

The Relationship between Tipping and Service Quality: A Comment on Bodvarsson and  
Gibson's Article

Michael Lynn  
School of Hotel Administration, Cornell University

Numerous studies have found that tip size is only weakly related to service quality. Bodvarsson and Gibson recently challenged this finding—criticizing previous research and reporting that consumers say they would tip substantially different amounts with different levels of service quality. This paper presents a critical response to Bodvarsson and Gibson's article. Contrary to Bodvarsson and Gibson's claims, the weakness of the tipping–service relationship in the existing literature is not just a reflection of restricted variability in service ratings. Nor is it attributable to other methodological problems with service ratings. Furthermore, the data from Bodvarsson and Gibson's role-playing survey is suspect, because what people say they would do in a given situation is often different from what they actually do in that situation.

Restaurant patrons often give small sums of money (called “tips”) to the waiters and waitresses who have served them. These voluntary after-service payments are supposed to be an incentive/reward for good service (Schein et al., 1984). However, researchers in psychology and economics have found only a very weak relationship (mean  $r = 0.13$ ) between tip size and evaluations of service quality (Bodvarsson and Gibson, 1994; Crusco and Wetzel, 1984; Lynn and Grassman, 1990; Lynn and Graves, 1996; Lynn and Latane, 1984; Lynn, 1988; May, 1978). The weakness of this relationship raises questions about the importance of equity motives as a determinant of consumer behavior and about the efficacy of tipping as an incentive/reward (Lynn and Graves, 1996).

Table 1. Mean and SD of Percent Tip (Along with Sample Size) at Each Level of Customer-rated Service Quality for Four Published Studies

Service Rating	<i>Bodvarsson and Gibson (1984)</i>	<i>Lynn and Grassman (1990)<sup>a</sup></i>	<i>Lynn and Graves (1996: Study 1)<sup>a</sup></i>	<i>Lynn and Latane (1984: Study 1)<sup>b</sup></i>
1 (poor)	8.8%	—	14.6%	10.8%
	—	—	7.23	8.43
	(1)	(0)	(6)	(6)
2	10.3%	—	12.9%	12.0%
	3.07	—	5.87	6.27
	(3)	(0)	(9)	(13)
3	13.9%	2.8%	14.5%	15.3%
	4.89	4.81	3.86	11.62
	(38)	(3)	(23)	(21)
4	13.0%	10.0%	14.3%	16.4%
	4.20	4.38	5.89	8.53
	(189)	(23)	(50)	(78)
5 (excellent)	13.8%	11.5%	16.6%	16.1%
	3.94	4.70	5.33	13.21
	(466)	(77)	(85)	(51)

<sup>a</sup> Service was rated on several dimensions and the average of all the ratings was rounded to the nearest whole number.

<sup>b</sup> Service ratings were made on a 1–10 scale and were converted to a 1–5 scale for this table.

In a recent article appearing in this journal, Bodvarsson and Gibson (1999) challenge the finding that tip size is only weakly related to service quality. They raise three objections to existing tests of this relationship. First, they argue that ratings of service quality exhibit little variability and that this restriction of variability “biases” the size of the correlation between service ratings and tip amounts. Second, they argue that the rating scales used to measure service quality are ordinal and that treating the scales as cardinal data (by using them in correlation and regression analyses) is inappropriate. Finally, they argue that asking people to rate service quality falsely assumes that such ratings have comparable meanings across raters. Bodvarsson and Gibson (1999) then address these problems by asking people how much they would tip under hypothetical conditions of having received “satisfactory,” “very good,” and “poor” service. Survey respondents reported that they would tip substantially more for very good service (approximately 20%) than for satisfactory service (approximately 14%) or for poor service (approximately 6%).

This paper presents a critical response to Bodvarsson and Gibson’s (1999) article. First, each of their three objections to existing research is examined and shown to be non-problematic. Then, the validity of their hypothetical role-playing methodology is challenged.

## Distribution of Service Ratings

Bodvarsson and Gibson (1999) point out that customer ratings of service quality are overwhelmingly positive. They argue that this restriction of variability in service ratings biases tests of the tipping–service relationship and that the relationship would be stronger if the bias were eliminated. There is some truth to this argument. An examination of data from four published studies reveals that consumers' ratings of service quality do fall disproportionately in the upper ranges of the scales used (Table 1). Clearly, the correlation between tip sizes and service evaluations would be stronger if there were more variability in service ratings. However, there are two problems with this criticism of existing research.

First, restrictions in the variability of service ratings affect the size of the tipping–service relationship only as measured in terms of variance explained. The absolute difference in mean tip size at different levels of service quality provides a measure of effect size that is independent of variability in service ratings. Examining the mean tip size when service is rated a 3 out of 5 with that when service is rated a 5 out of 5 in Table 1 supports the idea that service quality has only modest effects on tipping. The differences in these means range from -0.1% to 2.1% of bill size in the three studies with sufficient sample sizes to make this comparison meaningful. This belies Bodvarsson and Gibson's (1999) argument that the weakness of the tipping–service relationship in the existing literature is primarily due to a lack of variability in service ratings.

Second, the effect of low variability in service ratings on the size of the correlation between tip size and service quality cannot legitimately be called a bias if the ratings accurately reflect consumers' perceptions of service quality. To sustain their argument that restricted variability in service ratings bias tests of the tipping–service relationship, Bodvarsson and Gibson would need to provide some evidence that the restricted variability in service ratings is due to response or sampling biases. They provided no such evidence.

Readers who doubt that restaurant service is as uniformly good as suggested by the data in Table 1 may be tempted to attribute the positive ratings to response bias and, thus, to accept a version of Bodvarsson and Gibson's thesis. However, if there is a positive response bias in ratings of service quality, then ceiling effects should result in the highest ratings of service being given to a set of service encounters with greater variability in true service quality than is true of the sets of service encounters receiving lower ratings. This heterogeneity in the variance of true service quality at different levels of rated service should also affect the variability of tip size if service quality is strongly related to tip size. However, an examination of Table 1 reveals no systematic differences in the variability of tip size at different levels of rated service quality. This suggests either that there is no sizable positive response bias in service ratings or that there is no sizable positive relationship between tip size and true service quality.

#### Ordinal Versus Interval Scales

Bodvarsson and Gibson (1999) also argued that the rating scales used in existing tests of the tipping–service relationship were ordinal and, therefore, provided data unsuitable for the statistical tests performed. This argument has no merit. Numerical rating scales such as those used in evaluations of service quality are commonly regarded as having equal (or nearly equal) intervals between response options (Churchill, 1996). This assumption is clearly justified when only the end points of the numerical response options are given verbal labels, because the numbered intermediate points on the scale represent equal intervals between the two endpoints. Contrary to the impression that Bodvarsson and Gibson (1999) gave, the rating scales used by Lynn and Grassman (1990) and others did not have verbal labels attached to intermediate points on the scales. Thus, these numeric scales were legitimately treated as interval level data. Even if the numeric scales were only ordinal as Bodvarsson and Gibson suggest, they would still be acceptable for use in correlational analyses because such analyses are robust with respect to monotonic transformations of data (Nunnally, 1978).

## Interpersonal Comparability of Ratings

Bodvarsson and Gibson's (1999) final criticism of existing tipping research is that its use of customer service ratings falsely assumes that the meaning of the ratings is comparable across persons. Although service quality is in the eye of the beholder, Bodvarsson and Gibson are wrong to claim that service ratings have no interpersonal comparability. Rating scales with labeled end points provide a metric that allows the subjective experience of one consumer to be compared with that of another consumer because the end point labels (e.g., "poor" or "excellent") do have a shared meaning in our culture. Indeed, Bodvarsson and Gibson (1999) themselves relied on the shared meaning of similar labels when they asked people how much they would tip under conditions of "satisfactory," "very good," and "poor" service.

## Validity of Role-Playing Data

Bodvarsson and Gibson (1999) address the problems they see with existing tests of the tipping-service relationship by asking people how much they would tip under hypothetical conditions of "satisfactory," "very good," and "poor" service. However, the respondents to this survey clearly overstated the effects of service quality on tipping behaviors. In that study, regular restaurant patrons from St. Cloud reported that they would tip 13.3% for satisfactory service and 19.4% for very good service. Yet in an earlier study involving post dining interviews with restaurant patrons as they were departing the restaurant, Bodvarsson and Gibson (1994) themselves found that people from the St. Cloud area tipped 13.9% for service rated a 3 out of 5 and tipped only 13.8% for service rated a 5 out of 5. The later of these two studies is the most trustworthy, because previous research has clearly demonstrated that people are often poor at predicting and explaining their own behavior (Myers, 1990). What people think and say they will do in a given situation is often different from what they actually do (Freedman, 1969). On the other hand, people can accurately report their own attitudes and can

accurately recount actions they have taken just moments before. Thus, the post dining, exit-interviews that Bodvarsson and Gibson (1999) now criticize are actually better than the new methodology that they employ. Bodvarsson and Gibson's real contribution to the tipping literature is the earlier study (Bodvarsson and Gibson, 1994) that they now disparage—it is not the study recently reported in this journal (Bodvarsson and Gibson, 1999).

#### References

- Bodvarsson, O. B. and W. A. Gibson. (1994). Gratuities and Customer Appraisal of Service: Evidence from Minnesota Restaurants. *The Journal of Socio-Economics*, 23: 287–302.
- Bodvarsson, O. B. and W. A. Gibson. (1999). An Economic Approach to Tips and Service Quality: Results of a Survey. *Social Science Journal*, 36: 137–147.
- Churchill, G. A. (1996). *Basic Marketing Research* (3rd ed.). New York: Dryden Press.
- Crusco, A. H. and C. G. Wetzel. (1984). The Midas Touch: The Effects of Interpersonal Touch on Restaurant Tipping. *Personality and Social Psychology Bulletin*, 10: 512–517.
- Freedman, J. L. (1969). Role Playing: Psychology by Consensus. *Journal of Personality and Social Psychology*, 13: 107–114.
- Lynn, M. (1988). The effects of alcohol consumption on restaurant tipping. *Personality and Social Psychology Bulletin*, 14: 87–91.
- Lynn, M. and A. Grassman. (1990). Restaurant Tipping: An Examination of Three “Rational” Explanations. *Journal of Economic Psychology*, 11: 169–181.
- Lynn, M. and J. Graves. (1996). Tipping: An Incentive/Reward for Service? *Hospitality Research Journal*, 20: 1–14.
- Lynn, M. and B. Latane. (1984). The Psychology of Restaurant Tipping. *Journal of Applied Social Psychology*, 14: 549–561.

May, J. M. (1978). *Tip or Treat: A Study of Factors Affecting Tipping Behavior*. Unpublished master's thesis, Loyola University of Chicago.

Myers, D. G. (1990). *Social Psychology*. New York: McGraw-Hill.

Nunnally, J. C. (1978). *Psychometric Theory* (2nd ed.). New York: McGraw-Hill.

Schein, J. E., E. F. Jablonski, and B. R. Wohlfart. (1984). *The Art of Tipping*. Wausau, WI: Tippers International, LTD.