Concept Test Accuracy And Consumer Uncertainty: Propositions And Empirical Verification

Craig F. Kolb, Ask Afrika, South Africa

Abstract

While concept testing is commonly used in marketing research practice, the issue of concept ambiguity and respondent uncertainty, purchase/switch probability distribution form (a factor related to accuracy), and the underlying rationale for distribution form have not been extensively researched. There is also a dearth of research into the effects of consumer uncertainty on purchase/switch probability means – which are crucial to prediction. The study extends the literature by verifying the impact of uncertainty on distribution shape and mean; while contributing four propositions to explain how uncertainty affects concept test accuracy, for better or worse. Two of the propositions are tested and supported by the results. A simple uncertainty ratio measure is also proposed and evaluated.

Key words: Concept testing, predictive accuracy, probability scale, Juster scale, uncertainty, mobile number portability.

Introduction and Background

Concept testing is frequently conducted in the early stages of new product and service development to isolate the most successful concepts, and to provide diagnostic information to aid in the further development of new offerings. A crucial component of concept tests is the purchase interest/intention scale (Schwartz, 1987), or a probability scale such as the Juster scale (McDonald and Alpert, 2001). These scales are used to isolate relatively more successful concepts from a cluster of concepts, or are compared to category norms to assess consumer acceptance. While at times attempts are made to predict actual demand, a lack of accuracy is a key concern (Schwartz, 1987).

Previous research has highlighted the relationship between predictive accuracy and the probability scale distribution. Kuhlwani and Silk (1982) pointed out the relationship between distributional form as measured by the parameters of a beta distribution and test-retest reliability. More recently, Riebe (2000) took this a step further, showing that distributional form related to predictive accuracy.

McDonald and Alpert (2001) showed that the Juster scale did not perform well in predicting actual adoptions for a new product concept test, with the caveat that this may be have been due to intervening factors that respondents did not consider. Infosimo (1986) also found similar problems and noted that promotion and distribution effects needed to be incorporated in the model. Jamieson and Bass (1989) confirmed the efficacy of such a model, which modified intention scale results utilising factors such as awareness. These studies implicitly highlight the fact that probability/intention scales, are effective in a concept test setting - once contextual details which are normally known to the consumer in an existing product situation are artificially taken into account or modeled on behalf of the consumer. Alternatively, additional contextual detail incorporated in the statement itself can improve realism. Unfortunately, it is impossible to provide perfect realism. As a result it would be useful to determine the impact of any concept ambiguity, by measuring respondent uncertainty.

Ekstrand and Loomis (1998) highlight two possible ways of measuring uncertainty, firstly by using scales, or by asking respondents to indicate their level of certainty after responding to a dichotomous intention question. The second approach it is argued creates an unnecessary increase in questionnaire length. In this paper scale distributional form is used as a measure of aggregate consumer uncertainty. Uncertainty is defined here as a state where the information available deviates from the consumer's ideal information state. Since we rarely have perfect information or the ability to process all the information available, we resort to making assumptions and so take on risk (Morgenstern and Zechmeister, 2001). Different respondents, facing the same decision, may display differing propensities to take risks when faced with similar uncertainty in the same decision making situation, in order to move forward to action. According to Morgenstern and Zechmeister (2001), an individual's predisposition to accept risk is a key factor in differentiating individuals in terms of willingness to act under uncertainty.

This, it could be argued, determines a respondent's choice of items on a scale. When choosing from the extremes of the scale the respondent indicates a willingness to accept the uncertainty and risks they perceive and their willingness to act. When choosing the middle of the scale the respondent indicates a reluctance to act; having crossed an uncertainty threshold. The number of respondents choosing points toward the middle of the scale increases as the number of respondents who are uncertain increases.

To elaborate on the threshold concept, Bartels (1984) proposed a measure of voter uncertainty on candidate positions on an issue, based on individual variance per voter across items of a scale measuring a candidate's position. He was however unable to obtain the necessary data and instead utilised respondent refusal "to place the candidate" on a political issue, in order to operationalise the measure, reasoning that refusal occurred once the "unobserved" variance (uncertainty) exceeded an uncertainty threshold "T". While the present paper takes a similar view to Bartels (1984), it differs in that respondents are reliant mostly on the concept for information, and so do not suffer the variations in knowledge voters would of political candidates. Instead it is argued that each individual has their own predisposition toward accepting different risks (Morgenstern and Zechmeister, 2001).

Another issue which will be explored is the effect of concept wording, in particular the affect of uncertainty on the mean probability, which is used as a forecast of the proportion purchasing a service/product, or switching brands, in a specific time period. Other studies have examined the impact of wording, pictures, and product mockups on respondent interest; but none thus far has examined the effect of uncertainty. Some such as that by Armstrong and Overton (1971), found no difference between a brief written description and a comprehensive description, which included elaborate props and opportunities to ask questions. More recently, work by Wright et. al. (2004) showed that using a purely factual concept statement, versus more persuasive statements or statements enhanced by visuals resulted in little difference in the results. Similarly, Dickinson and Wilby (1998) found that a concept test with a product placement, versus a concept test without (for an existing product) produced no difference in scores, highlighting the fact that unless an alternative concept test offers additional information it is not likely to have a different impact. In contrast to these authors others have found different mean ratings in response to different wording. Hally and Gatty (1971) found significant differences across different concept writers executions of the same concept. Lewis (1984) found differences upon in home placement, although it appears this may have been a methodological artifact. The present study extends the literature by examining the impact of concept ambiguity on mean purchase probability scores.

Four propositions are put forward to contribute a theoretical framework for explaining the impact of uncertainty on predictive accuracy. Firstly it is argued that the threshold of uncertainty "T" varies across individuals in the same situation. This could be represented by the midpoint of the probability scale, since; at an individual level the midpoint of the scale represents the *maximum level of uncertainty of action.* Secondly, at an aggregate level, the proportion that cross the threshold and indicate a *maximum level of uncertainty* will increase as the perceived costs of a decision increase relative to the benefits in different situations, while the proportion indicating a *minimum level of uncertainty of action* will decrease relatively. Thirdly, whether the threshold is crossed in a particular situation depends on a predisposition to accept risk at an individual level. The fourth and last proposition aims to explain the effect of uncertainty on mean switch/purchase probability, a statistic commonly used as a forecast of the proportion purchasing or switching. The mean switch/purchase probability of a new concept can increases or decrease relative to a before measurement of existing switch/purchase rates - in part due to an increase in aggregate uncertainty (represented by the respondents crossing the threshold and choosing the mid-point) relative to the *before-measurement* level of uncertainty. Assuming mean probability for switch/purchase in the existing product/service situation was below 0.5, increased uncertainty will increase the mean (and vice versa if above 0.5).

The first proposition is regarded as axiomatic, the second is evaluated by examining differential response to the two versions of a concept statement describing the mobile number portability concept (MNP). The third proposition is not evaluated in the present study. The fourth and last proposition is evaluated by testing for significantly different means in response to the plain and ambiguous concepts. Limitations and future research directions are then discussed.

Methodology

Instrument & design

Experimental and control groups were utilised. In the experimental group, the before-treatment measure, a 11 point likert scale, was used to establish the likelihood of switch prior to exposure to the mobile number portability (MNP) concept. Respondents were then exposed to the treatment, labeled the *ambiguous* concept test statement. This included a mention of charges being levied if a switch occurred before the expiration of the mobile contract, but failed to mention the exact charge. A second aspect, mentioned a “better deal” being offered by a competitor, but without specifying what it was. As an after-treatment measure, respondents had to indicate the likelihood of switch. Respondents in the control group were measured using exactly the same before-treatment likelihood of switch measure as those in the experimental group. They were then exposed to a *plain* concept statement. This plain statement only mentioned that they could keep their mobile phone number – no mention of possible charges or a “better deal” being made. Respondents then had to indicate the likelihood of switch as a post-treatment measure.

Sample

Two telephonic surveys were conducted in South Africa using a CATI system. Two samples were drawn using random digit dialing of the postpaid subscribers of a mobile phone operator, labeled for the purposes of this paper as Operator A. The first sample of 217 respondents was drawn in June 2004, and exposed to the *ambiguous* concept statement. The second control sample of 188 postpaid subscribers, was drawn in March 2005 and exposed to a *plain* concept statement.

Results

Using beta distribution fit as an indicator of accuracy in concept testing

The work of Kuhlwani and Silk (1982), highlight the relationship between the parameters of the beta distribution, P & Q and test retest reliability. The more $P + Q \rightarrow 0$, the more reliable. Decreasing P + Q tends to result in the mass of the distribution moving outwards. Riebe (2000) described this in more graphical terms as a U or J shaped distribution, which was shown to relate to greater predictive accuracy. This is in line with propositions one and two, stated in the introduction, as increasing certainty should produce shifts away from the middle of the distribution.

Beta distributions were fit manually using the method of moments (shown in brackets in table 1), while NCSS was used to fit using maximum likelihood. The larger negative log likelihoods for the plain concepts indicate better fit. As can be seen P + Q is closer to zero in the case of the plain concept than in the case of the ambiguous concept. A comparison of confidence intervals for the maximum likelihood based parameter estimates; shows that while the P parameters show significant differences (0.05 significance level) between the ambiguous and plain concept distributions (for both operators), a difference between the Q parameters could not be confirmed. A simpler alternative was then explored.

Table 1: A Comparison of the Beta Distribution Parameters P & Q

	Log – likelihood	P	Q	P+Q	Mean	Std dev.	Lower Limit (0.05 level)	Upper Limit (0.05 level)
Operator A - Ambiguous (n=188)	-660.81	0.1477 (0.4994)	0.1821 (0.3936)	0.3298 (0.893)	5.5926	3.6089	P=0.1241	P=0.1713
							Q=0.1509	Q=0.2133
Operator A - Plain (n=216)	-1011.2	0.0929 (0.2203)	0.2011 (0.4068)	0.2940 (0.6271)	3.5137	3.7429	P=0.0778	P=0.108
							Q=0.1588	Q=0.2434

A simple alternative to the beta parameters

In line with propositions one and two in the introduction, a parsimonious measure of uncertainty is proposed and labeled the uncertainty ratio. This is simply calculated as the ratio of the frequency of the mid point of the scale (a 5 in this case) to the sum of the frequencies of the extremes of the scale (a '0' and '10' in this case). Ratios of 0.1552 and 0.2913 were obtained for the simple concept and ambiguous concepts respectively. Fisher's exact test for a difference between proportions was performed, and a significant difference in proportions ($p<0.01$, two tailed) was found, supporting the proposition that the aggregate uncertainty ratio be used to measure uncertainty.

While this approach does not take into account the entire distribution and therefore loses some information, it has advantages over the use of beta parameters. Firstly it is more easily calculated, secondly this approach is not subject to the problem of imperfect Beta distribution fit, which affects the accuracy of the beta parameters. While Riebe, Danenberg, Sharp and Rungie (1999) suggest that lack of fit may itself act as an indicator of inaccuracy, which makes sense given the pronounced un-beta like w shape which occurs as the uncertainty ratio increases, the corollary, that "fit equals accuracy" would not apply (as the beta can take on shapes close to a normal distribution, where mass moves to the centre). Lastly, it is possible to set P + Q equal and yet have distribution shapes suggesting different predictive accuracy levels.

Evaluating the impact of uncertainty on the mean probability of switch

While the before-treatment measures could have served as covariates, ANCOVA was abandoned due to the failure of the linearity assumption. One way ANOVA's were conducted instead. No significant differences were found for Operator A in the before-treatment measures between the two groups, verifying that no uncontrolled factors impacted on the results. The ambiguous concept mean probability was significantly (0.05 significance level) higher than the plain version. The null hypothesis of equal variance was not rejected in either case ($F=31.773$, $df=1$, $p<0.000$). This result supports proposition four, as the increase in the proportion of respondents crossing the uncertainty threshold 'T' contributed to the shift in the mean. In addition, a change in the ratio of those selecting a '10' to those selecting a '0' increased from 0.38, to 1.14, contributing to the shift. It seems plausible that a greater number assumed the positive aspects of the ambiguous concept outweighed the negative aspects. The positive aspect – an unspecified "better deal" - was assumed to be sufficiently attractive by those who chose a 10. The negative aspect – an unspecified charge for switching before the contract ended - was seen as less of a barrier as respondents could wait for the contract to end. This was confirmed by a question which asked respondents if they would wait. As indicated by the uncertainty ratio, the respondents who indicated the *minimum level of uncertainty* decreased relative to those who crossed the threshold and chose the midpoint, to indicate a *maximum level of uncertainty*.

Discussion and Practitioner Implications

This study demonstrated that concept statement ambiguity is a factor driving changes in distributional form, which has an established relationship with predictive accuracy. This confirms the importance of clear concept wording, while also showing that practitioners can measure the level of uncertainty in a concept test, allowing for continuous improvements in wording with judicious use of norms. A simple aggregate uncertainty ratio was shown to be able to detect distribution changes, and would be useful as a measure of uncertainty in practice. Four propositions were put forward to explain consumer response to ambiguity in concept test wording, in particular how this either decreases or increases the mean purchase probability, which has obvious implications for accuracy. Practitioners not aware of this may wrongly conclude that a new concept will be more or less successful than an existing product/service, by incorrectly attributing changes in switch/purchase rates to the new concept when in fact it may in part be driven by an increase in uncertainty. Two of these propositions were tested in the present study and supported by the results. In part this offers an explanation for the decrease in Juster accuracy in the new concept test situation, a scale otherwise successfully applied in existing product service markets.

Besides replications, the present study could be extended by quantifying the relationship between measures of uncertainty and predictive accuracy, by testing proposition three and evaluating the relationships between risk aversion, uncertainty and predictive accuracy.

References

Armstrong, J.S., and Overton, T., 1971. Brief vs. Comprehensive descriptions in measuring intentions to purchase. Journal of Marketing Research. Vol. VIII (February). 114-117.

Bartels, L.M., 1984. Issue voting under uncertainty. American Journal of Political Science. 30 (40), 709.

Dickinson, J.R., and Wilby, C.P., 1998. Concept testing with and without product trial. Journal of product and innovation management. 14 (2), 117-125.

Hally, R.I., and Gatty, R., 1971. The trouble with concept testing. Journal of marketing research. Vol. 8 (May), 231-232.

Infosimo, W.J., 1986. Forecasting new product sales from likelihood of purchase ratings. Marketing Science. 5 (4) Special issue on consumer choice models, Autumn, 372-384.

Jamieson and Bass., 1989. Adjusting stated intention measures to predict trial purchase of new products: A comparison of models methods. Journal of marketing research. Vol. 26 (August). 336-345.

Kuhlwani, M.U., and Silk, A.J., 1982. On the reliability and predictive validity of purchase intention measures. The Institute of Management Sciences.

Lewis, I.M., 1984. Do concept scores measure the message or the method. Journal of Advertising Research. 24 (1).

Ekstrand, E.R., and Loomis, J., 1998. Incorporating Respondent Uncertainty When Estimating Willingness to Pay for Protecting Critical Habitat for Threatened and Endangered Fish. Water Resources Research, 34, 3149-55.

McDonald, H., and Alpert, F., 2001. Using the Juster Scale to Predict Adoption of an Innovative Product. Proceedings of the Australian and New Zealand Marketing Academy Conference 2001, 1-8, Massey University, Auckland, NZ.

Morgenstern, S. and Zechmeister, E., 2001. Better the devil you know than the saint you don't: Risk propensity and vote choice in Mexico. The Journal of Politics. 63 (1), 193-119.

Riebe, E., 2000. Identifying Variations in the Accuracy of Probabilistic Predictions. Proceedings of the Australian and New Zealand Marketing Academy Conference 2000.

Riebe, E., Danenberg, N., Rungie, C., and Sharp, B., 1999. Verifying the distribution of probability scales. Australian and New Zealand Marketing Academy Conference. Sydney, Australia, University of New South Wales.

Schwartz, D., 1987. Concept testing: how to test new product ideas before you go to market. New York, NY: American Management Association.