

Illustration of the Choice Set in Experiment 1

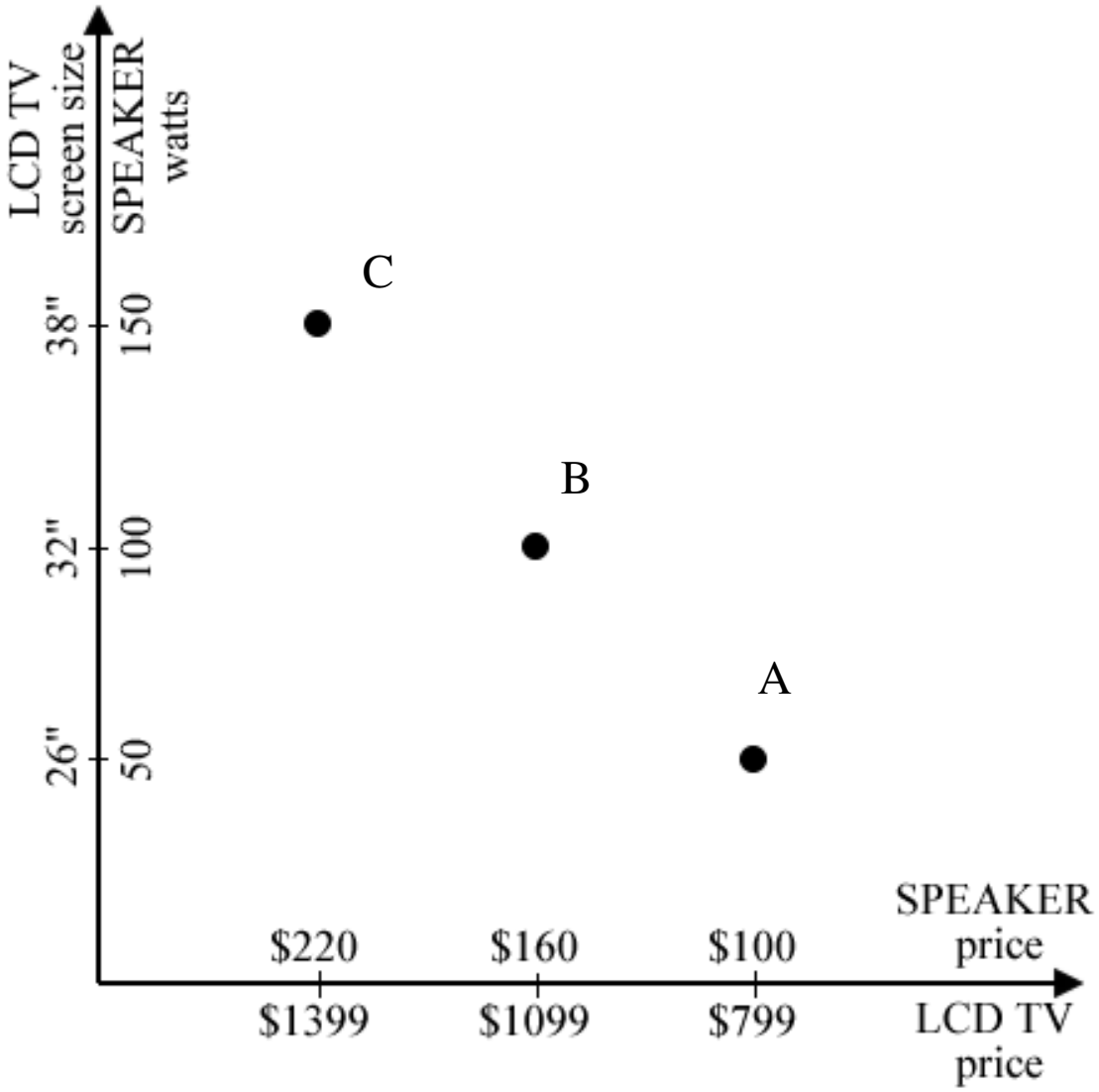


Figure 4

Attributes and Levels for Experiment 2

Salary

20% above expected salary
10% above expected salary
expected salary
10% below expected salary
20% below expected salary

Advancement Opportunity

Rapid
Moderate

Commute Time

100% above expected
50% above expected
Expected commute time
50% below expected
No commute time

Company Growth

Gross National Product plus 10%
Gross National Product
Gross National Product minus 10%

Region

Los Angeles
Orange County
San Francisco
San Diego

People

Flexible
Competitive

Travel

1 night per month or less
2-5 nights per month
6 nights per month or more

Functional Activity

Accounting/Auditing
Consulting
Marketing
Finance
Operations

Table 1

Experiment 1: Percentage of holdout choices correctly predicted

Holdout Choices	Latent Reference Dependent Model [4]	Chance Model
Triplets	52.82%	33%
Pair AB	68.55%	50%
Pair AC	66.94%	50%
Pair BC	67.34%	50%

Table 2
Experiment 2: Percentage of holdout choices correctly predicted

Holdout Choices	Latent Reference Dependent Model	Traditional Self-Explicated Conjoint Model
Q1 (ABC*)	47.89	36.21
Q2 (BC*D)	50.70	34.48
Q3 (C*DE)	50.70	25.86
Overall	**49.77**	**32.18**