REFERENCE AND MEMORY IN CUSTOMER SATISFACTION STUDIES

James H. Drew, GTE Laboratories Inc.

ABSTRACT

Revealed customer opinion is not necessarily static over time for a variety of reasons, many of which are functions of consumer memory and not directly related to objective levels of service provision. In the absence of empirical data, corporate folklore has developed several divergent views of the effect of customer memory and past experience on current satisfaction levels. Through a general statistical model of service quality evaluation and its change dynamics, objective mediation among these competing hypotheses is possible. This paper presents two longitudinal studies of divergent customer bases whose data modeling sheds light on several aspects of customer memory as it relates to service quality evaluation and modification.

INTRODUCTION

Customer satisfaction researchers have developed extensive theories for the nature and formation of the satisfaction affect at a single point in time. Most companies, though, record and track quality and satisfaction statistics over time, often through survey designs which gather periodic data from the same customers. It is then possible to examine and model how satisfaction and quality judgments change over time and in relation to which stimuli. In particular, since both sorts of affect depend on reference standards, it is important to also model the nature, time lag and dynamics of such references.

Opinion Updating Models

We will examine some of the models of opinion dynamics which have been developed in the marketing, cognitive psychology and survey research literature. Insofar as many of these models were constructed using laboratory or non-experiential data, the results reported here afford tests of some well-known models and opportunities for their refinement.

Models from Marketing Science. There are several models in the marketing literature which explain changes in customer satisfaction through successive updates in expectation levels. Drawing on the adaptation level theory of Helson (1964), LaBarbera and Mazursky (1983) tested a model in which prior satisfaction standards are adapted to current service experience, thus creating an updated standard. Boulding et al, (1993) have developed a fuller model of the interplay between customers' expectations, perceptions and evaluations, and how these constructs are updated over successive service experiences.

In their model, "would" (i.e. predictive) expectations, and "should" (i.e. normative) expectations are updated over successive service experiences, and evaluated along with service attributes to produce a customer satisfaction level via the comparison of expectations and performance. Delivered service at time t, for example, is positively related to the "would"-expectation at time $t+1$, which in turn is positively related to delivered service at time $t+1$. This model thus explains correlations in customer perceptions between times t and $t+1$ through delivered service at time t.

Note that although both the Boulding model and the adaptive expectations model will behave similarly in the sense that a prior positive experience will have a positive effect on the current rating, and a negative one will depress it, the former model links consecutive aggregate ratings only through delivered service attributes (which drive the updating of expectations), while the latter allows for a direct link.

Models from Cognitive Psychology. We have outlined two adaptation models already in use in the marketing tradition. An even more explicit model for belief updating was developed by Hogarth and Einhorn (1992), which draws on research in cognitive psychology. This model was constructed to test the classic primacy model , where the current rating is completely dependent on any initial rating (e.g. Nisbett and Ross, 1980), and the classic recency model, where the current rating completely supersedes any prior ratings (e.g. Davis, 1984).

They postulate an anchor-and-adjustment model, wherein attitudes are updated with each discrete new experience X_t, and current attitude A_t is a weighted average of the prior attitude and the difference between the current experience and one of several reference points:

$$A_t = A_{t-1} + w_t (X_t - R).$$

In this formulation, R can be a constant, rather like a "should" expectation, in which case the ordering of the experiences is irrelevant and there are no primacy or recency effects; or R can depend on the prior attitude A_{t-1} itself, in which case the updating is like a moving average and the order of experiences is quite important. Further, the weight w_t of the second, "disconfirmation" term in parentheses is related to the size and parity of the disconfirmation EX_t - R: if the experience is more favorable than the reference, then the weight is inversely proportional to the anchoring attitude A_{t-1}, and if the experience is less favorable then it is directly proportional to A_{t-1}. This is a contrast effect, and specifies the notion that adjustments are larger in the direction toward $EX_t - R = 0$.

Models from Survey Research. Two contrasting forms for attitude responses have been mentioned in the survey literature. Waterton and Lievesley (1989) mention two such models. One is that repeated interviewing "freezes" attitudes, or that repeated interviewing speeds up the process of forming a stable attitude, so that attitudes are consistent over observations. The second postulates raised consciousness through repeated interviewing, so that adjustments in attitude, particularly from the first to the second interview period are common. In their analysis of the British Social Attitudes Survey, Waterton and Lievesley find no strong evidence of either effect.

Other models from the survey literature focus directly on the role of the instrument. Simmons et al. (1993) discuss the role of the questionnaire in creating attitudes not existing prior to the subject's response. They also postulate conditions under which responses to one item may affect responses to a subsequent item. Schul and Schiff (1993) also address the issue of inter-item effects, and develop an argument for the longevity of the memory of inadequate service encounters. This is a restating of the folklore that customers remember service experiences for long time periods, and that a service problem is remembered, and has an effect on overall opinion for a longer time than service without incident.

These models, and some of their salient or distinguishing features, are summarized in the following table:

Model	Source	Important Features
Adaptive Expectations	Helson (1964)	Direct linkage of aggregate rating
"Would"/"Should" Expectation Updating	Boulding et al. (1993)	Rating linkage through expectations, service attributes, <u>not</u> aggregate ratings
Primacy	Nisbett and Ross (1980)	First rating determines subsequent ratings
Recency	Davis (1984)	Prior ratings completely overcome by current rating
Anchor/Adjustment	Hogarth-Einhorn (1992)	Adjustment amount from previous reference point depends on reference level
Memory Freezing	Waterton and Lievesley (1989)	No attitude change from early ratings
Respondent Sophistication	Waterton and Lievesley (1989)	Largest change in attitude between first and second survey wave

Data From Study I

The data that we analyze in this study were edited from a large set of business customer opinion survey interviews conducted from 1989-1992. These surveys make annual contact with identified telecommunications decision makers in medium and large businesses in telephone company franchise areas. Substantial editing was necessary to delete businesses whose interviews were not matched during this period. Further, provision had to be made for separate interviews for a given business which happened to be conducted in the same calendar year. (This can occur through errors in the survey frame or in interviewing schedules.) We assume such deletions are not related to survey responses or their interactions. Previous studies of cross-sectional data have found, in fact, that for these types of customer opinion surveys, there is no measurable effects due to nonresponse; see Drew (1990). The combined dataset contained approximately 3000 customers who furnished two or more interviews during this time period.

Anchor-Adjustment Models. A popular model for attitude formation, mentioned by Kahneman and Tversky (1973), makes the current attitude an adjustment from some form of reference value. Updates in attitude are due to changes in current perceived performance and an updating of the reference value from which current performance generates an adjustment. The updating of the reference level has the form:

$$\begin{aligned} REF_{t+1} &= REF_t + (1-\rho)(x_t - REF_t) \\ &= \rho\_REF_t + (1-\rho)\, x_t \end{aligned}$$

where REF_t is a reference value perhaps depending on previous experience, x_t is a measure of current (i.e. time t) perception, and ρ is a weight measuring the relative contribution of x_t and REF_t. Note that $\rho\_ = 1.0$ yields the recency model, while $\rho = 0.0$, with an initial reference value set equal to the first rating yields the primacy model, with the added condition that no other current attributes enter the model as explanatory variables. Hence, these two models can be evaluated in the context of the anchor-adjustment model.

As mentioned earlier, in the Hogarth-Einhorn model, the weight $\rho_$is taken to be directly proportional to the reference value when the current rating exceeds the reference value, and inversely proportional to the reference value when the current rating is smaller.

An extended model can also incorporate the features of the memory freezing model. By postulating that the reference value will not be updated unless the current perception is much more favorable or much less favorable, one can simultaneously allow freezing, as well as the Hogarth-Einhorn form of reference adjustment. This is accomplished by allowing these two different models to operate in mutually exclusive rating response intervals. As noted above, the Boulding model can be tested by allowing the inclusion of attributes from prior time periods.

In our data, the rating at hand is QREP, the customer's evaluation of overall repair quality, and its significant attributes were found (from earlier studies) to be RFX and RDEAD (Fixing the Problem the First Time, and Meeting Deadlines, respectively.)

Consider the specification below, taken from Drew and Bolton (1996). The basic form of this model is anchor-adjustment as postulated by Hogarth-Einhorn, with the anchor for time t represented by REF_t and the adjustment a function of repair attributes for the current (time t) period. The anchor REF_t consists of a weighted average of the preceding general rating $QREP_{t-1}$ and the preceding anchor, and the attribute ratings of the preceding period, as the Boulding model and a test of memory sophistication require. The weighting function of the preceding anchor and general rating take on a slightly more general form than suggested by Hogarth-Einhorn: first, non-zero weighting takes place only when the preceding rating differs from the preceding anchor by more than k units; and second, the weight is based on a power of the preceding reference. Hogarth-Einhorn specify that k=0 and $\alpha_+ = -\alpha_- = 1.0$, and a natural alternative is $\alpha_+ = \alpha_- = 0.0$.

$$\text{QREP}_t = \begin{cases} \text{REF}_t + \beta_1 \text{RFX}_t + \beta_2 \text{RDEAD}_{t1} + \epsilon_{it} & \text{if QREP}_{t-1}\text{-REF}_{t-1} \; 5 \; \text{-k} \\ \text{REF}_{t-1} + \beta_1 \text{RFX}_t + \beta_2 \text{RDEAD}_{t-1} + \epsilon_{it} & \text{if } |\text{QREP}_{t-1}\text{-REF}_{t-1}| < \text{k} \\ \text{REF}_t + \beta_1 \text{RFX}_t + \beta_2 \text{RDEAD}_{t-1} + \epsilon_{it} & \text{if QREP}_{t-1}\text{-REF}_{t-1} \; ; \; \text{k} \end{cases}$$

where

$\text{REF}_t = \rho_- \text{REF}_{t-1} + (1-\rho_-) \text{QREP}_{t-1,}$ if $\text{QREP}_{t-1}\text{-REF}_{t-1}$ 5 -k,

$\text{REF}_t = \rho_+ \text{REF}_{t-1} + (1-\rho_+) \text{QREP}_{t-1,}$ if $\text{QREP}_{t-1}\text{-REF}_{t-1}$; k, and

$\rho_- = \beta_- \text{REF}_{t-1}^{\alpha-}$ and $\rho_+ = \beta_+ \text{REF}_{t-1}^{\alpha+}$

Since the Hogarth-Einhorn model includes recency and primacy as special cases, and the inclusion of QREP_{t-1} in the anchor subsumes adaptive expectations, our model thus includes the characteristics of each of the rating change models given in the table above, except perhaps for the freezing and sophistication of memory, which do not seem viable in this context. Note too that by allowing the form of the current reference to be a function of the prior survey variable, we capture the notion that responses are potentially artifacts of the item form., that is, that QREP_t should look much like QREP_{t-1}.

Since the join points of this segmented regression are unknown, this model must be fit by a nonlinear routine. The complexity of this model, along with the response patterns of the 162 customers sampled, did not allow the unrestricted fitting of α_- and α_+, so that only the two alternatives $\alpha_+ = -\alpha_- = 1.0$ and $a_+ = \alpha_- = 0.0$ were fit and compared. The resulting loss functions were nearly identical, and there is no evidence to reject the Hogarth-Einhorn form, but for our purposes it is more interpretable to allow $\alpha_+ = \alpha_- = 0.0$. The period t=3 is tested, and to test the constancy of the anchor the initial anchor REF_2 is given the linear form

$$\mu_0 + \beta_0 \text{QREP}_{1.}$$

The model was simplified in other ways. Fitting a non-zero value for k invariably resulted in estimates for $\mu_0 + \beta_0 \text{QREP}_1$ and k which led to sparse or empty model segments. Interpreted as an indication that these data do not support three segments, k was fixed at zero. As noted below, the current attributes were also ignored in the results reported here. Further, a preliminary model gave nearly identical estimates for β_- $(=\rho_-)$ and β_+ $(=\rho_+)$ and the common value is denoted by ρ.

A simplex routine (the O'Neill, 1971 algorithm as implemented in SYSTAT) yields the following least squares estimates when t = 3, and $a_+ = \alpha_- = 0$, which were confirmed using several different starting values for the algorithm, and where the estimated asymptotic standard errors are given in parentheses:

Paramete r	Estimate (a.s.e.)
μ_0	0.880 (0.236)
β_0	0.150 (0.060)
ρ	0.968 (0.061)
β_1	0.271 (0.061)
β_2	0.323 (0.065)

All coefficients are significantly different from zero at the 0.001 level. It is possible, then, to comment on aspects of customer memory that have been previously discussed. Our estimates imply that:

! $QREP_t$. does not depend on its value from the preceding time period nearly as much as it depends on current period attributes.

! There is no evidence that k is significantly larger than zero, so the updating mechanism is uniform across waves.

! The single anchor update weight is close to 1.0. Therefore, updating of the anchors heavily favors the initial anchor. Further, the initial anchor is only slightly dependent on historic perceived service, and thus not far from being a constant.

! The overall repair rating QREP is unlikely to be a survey artifact, in the sense that the previous period QREP variable and the original anchoring value $QREP_1$ have little influence on the current period rating.

Finally, it is important to note some indications from unsuccessful estimation runs. Models with both the updating mechanism and the current period attributes RFX_3 and $RDEAD_3$ generally either did not converge to estimates with reasonably small asymptotic standard errors, converged to estimates with k very large or very small, converged to estimates with the attribute coefficients indistinguishable from zero, or did not converge at all. Presumably, this shows that updated past evaluations carry about the same information as current attribute ratings. This, of course, has important implications for the use of the panel designs that are necessary to acquire time series data from customers.

Data from Study II

The preceding results should perhaps be viewed cautiously, on the grounds that survey responses were collected no more frequently than at annual intervals, that POTS (Plain Old Telephone Service) evaluation is not cognitively stimulating, or that the memory lag and updating structure is specific to business customers only.

The following study is based on three waves of telephone surveys with cellular telephone customers surveyed during the time periods January-March 1992, August-September 1992 and April-May 1993. Unlike the business customers of our first study, these customers were mostly personal and small-business users, and few made telecommunications decisions for anyone other than themselves.

A small number of customers were dropped from the analysis because they chose to terminate service. Most such termination was not related to issues of cellular quality, and a logistic regression revealed that termination was not statistically related to prior quality ratings. The number of customers interviewed for all three waves of the study was 245.

The questionnaires from this survey program each contained items associated with service quality and its attributes; the following items ask for the overall satisfaction (on a 1-5 scale) with:

OVQ_i: the overall services you received
$MAKE_i$: the calls you make and receive
$FIRST_i$: calls going through the first time
$PRICE_i$: the prices charged,
$CLEAR_i$: the clarity of the calls, and
$CALLCS_i$ indicates whether a call was made to Customer Service.

The subscript refers to the survey wave (i=1,2,3) in which the data were furnished.

Consider first this simplification of the model introduced above:

$$OVQ_3 = \begin{cases} (1-\rho_-)\,OVQ_2 + \rho_-\,(\mu_0 + \beta_0\,OVQ_1)\ , & \text{if } OVQ_2 < \mu_0+\beta_0\,OVQ_1-k \\ (1-\rho_+)\,OVQ_2 + \rho_-+(\mu_0 + \beta_0\,OVQ_1)\ , & \text{if } OVQ_2 > \mu_0+\beta_0\,OVQ_1+k \end{cases}$$

In a preliminary analysis, it was found that k was effectively zero, and there was no significant difference between the reference weights ρ_- and ρ_+ in the two model segments. In the full model below, then, we set k=0 and $\rho_- = \_\rho_+ = \rho$.

With this simplification, it is feasible to fit a fuller model with the current-period quality attributes listed above:

$$OVQ_3 = \beta_{Make}\,MAKE_3 + \beta_{First}\,FIRST_3 + \beta_{Clear}\,CLEAR_3 + \beta_{Price}\,PRICE_3 + \beta_{CallCS}\,CALLCS_3 + \rho\,OVQ_2 + (1-\rho)\,(\mu_0 + \beta_0\,OVQ_1)$$

The estimated parameters are:

Parameter	Estimate (Standard Error)
β_{Make}	0.447 (0.066)
β_{First}	0.123 (0.061)
β_{Clear}	0.026 (0.050)
β_{Price}	0.065 (0.039)
β_{CallCS}	-0.073 (0.097)
ρ	0.041 (0.045)
μ_0	1.181 (0.296)
β_0	0.071 (0.056)

Note again that the once-lagged quality rating OVQ_2 has only a small effect on the reference value, which is almost a constant. The reference value itself is small in relation to the contribution of the current-period quality attributes $MAKE_3$ and $FIRST_3$. This is a finding similar to that of Study I.

Further information on reference points is afforded by analysis of the variable $EXPSER_i$ (i=2,3), which compares overall service perception with expectation. This is effectively the expectency-disconfirmation measure that Oliver (1980) and others identify as a distinguishing ingredient in the customer satisfaction affect.

SUMMARY

We have considered the modeling of data from two studies of customer satisfaction, in which customer evaluations were measured at several points in time. The customer base (business vs. residential customers), the passivity of the service (local telephone vs. cellular), the service aspect (repair service vs. overall service) and interviewing interval (annual vs. biannual) were different in the studies.

These data afford an analysis of the cognitive mechanism by which service evaluations are created and updated. The language for this analysis comes, variously, from the disciplines of cognitive psychology, marketing science and survey research. From these two datasets, we have found that:

! Reference points, with which current evaluations are contrasted, seem to exist and are heavily influenced by initial or Aprehistoric@ quality evaluations.

! Initial reference points, formed before the survey period, can vary greatly among a customer base and can survive through at least 18 months of service.

! There is no evidence of any quality Ainertia@ based on a reference point, which requires an unusual quality change to overcome.

! Current period attributes play a far greater role in service evaluation than any reference point.

It follows that the general level of service evaluation is partly influenced by prior reference values, but this does not preclude incremental rating improvements based on current service attributes. In other words, aggregate survey results should not breed the fatalism characterized by the belief that opinions are preordained; current (and future) ratings follow service delivery to a very large extent. Instead of their traditional reliance on independent cross-sectional surveys, these results show that the wise company will develop satisfaction measurement programs to follow changes in time series ratings to assess their performance. It is improvement upon goodness, rather than goodness itself, that is most import in quality management.

REFERENCES

Boulding, W., A. Kalra, R. Staelin and V. Zeithaml (1993), "A Dynamic Process Model of Service Quality: From Expectation to Behavioral Intentions," *Journal of Marketing Research*, XXX (February 1993), 7-27.

Davis, J. H. (1984), "Order in the Courtroom," in *Perspectives in Psychology and Law*, D. J. Miller, D. G. Blackman and A. J. Chapman (eds.), New York: Wiley.

Drew, James H. (1990), "Current Findings Regarding BCOS Nonresponse and Stratification," GTE Laboratories Technical Memo.

Drew, James H. and Ruth N. Bolton (1996), The Dynamics of Customer Opinion, working paper.

Helson, H. (1964), *Adaptation Level Theory: An Experimental and Systematic Approach to Behavior*, New York: Harper and Row.

Hogarth, R. and H. Einhorn (1992), "Order Effects in Belief Updating: The Belief-Adjustment Model," *Cognitive Psychology*, 24, 1-55.

Kahneman, D. and A. Tversky (1973), "On the Psychology of Prediction," *Psychological Review*, 3, 430-454.

Kalton, G. and C. Citro (1993), "Panel Surveys: Adding the Fourth Dimension," *Survey Methodology*, 19, 2, 205-215.

LaBarbera, P. A. and D. Mazursky (1983), "A Longitudinal Assessment of Consumer Satisfaction/ Dissatisfaction: The Dynamic Aspect of the Cognitive Process," *Journal of Marketing Research*, XX (November 1983), 393-404.

Nisbett, R. and L. Ross (1980), *Human Inference: Strategies and Shortcomings of Human Judgment*, Englewood Cliffs, NJ: Prentice-Hall.

Oliver, Richard L. (1980), "A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions," *Journal of Marketing Research*, XVII (November), 460-9.

O'Neill, R. (1971) "Function Minimization Using a Simplex Procedure, Algorithm AS 47," *Applied Statistics*, 338.

Schul, Y. and M. Schiff (1993), Measuring Satisfaction with Organizations," *Public Opinion Quarterly*, 57, 4, 536-551.

Simmons, C., B. Bickart, and J. Lynch, Jr. (1993). "Capturing and Creating Public Opinion in Survey Research," *Journal of Consumer Research*, 20, 2, 316-329.

Waterton, J. and D. Lievesley (1989), "Evidence of Conditioning Effects in the British Social Attitudes Panel," in *Panel Surveys*, eds. D. Kasprzyk, G. Duncan, G, Kalton and M. Singh, New York: John Wiley and Sons.

Send correspondence regarding this article to:
James H. Drew
GTE Laboratories Inc.
40 Sylvan Road
Waltham, MA 02254 USA