Managers the world over believe that improving customer satisfaction leads to improved business results—and not without reason: There’s a lot of research that proves the relationship. Yet the same research shows that the relationship is weak. As one researcher noted in an article published in the *Journal of Retailing* in 2003, overall customer satisfaction seems to have only a modest impact on the share of wallet retailers get from their customers. And, in an award-winning article published in the *Journal of Service Research* in the same year, Keiningham and colleagues found a significant but weak relationship between customer satisfaction and share of wallet in financial services. In fact, research appears to show that up to 80 percent of what individual customers do with their money is unrelated to how satisfied they say they are in surveys.

Typically, two reasons are suggested for the weak relationship between customer satisfaction and individual behavior. The first is heterogeneity: Different kinds of customers appear to differ in their sensitivity to variations in satisfaction, or at least in their use of scales. The second is non-linearity: Research suggests that customer loyalty only increases when customers are “extremely” satisfied. Taking these factors into account leads to improved models, but some 70 percent of what people do with their money typically remains unexplained. Why should this matter?

A relatively simple logic underpins customer satisfaction research: key drivers ➔ customer satisfaction ➔ business outcomes ➔ (e.g., share of wallet). Marketers invest in customer satisfaction research in the hopes that it will lead to good business outcomes. But the poor correlations lead to a question: What’s the use of knowing what drives customer satisfaction if it is poorly correlated with customer spending? Our goal is to help marketing researchers get a better return on their customer satisfaction research by offering a better dependent variable. As we’ll show, the problem is easy to fix.
What’s the Correlation?

Let’s begin with a personal anecdote. Recently, we were doing business in Chicago and we stayed at a Westin hotel. When we checked out, we were asked to complete a short customer satisfaction survey. It asked if we were satisfied. We answered “extremely.” It asked if we would stay at a Westin hotel sometime in the future. We answered “definitely.” It asked if we’d recommend the Westin to others. Again we answered “definitely.” We answered every question at the top of the scale. The survey went on to ask some simple questions about touch-point experiences, and that was that.

As marketing researchers we know what will happen next—the data will go to researchers who’ll use it to update a “loyalty” index of some kind. The index is probably being tracked, and there’s a good chance that management key performance indicators (KPIs) will be linked to it. The researchers will perform a touch-point driver analysis to identify where managers need to focus if they want to improve the index. This will be used to guide management action.

But how well are our ratings of the Westin likely to correlate with our future behavior? The answer is “not very,” and the reason is simple: We both travel a lot and stay at many hotels, one of which is the Intercontinental. As it happens, we tend to rate the Intercontinental as highly as we tend to rate the Westin. Our equal ratings for both leads to a detachment from the individual customer satisfaction and behavior—from actual share of wallet as recorded on the panel. We were therefore able to analyze the relationship between a customer’s satisfaction with each retailer and the share of wallet he really allocated to each retailer.

To check our hypothesis, we recently did a survey among grocery customers in Italy. The respondents were members of a well-known retail consumer panel. This means we were able to combine each panelist’s attitudinal satisfaction as expressed in the survey with his actual share of wallet as recorded on the panel. We were therefore able to analyze the relationship between a customer’s satisfaction with each retailer and the share of wallet he really allocated to each retailer.

We used the following question to measure customers’ ratings of retailers: “Based on what you have seen or heard about each of the following retailers, even if you’ve never been there or never bought directly from them, I would like to know how you rate each retailer in terms of their ability to meet your needs when you do your shopping.” We used a 10-point scale anchored by the words “terrible” (1) and “outstanding” (10). Panelists were asked to rate the retailers they used or would consider using—in other words, their evoked set. There were 16 retailers in total. The average panelist used 3.2 retailers and considered a further 5.8. We therefore got ratings for an average of 9.0 retailers from each panelist.

Like researchers before us, we found a significant but poor relationship between individual customer satisfaction scores and shares of wallet (R = 0.21). However, when we transformed the ratings into relative ratings at respondent level, we got significantly better results. To do this we calculated the mean rating each respondent gave to the retailers they rated and then subtracted the mean from each retailer’s rating. Mean-centering in this way improved the correlation between individual customer satisfaction and behavior—from R = 0.21 to R = 0.40.

Improving the correlation by rescaling with a ranking approach. The main weakness of the mean-centered approach is that it’s not psychologically “truthful.” Mean-centering is equivalent to having a person compare each option with a mental average. In this way it is both comparative and takes competitor scores into account. But it’s not what real customers do. Real customers go directly to a comparison of their options. They don’t compare via an “average.” A process that gets closer to direct comparison than mean-centering is rescaling through ranking.

Ranking through rating involves the following: Ask each customer to rate the products, services or brands in their evoked set. Use the resulting scores to create a ranking at respondent level. Then calculate the relationship between individual behavior and the ranks.

Let’s illustrate this using our retail data. The retailer with the highest score gets a rank of “1” at respondent level. The one that comes second is ranked “2” and so on. This continues until all the rated options (i.e., the options in a respondent’s evoked set) are ranked. Non-rated options are automatically assigned a share of wallet of zero based on the assumption that any retailer that doesn’t get into the evoked set is unlikely to be used.

An important question remains: Why go to the trouble of asking a rating question first and then only ranking? Why not ask respondents to rank their options directly? The answer...
is that, in real life, share-of-wallet ties are ubiquitous. In consumer panel data for packaged goods, for instance, we’ve found that up to 35 percent of panelists use two or more brands in their repertoires equally. In the retail data we’re using in this example, some 31 percent of panelists had retailers in their repertoires that tied for share of wallet. We therefore have to allow for ties. Respondents find it easier to capture this indirectly through tied ratings than directly through saying which products, services or brands are tied. The question then becomes: How do you turn tied ratings into ranks?

There are two main methods for dealing with ties, namely, the “classroom” and the “prize money” methods. With the classroom method, if two options are tied for second place (for example), then they both get a rank of “2” and the next best option gets a rank of “4.” This is called the classroom method because it’s the way students are ranked in exams. With the prize money method, the two tied options would have occupied second and third spots had they not tied—so they share the ranks and each get “2.5.” This is the way prize money is usually shared.

Our research shows that the prize money method leads to better results. In ranking, higher satisfaction scores lead to lower (meaning better) ranks. In our data, for instance, the best score a business can get from a customer is “10,” but the best rank it can be is first. As a result, the correlation between the rank and share of wallet a business gets from its customers is negative. We can flip the negative into positive for comparative purposes. When we tested the prize money ranking method on our data, we got an improved correlation with share of wallet of \( R = 0.46 \).

By asking each respondent to rate relevant competitors, we improved our ability to predict the share of wallet each individual customer would give to each retailer by more than 115 percent, from \( R = 0.21 \) (old-fashioned satisfaction method), through \( R = 0.40 \) (mean-centering method) to \( R = 0.46 \) (prize-money ranking method). The line graphs below show the relationship between share of wallet and both old-fashioned satisfaction and ranks. At aggregate level, the relationship between satisfaction and share of wallet is very strong: Retailers got an average share of just 8 percent from customers who rated them “5” or less and 41 percent for retailers that were rated “10” (an increase of six notches). This is an improvement of more than 500 percent. But the relationship between rank and share of wallet is even stronger. Taking a retailer’s rank from sixth to first (also six notches) leads to an increase in share of wallet of more than 1000 percent—from about 7 percent to 72 percent. Ranking therefore leads to better results in at least two respects: (1) The correlation with individual behavior is some 115 percent better, and (2) the payback curve is steeper.

When scores are ranked, the relationship with scores disappears. One of the questions one might ask when seeing these
data is: Why, when the relationship between old-fashioned satisfaction and average share of wallet is so strong, is the relationship between individual satisfaction scores and share of wallet so weak \((R = 0.21)\)?

The problem, as we show in the line graph below, is that the increase in share of wallet isn’t explained by the increase in rating per se. It’s explained by the fact that the better a retailer’s rating, the higher its rank is likely to be. In other words, the improvements in share of wallet came mainly from improved ranks, not improved scores. No matter what a retailer’s score, if it was a panelist’s top-ranked retailer and that panelist only had two retailers in his evoked set, it got between 59 percent and 68 percent share of wallet. By contrast, the second-ranked retailer only got about 32 percent share of wallet no matter what its score.

Put another way: If the top-ranked retailer was rated “8,” it got 68 percent share of wallet as long as it was top-ranked. By contrast, even if the second-ranked retailer got “9,” it only got 32 percent share of wallet as long as another retailer—the top-ranked one—got “10.” Among panelists using three retailers, the share that went to the top-ranked retailer fell—from about 68 percent to about 58 percent. But the pattern remained the same. The second- and third-ranked retailers scored about 26 percent and 16 percent, respectively, irrespective of their scores. The fact that a high score such as “8” can sometimes be associated with being the best, but at other times with being second or third, leads to the poor individual correlations for the old-fashioned rating approach.

These results show that a business will do little to change the share of wallet it gets from individual customers unless that improved score leads to an improved rank. It’s about being first, not scoring high—though scoring high helps a business to be first. This is also why “extreme satisfaction” appears to drive high loyalty: The higher the satisfaction score, the more likely the business is to be a person’s top-ranked option. Some marketers erroneously concluded that it was only about delighting their customers. What it was actually about was being seen as No. 1. Delight might play a role in achieving this, but it’s not the main reason extreme scores lead to more business.

The relationship between rank and share of wallet is governed by a power law. The “business rank and share of wallet” line graph on page 10 shows a pattern of escalating ROI for higher ranks at individual customer level. This suggests that the relationship might be governed by a power law. That leads to the final refinement in this approach, namely, the transformation of the ranked customer satisfaction score into an implied share of wallet by the use of a power law. We call the resulting number the business’s attitudinal equity. The equation is:

\[
AE_i = \frac{100}{\sum_{i=1}^{n} \left( \frac{1}{\text{Rank}_{i}^{1.35}} \right)}
\]

where \(n\) is the number of options in a customer’s evoked set and \(i\) is the \(i\)th option in that set. Using the solver function in Excel, we found that the optimum value for \(s\) is 1.35.

The output from this equation is a number between one and 100 for every option—both client and competitors—in a customer’s evoked set. When all the scores are summed, they add up to 100. In other words, the equation turns ranks into a share of mind. The correlation of this attitudinal share
of mind with share of wallet is $R = 0.54$. Overall, by taking competition into account, creating rankings through ratings and using a power law, we improve the predictive power of satisfaction measurement by more than 150 percent, from $R = 0.21$ to $R = 0.54$.

Moreover, we do this with just one question, not three. And we get scores for both the client and the competitors.

Many managers may feel that getting additional ratings will make the questionnaire too long. In fact, however, it turns out to be quicker than conventional customer satisfaction approaches. Most commercial approaches measure customer satisfaction with three or more questions, typically, “satisfy,” “recommend,” “buy again.” These three questions are then used to derive an overall score or index for the client. In our experience it takes time for respondents to read each question and then rate the client. By contrast, our approach requires only one question. Less time is spent reading and, although more time is spent rating, people are so much faster at rating than reading that they take less time overall.

The Dangers of Delight

Why the doctrine of delight is a spending trap. We should pause for a moment to consider the doctrine of delight. The challenge that delighting customers poses for managers can be summed up in a word: habituation. As humans, we become used to the things that customer experience managers do to delight us. It’s a tendency with which we’re born, and it’s partly responsible for our evolutionary success. If it weren’t for habituation, we might never get restless or bored, and we might never learn anything. Habituation impels us forward and motivates us to improve things.

But for business managers, the fact that customers become habituated is a problem because it creates a constantly shifting target. We (your authors) still remember the first time we walked into a hotel room in the 1990s and found a flower and a chocolate with a welcome note on our pillows. It was a surprise. We were delighted. But quite quickly the novelty wore off, and very quickly all good hotels did the same. Soon we came to expect flowers and chocolates. When it didn’t happen, we became annoyed. And that’s the problem: A customer’s state of mind becomes, “Oh, that’s OK, but what are you going to do next?” Finding ways to delight customers creates new forms of spending that are quickly taken for granted, and just as quickly matched by competitors if they pose a competitive threat. Badly managed, the doctrine of delight can create a spending treadmill with a limited payback.

As an example, consider the airline industry. When British Airways introduced flat beds in the 1990s, it created a stir and lots of good publicity, and it considerably enhanced the experience of flying long distances. Soon other airlines had to match the offer, and over time they did. But the spending that the constant focus on enhanced customer experiences created in the airline industry opened up a gap for “no frills” airlines such as Ryan Air. These airlines offer just one thing: They’re cheap. Flying with them isn’t delightful. But they have a competitive advantage when it comes to something that’s important to people—namely, price.

Benchmarking—or how not to take competition into account. Benchmarking is the practice of comparing a business’s satisfaction scores with the scores collected for similar businesses in similar surveys. Scores above the benchmark are good. Scores below are obviously bad. Benchmarking requires a database and an investment in quality control when building the database to ensure that the correct benchmarks are chosen. And it creates expenses because of the logistics involved.

There are two problems with benchmarks. First, they’re backward-looking—by definition, you cannot have benchmarks for anything new a business might do to create competitive advantage. Second, they don’t get us closer to understanding individual customer behavior. Even when a business beats its benchmark and does better business, the weaknesses of the basic approach remain and the correlation of satisfaction with individual behavior stays poor. Customers are constantly comparing what they get from one product, service or brand with what they get from others. And it’s that comparison that needs to be taken into account if we’re to get better at predicting how individuals will behave.

The comparison doesn’t even have to come from competitors in the category. In the U.K. when Tesco (the retailer) announced that it would make sure queues at the checkout were never longer than three people, banks were put under pressure because people expected a similar attention to queues from the banks.

Like the doctrine of delight, benchmarks create a trap for customer experience managers. They lock managers into a backward-looking approach to customer management and into the use of dominant suppliers because good databases are needed to make them work. Further, it’s in the interests of suppliers to keep their clients backward-looking because that’s where the benchmarks are.

Our research suggests that businesses will improve the share of wallet they get from their customers, not merely by improving their customer satisfaction ratings, but by improving their ranks. Improved results only appear to flow from improved ratings because higher ratings tend to be associated with higher ranks. This being so, the idea that delight alone drives success is wrong. What drives success is being No. 1. Businesses are on a treadmill, the treadmill of maintaining a perceived competitive advantage in a world where innovation coupled with habituation creates a constantly rising performance bar.

Jan Hofmeyr is the international director of Synovate Laboratories and is responsible for research and development in the customer experience and branding spaces. He may be reached at Jan.Hofmeyr@Synovate.com. Ged Parton is a member of the Synovate Global Executive Board and Global CEO of all Synovate Practices and Capabilities. He may be reached at Ged.Parton@Synovate.com.