

THE NORM OF RESTAURANT TIPPING

Michael Conlin
Department of Economics
Syracuse University

Michael Lynn
School of Hotel Administration
Cornell University

Ted O'Donoghue
Department of Economics
Cornell University

February, 2003

Abstract

Using survey data, we identify a variety of factors that influence tipping behavior and in the process lay out a simple theoretical framework to help to interpret our empirical observations. We first investigate the efficiency of observed tipping behavior. While there are elements of efficiency — notably, percent tip depends on service quality — it does not appear fully efficient. We then posit a model in which customers trade off material well-being against disutility from not adhering to the norm, and we use this model to reinterpret initial empirical findings and make additional empirical predictions.

JEL Classification: Z13, A12, D82.

Keywords: Fairness, Gift Giving, Reciprocity, Social Norm.

Acknowledgments: We are grateful to Steve Coate for numerous discussions, and we also thank J. Barkley Rosser, Jr. and two anonymous referees for helpful comments, and Mandar Oak for research assistance. For financial support, O'Donoghue thanks the National Science Foundation (Award SES-0078796).

Contact: Ted O'Donoghue; Department of Economics; Cornell University; 414 Uris Hall; Ithaca, NY 14853-7601. Phone: (607) 255-6287. Fax: (607) 255-2818. Email: edo1@cornell.edu.

I. Introduction

People tip billions of dollars per year in restaurants, both those that they frequent and those that they will never visit again. In fact, Kahneman, Knetsch, and Thaler (1986) report that in telephone surveys people state that the appropriate tip for a \$10 bill does not depend on whether it is a restaurant the person frequents or a restaurant the person will never visit again. From an economics perspective, this tipping behavior is hard to understand in terms of pure self-interest. In this paper, we examine restaurant tipping as a behavioral norm. Using survey data, we identify a variety of factors that influence tipping behavior, and in the process we develop a simple theoretical framework to help to interpret our empirical observations.

Many researchers have studied the role of “norms” for economic behavior; we briefly discuss this literature in Section 2. An important question in this literature is why a society would choose to inculcate a norm into its members. Some economists have argued that norms promote efficiency in realms where the market system fails (e.g., Arrow (1971)). A second question is what determines a person’s actual behavior in the presence of a norm. We discuss our empirical results in terms of both of these questions.¹ Our interpretation of restaurant tipping behavior is closely related to the literature on norms of gift giving and reciprocity (Akerlof (1982) and Fehr, Kirchsteiger, and Riedl (1993)); our results provide a real-life illustration of the behavioral power of these norms.

Our data consist of survey information collected outside of 39 restaurants in Houston, Texas. Information was obtained on the size of the tip, the size of the bill, the number of individuals in the party, characteristics of the meal, service quality, demographic characteristics of the server and tipper, the number of times the tipper frequents the particular restaurant, and the number of times the tipper frequents any restaurant. The survey data was augmented with information on the number of seats at the restaurant and whether the meal occurred on the weekend. We describe this data in more detail in Section 3.

In Section 4, we investigate whether the tipping behavior we observe appears to be efficient. There is an obvious reason why the institution of restaurant tipping might be efficiency enhancing. If efficiency requires the server to exert some effort, the server must have an

¹ A third question often asked in the norms literature is how a norm arose. Our analysis has very little to say about the origins of tipping, but we refer the reader to interesting discussions in Hemenway (1984) and Lynn, Zinkhan, and Harris (1993).

incentive to exert this effort. In principle, this incentive could be provided by a service contract between the customer and the server. However, since such contracts would involve prohibitive transactions costs, the norm of restaurant tipping may serve as a substitute. Our focus in Section 4 is whether our observed tipping behavior resembles what one would expect to see if tipping were serving as an implicit service contract.

The most obvious feature of an efficient tipping contract is that the tip should increase with service quality, and this relationship is supported in our empirical analysis.² However, we also investigate other features of an efficient tipping contract. First, we examine how tipping behavior depends on whether the meal occurs on a weekend or a weekday. Second, we examine how tipping behavior depends on repeated interaction. Our empirical results suggest that tipping behavior responds to these factors in ways that are inconsistent with an efficient tipping contract. We conclude that while there may be elements of efficiency in the norm of tipping, it is not fully efficient.

Given that tipping behavior is not determined by an explicit tipping contract, but by the degree that people adhere to a norm, it is not surprising that tipping behavior is not fully efficient. In Section 5, we focus more directly on why people tip. In particular, we posit a simple theoretical model for the determinants of tipping behavior and then investigate whether our observed tipping behavior is consistent with this model. Our model is based on two hypotheses. First, we posit that a person experiences disutility from not adhering to the norm, and in deciding how much to tip, the customer trades off material payoffs against this disutility. Second, we posit that the norm itself is based on internalized concerns for fairness (i.e., that the server deserves to be rewarded for his efforts). This model suggests that tipping behavior should depend on features of the environment that influence the degree to which the customer cares about adhering to the norm. This model also suggests that tipping behavior should depend on factors that influence the customer's perceptions of the server's effort because such factors change the norm (i.e., the appropriate tip). We use this model both to reinterpret some of our initial empirical findings from Section 4 and to make additional empirical predictions that we test in Section 5. We conclude in

² Previous empirical studies have also found that tip increases with service quality (see in particular Bodvarsson and Gibson (1997), and Lynn and McCall (2000)).

II. Norms and Economic Behavior

Many researchers have studied the role of “norms” for economic behavior. One strand of literature focuses on norms as a form of equilibrium selection (see, for instance, Sugden (1989), and Basu (1997,2001)), which are perhaps better called “conventions”. A second strand of literature focuses on “social norms” that are enforced by the threat of direct social sanctions. Under this conceptualization, if a person violates the norm, she will be directly punished by other people. Such punishments might take the form of membership denial in social groups, or less favorable social treatment (e.g., failure to accord the violator with the usual social respect) or literal punishment (see, for instance, Akerlof (1980), Axelrod (1986), Elster (1989*a*,1989*b*), Besley and Coate (1992), Bernheim (1994), Lindbeck et al (1999), Posner and Rasmusen (1999) and Fehr and Gächter (2000)). A third strand of literature analyzes “internalized norms”: Even when a person has no concerns about direct social sanctions, she may feel embarrassment, anxiety, guilt, shame, or some other negative feeling if she violates a norm (see Elster (1989*a*,1989*b*), Basu (1997,2001), Posner and Rasmusen (1999)).

Restaurant tipping is clearly not an equilibrium-selection norm. While it is possible that direct social sanctions play a minor role in restaurant tipping behavior, it cannot be the major enforcement mechanism. People too often tip in situations where they will never incur social sanctions. We suspect that restaurant tipping is primarily an internalized norm, and much of our analysis, particularly in Section 5, proceeds under this interpretation.

An important question in the literature on norms — particularly for social and internalized norms — is why a society would choose to inculcate a norm into its members. A common answer among economists is efficiency, or collective optimality. An early proponent of this view was Arrow, who wrote (p. 22), “I suggest as one possible interpretation that they [norms of social behavior, including ethical and moral codes] are reactions of society to compensate for market failures;” and (p. 22) “There is a whole set of customs and norms which might be similarly interpreted as agreements to improve the efficiency of the economic system (in the broad sense of satisfaction of individual values) by providing commodities to which the price system is

inapplicable.”³ In Section 4, we analyze whether the norm of restaurant tipping promotes efficiency.

A second question — again particularly relevant for social and internalized norms — is what determines a person’s actual behavior in the presence of a norm. Do people merely behave exactly as the norm requires, or do they trade-off adhering to the norm against material payoffs? What behavior does the norm call for in the first place? By postulating answers to these questions, one can make predictions about how norms influence behavior; we take precisely this approach in Section 5.

Our approach to restaurant tipping behavior is closely related to the literature on norms of gift giving and reciprocity. Akerlof (1982), for instance, investigates how such norms might operate in labor markets. Specifically, he describes how, for some occupations, norms call for workers to labor in excess of minimum work standards and call for firms, in return, to pay these workers a wage in excess of their outside option. Akerlof builds a simple model in which norms for worker effort are determined endogenously, worker effort decisions are influenced by these norms, and firms take into account both effects when making wage offers. Although Akerlof’s motivating real-world example — the behavior of “cash posters” at Eastern Utilities Co. (reported by Homans (1953,1954)) — involves both repeated interaction and the possibility of direct social sanctions, Fehr, Kirchsteiger, and Riedl conduct an experimental test of Akerlof’s theory that eliminates such concerns. They find striking evidence of reciprocal (gift-giving) behavior that is inconsistent with pure self-interest. Our results on restaurant tipping behavior provide a further, real-life illustration of the behavioral power of these norms.⁴

Finally, our analysis in Section 5 is closely related to the recent literature on social preferences (see Rabin (1993), Fehr and Schmidt (1999), Bolton and Ockenfels (2000), and Charness and Rabin (2002)). This literature proposes a variety of ways that people might care about social outcomes and not just private outcomes. The usual structure assumed in these models involves people trading off their own material payoffs against their concerns for social

³ Of course, efficiency does not provide an explanation for how the norm came into being in the first place, nor does it provide a justification for why the norm persists. Indeed, Elster (1989*a*,1989*b*) explicitly argues that many norms are socially inefficient.

⁴ See Berg, Dickhaut, and McCabe (1995) for additional experimental results demonstrating reciprocal (gift-giving) behavior in a one-shot, anonymous situation. The behavior observed both by Fehr, Kirchsteiger, and Riedl and by Berg, Dickhaut, and McCabe is qualitatively similar to restaurant tipping behavior: First, one party makes a “gift” to a second party, and then the second party makes a return “gift” to the first party.

outcomes, and we use a similar structure.

III. Data Description

The data consist of survey information obtained in 112 survey sessions outside of 39 restaurants in Houston, Texas. The survey was conducted at a wide variety of restaurants including those serving Mexican, Italian, Thai, and American food. Each survey session was conducted between 6pm and 10pm at a single restaurant. The interviewer surveyed individuals leaving the restaurant who paid the bill and left the tip.

For each tipper, information was obtained on the size of the tip, the size of the bill, the number of individuals in the party, the number of courses ordered, whether alcohol was consumed, the food quality (appearance, portion size, taste, temperature, and price), the service quality (appearance of server, knowledge of menu, friendliness, speed, and attentiveness), gender of server, gender of tipper, age of tipper, the number of times the tipper frequents the particular restaurant, and the number of times the tipper frequents any restaurant.⁵ The survey data was augmented with information on the number of seats at the restaurant and whether the survey was conducted on the weekend (i.e., Friday, Saturday or Sunday). A detailed description of the survey is contained in the Appendix.

Summary statistics are contained in Table 1. The survey resulted in 1,393 observations, with the average tip size being \$4.30 and the average bill size being \$26.42. Rather than tip in dollars, we shall mostly focus on percent tip (tip/bill); the average percent tip is 17.56%, or slightly more than the “norm” of 15%.⁶ The number of individuals in a party ranges from one to five and averages 2.37 in our sample.⁷ The average number of courses (where a course consists of an appetizer, soup, salad, entrée, or dessert) is 1.96, and alcohol is consumed in slightly more than half the observations. The food quality and service quality measures are each based on a

⁵ The interviewer also asked if there were separate checks distributed at the table and if more than one person contributed to a single check. As discussed in the Appendix, the observations with separate checks or when more than one person contributed are not used in the estimation.

⁶ We put quotation marks around “norm” here to emphasize our belief that the underlying norm is more complex than merely being 15% of the bill. We note that there may be significant “rounding off” in our data. Of the 1393 observations, 904 have a tip of a whole dollar amount and 769 have the sum of the tip and bill as a whole dollar amount. Because we do not believe this rounding off is correlated with the independent variables in our specifications, it should not bias our empirical results.

⁷ As discussed in the appendix, observations where the size of the party is greater than five are not used in the

five point scale with 1 being poor and 5 being excellent. Slightly over a third of all tippers had a female server and a quarter of the tippers in our sample are female. The age of the tipper is classified into four categories: teenager (1), young adult (2), middle aged (3), and elderly (4). Based on these numerical designations, the average of the age variable is 2.68. The average number of times per month that the tipper frequented the particular restaurant she just left is 2.09. Forty percent of tippers frequented the particular restaurant more than once a month. The average number of times per month that the tipper frequented any full-service restaurant is 11.55. Eighty-four percent of tippers frequented a full-service restaurant at least four times a month.⁸ Only eight percent of the interviews were conducted on a Friday, Saturday or Sunday and the average number of restaurant seats is 212.

Our empirical results focus on the determinants of percent tip and service quality.⁹ Table 2 presents our most basic empirical results that show how percent tip and service quality depend on various control variables. To examine the determinants of percent tip, we estimate models of the following form:

$$\text{Percent Tip} = \alpha_i + \beta \mathbf{X} + \varepsilon.^{10}$$

The vector \mathbf{X} consists of a variety of control variables: the size of the bill, the size of the bill squared, the size of the group, the number of courses ordered, the consumption of alcohol, the food attributes, gender of server and tipper, age of tipper, and the frequency of the tipper's visits to full-service restaurants. The variable α_i represents the fixed effect associated with interviewer i . Because each interviewer surveyed tippers from only one restaurant and conducted these surveys from 6pm to 10pm, these fixed effects not only capture the characteristics of the interviewer but also the characteristics of the restaurant and the evening on which the interviewer was conducting the survey. The unobserved factors, ε , are assumed to be normally distributed and independent of \mathbf{X} . These unobserved factors are allowed to be correlated for observations from the same interviewer.

estimation. Often, the tip for these larger parties is automatically included in the bill.

⁸ Because our sample is not randomly selected but instead consists of tippers at restaurants, frequent diners are likely over-represented in our sample. This explains the seemingly high restaurant usage by people in our sample.

⁹ We use percent tip as our dependent variable because we suspect that most people think about tips in percentage terms. Indeed, Mills and Riehle (1987) report from survey evidence that 70 percent of people figure their tips based on a percentage of the check.

¹⁰ Although the dependent variable has a limited range, we feel a standard linear specification is appropriate given

The results of this specification are presented in the first column of Table 2.¹¹ The coefficient estimates associated with bill size, group size, the consumption of alcohol, the age of the tipper, and the frequency of the tipper's visits to full-service restaurants are statistically significant and have relatively large marginal effects on percent tip. In particular, percent tip decreases with bill size at a decreasing rate;¹² percent tip increases with group size and consumption of alcohol; percent tip decreases with the age of the tipper; and percent tip is larger for individuals who frequent full-service restaurants more often. This last variable may be a proxy for the individual's income, since higher income individuals are likely to eat more often at full-service restaurants. These results are robust to all of our specifications. In Section 5, we interpret some of these coefficients in relation to our model of the determinants of tipping behavior.

Four out of the five coefficients associated with the food measures are positive but none are statistically significant. However, the coefficients are jointly significant at the five percent level. We often aggregate these variables into a single food measure obtained by taking the average of the five variables. Column 2 of Table 2 contains the results when the aggregate food measure is included as an independent variable. The coefficient associated with this variable is positive and statistically significant at the five percent level.¹³ Note also that the coefficient estimates associated with the other independent variables are similar whether we use the individual food measures or the aggregate food measure.

To examine the determinants of service quality, we estimate models of the following form:

$$\text{Service Quality} = \alpha_i + \beta \mathbf{X} + \varepsilon .$$

that only 6 observations have zero tip and only 11 observations have less than a 5% tip.

¹¹ The empirical results when tip is the dependent variable are similar to those when percent tip is the dependent variable. The main differences are (i) the coefficients associated with bill size and bill size squared are both positive and statistically significant, (ii) the coefficient associated with group size remains positive but is not statistically significant, and (iii) the coefficient associated with whether the tipper is female is positive and statistically significant.

¹² The parameter estimates imply that percent tip is decreasing in bill size for any bill under \$90.50. Bill size was larger than this amount in only 19 of the 1,393 observations.

¹³ Of course, the question arises why *should* percent tip depend on food quality. Indeed, we suspect that in Table 2 food quality is related to percent tip because it is serving as a measure of repetition — that is, people who like the food at a particular restaurant are more likely to frequent that restaurant. This suspicion is supported by the fact that the coefficient associated with the food quality measure decreases substantially once we control for repetition.

The dependent variable is a measure of service quality obtained by taking the average of the scores for appearance of server, friendliness of server, speed of service, and attentiveness of service.¹⁴ The vector \mathbf{X} contains the same variables as in the prior specification, and again we include indicator variables for each interviewer. The results of this specification are presented in the third column of Table 2. The coefficient estimates suggest that many of the factors that are important for percent tip are not important for service quality; in particular, bill size, the consumption of alcohol, age of tipper, and the frequency of tipper's visits to full-service restaurants are not economically nor statistically significant.

IV. Is Restaurant Tipping Efficient?

In this section, we investigate whether our observed tipping behavior appears to be “efficient.” We consider a narrow definition of efficiency purely in terms of economic variables such as server's effort, service quality, and monetary payments. We do not incorporate into efficiency behavioral factors such as utility from altruism, guilt, or concerns for fairness. Our approach in this section is to assume that people's tipping behavior is being driven by some norm and to ask whether the observed behavior seems consistent with that norm promoting efficiency.

4.1 A Theoretical Framework

There is an obvious reason why the institution of tipping might be efficient in the restaurant industry. While service quality is an integral part of the customer's dining experience, service quality requires effort on the part of the server, and hence an appropriate compensation scheme with monitoring is required to induce the optimal level of effort. Because the customer is in a better position to observe the quality of service than is the restaurant owner, absent transaction costs, social optimality requires that the customer and the server write a service contract. However, since writing a service contract upon every visit to a restaurant would involve

¹⁴ We do not include the score for knowledge of the server when calculating this average. As we discuss in Section 4, knowledge of server may have been interpreted by the interviewee in a negative manner resulting in it being a poor measure of service quality. The empirical results do not change appreciably when knowledge of server is included in the service quality measure.

prohibitive transaction costs, the norm of restaurant tipping may serve as a substitute.¹⁵

The most basic feature of an efficient tipping contract is that, under the reasonable assumptions that the customer cares about service quality and that the customer and server are both risk-neutral over the tip paid, tip should increase with service quality. While this implication is rather intuitive, it will prove useful to develop a simple theoretical framework. Suppose that each time a customer dines at a restaurant, the server must choose how much effort e to exert. The customer cannot observe the server's effort level, but does experience service quality Q that is related to the server's effort. In particular, if the server chooses effort level e , then the service quality Q experienced by the customer has cumulative distribution function $F(Q|e, \theta)$, where θ represents other factors over which the server and customer have no control. Throughout we shall use $E(\bullet|e, \theta)$ to be the expectation operator given that Q is distributed according to $F(Q|e, \theta)$.

Because the magnitudes of tips in our dataset are very small (the mean tip in our sample is \$4.30) we assume for simplicity that both the customer and the server are risk-neutral.¹⁶ Letting T denote the tip paid by the customer to the server, we assume the customer's preferences are characterized by utility function $U^C(Q, T, \theta) = v(Q, \theta) - T$, where v is increasing and concave in Q . The server's preferences are characterized by utility function $U^S(T, e) = T - c(e)$, where c is increasing and convex. The customer and the server both being risk-neutral implies that any tips paid are merely monetary transfers and do not affect efficiency *per se*. Hence, the efficient level of effort is $e^* = \operatorname{argmax}_e [E(v(Q, \theta)|e, \theta) - c(e)]$.

Because the customer cannot observe the server's effort, a tipping contract can only specify the tip to be paid as a function of service quality. Let $T(Q)$ denote the tip paid as a function of perceived service quality Q . It is straightforward to derive that in order to induce the efficient effort level, the tipping contract should internalize the customer's marginal preference

¹⁵ Jacob and Page (1980) build an abstract model in which a firm's buyers care about the behavior of the firm's employees and can better monitor the behavior of the firm's employees. They show that it can be optimal for the buyers to contract with the employees, and they suggest restaurant tipping as an (implicit-contract) example. Also see Ben-Zion and Karni (1977).

¹⁶ Under expected-utility theory, any risk aversion over such small stakes must be negligible (Rabin (2000)). Even so, we discuss in certain places how risk aversion might change qualitative predictions of our theoretical framework.

for service quality; that is, the optimal tipping contract should take the form $T(Q) = T_0 + v(Q, \theta)$.¹⁷ Hence, if v is increasing in Q — the customer prefers higher service quality — then tip should be increasing in service quality.

Before we examine empirically whether tips depend on service quality, we discuss a few caveats. First, our theoretical framework above ignores the fact that a single waitperson usually provides service for multiple customers. Determining the efficient (from a collective perspective) tipping contract in such a situation is similar to determining the optimal contract in the multitask principal-agent model of Holmstrom and Milgrom (1991) or multiprincipal-and-single-agent model with “united principals” of Dixit (1996). These authors’ analyses imply that, if the server were *risk averse*, then the strength of the incentives would be decreasing in the number of customers. However, given our assumption that the server is risk neutral, our conclusion above still holds: for each customer, the tipping contract should internalize that customer’s marginal preference for service quality. More generally, we suspect that, for any plausible degree of risk aversion, our qualitative conclusions about efficient tipping behavior would hold even if we explicitly incorporated the fact that one waitperson serves multiple customers.

A closely related issue, although orthogonal to the question of efficiency, is whether the existence of multiple customers for one waitperson creates unhealthy competition between customers. In other words, customers might use the lure of a higher tip to convince the server to exert more effort towards them and less effort towards others.¹⁸ While some customers certainly behave in this fashion, we suspect this issue is not a major concern for our analysis because customers rarely discuss potential tips with waitpeople prior to meals. Of course, whether such competition is likely to occur may help to explain why tipping arises in some arenas (and countries) but not others.

A final caveat involves the fundamental asymmetry in our model between customers and servers. Our analysis assumes that, in contrast to pure self-interest, customers leave tips because their behavior is disciplined by a tipping norm. This raises the question of why a high-effort

¹⁷ If $T(Q) = T_0 + v(Q, \theta)$, the server will choose $e^s = \operatorname{argmax}_e [T_0 + E(v(Q, \theta)|e, \theta) - c(e)]$, and clearly $e^s = e^*$ for any T_0 , v , F , and c . For some examples (e.g., if Q or e were chosen from a discrete set) other tipping schemes could also induce the efficient effort level. Even for such cases, we believe the optimal tipping contract that internalizes the customer’s preferences is the natural contract to examine because it does not require any additional information about the exact nature of v , F , and c .

¹⁸ One can see this formally by considering the game in which customers simultaneously announce (binding) tipping contracts and then the server chooses a vector of efforts (similar to the “separate-principals” case in Dixit’s analysis).

norm cannot discipline server behavior, making the tipping norm unnecessary. We suggest two possible justifications for this asymmetry. First, there is an asymmetry in how easily people could simultaneously disobey a norm while convincing themselves that they had obeyed it. If a high-effort norm were used, servers might justify (to themselves) putting in low effort by deciding, for instance, that they feel ill or depressed. While such “excuses” also exist for a tipping norm, they seem less readily available. A second justification is an asymmetry in the source of material payoffs. Obeying a tipping norm means customers must pay a few extra dollars, whereas obeying a high-effort norm means servers must exert more physical effort. Under our interpretation, a norm can discipline behavior only if the negative emotions experienced upon disobeying the norm are sufficient to induce the person to forego material payoffs. If negative emotions are more effective at creating incentives to leave a few extra dollars than at creating incentives to exert more physical effort, then a tipping norm is more effective than a high-effort norm.

4.2 Tips Depend on Service Quality

To examine empirically whether tips depend on service quality, we estimate the following model:

$$\text{Percent Tip} = \alpha_i + \beta \mathbf{X} + \chi(\text{Service Quality Measures}) + \varepsilon.$$

This model is identical to that in our basic regression except that we now include as independent variables the five service measures.

The results of this specification are presented in the first column of Table 3. The coefficient estimates for the independent variables \mathbf{X} are much as before. The coefficients of interest are those associated with the service measures. Our results indicate that knowledge, friendliness, and speed of the server have a much larger effect on the tip than appearance and attentiveness of the server. The marginal effect of increasing the friendliness rating by one point is to increase tip by 1.336 percentage points, and the marginal effect of increasing the speed rating by one point is to increase tip by 1.196 percentage points. While the coefficients associated with the friendliness and speed variables are positive (as expected), the coefficient associated with knowledge is negative. When the interviewers asked the tippers how

knowledgeable their server was, the tippers possibly interpreted the question in a negative manner, explaining the negative coefficient. For instance, perhaps the more knowledgeable a tipper rated the server, the more the server bothered the tipper with unimportant information. Another possibility is that the knowledgeability question was picking up the degree to which the tipper perceived the server to be a “snob”.

It will often be convenient to use a single measure of service quality that we obtain by taking the average of the scores for appearance of server, friendliness of server, speed of service, and attentiveness of service.¹⁹ The results when the aggregate service quality measure is included in the prior specification are in Column 2 of Table 3. The coefficients associated with the non-service variables do not change appreciably from the prior specification. The coefficient associated with the aggregate service quality measure is positive and both economically and statistically significant. These results provide evidence that tips do increase with the level of service. People are not merely leaving a 15% tip; they are tipping as a function of service.

The reader may wonder whether people tip as a function of service only if they are concerned about future interactions. While we shall shortly discuss repetition in some detail, here we briefly illustrate that tips depend on service quality even for people who cannot have large concerns about the future. In particular, it seems reasonable that future interactions are not a major concern for those tippers who frequent the particular restaurant less than once a month. Hence, we estimate the specification above using only those observations where the tipper frequented the particular restaurant less than once a month. The coefficient estimates based on these 874 observations are presented in Column 3 of Table 3. While the standard errors of several coefficients increase as the result of the smaller sample size, the magnitudes of the coefficient estimates (including the service quality measures) do not change appreciably.

Hence, our data are consistent with the most basic feature of an efficient tipping contract: Tip depends on service. Previous empirical studies have confirmed this prediction (see in particular Bodvarsson and Gibson (1997) and Lynn and McCall (2000)). However, we now test for other features of an efficient tipping contract.

¹⁹ We do not include the score for knowledge of the server when calculating this average. As mentioned, knowledge of server may have been interpreted by the tipper in a negative manner, resulting in it being a poor measure of service quality. The empirical results presented in Tables 3 and 4 do not change appreciably when knowledge of

4.3 Tips Depend on “Noise”

Any factor that alters the relationship between effort and service quality can potentially alter tipping behavior. One such factor in our dataset is whether the meal occurs on a weekday or a weekend. The relationship between the server’s effort and perceived service quality depends on things such as the demands of other customers and the speed and sequence with which meals come out of the kitchen. Because restaurants are normally busier on weekends, we expect the relationship between effort and the speed of service variable to be weaker for meals that occur on a weekend rather than a weekday. We often speak of this as an increase in “noise.”

To explore the theoretical implications of noise for the efficient tipping contract, we return to our earlier framework. An increase in noise will change the relationship between effort and service quality, that is, it changes $F(Q|e, \theta)$. However, since the efficient tipping contract merely internalizes the customer’s marginal preference for service quality and is independent of F , an increase in noise should not change the efficient tipping contract. Hence, our theoretical framework implies that in the efficient tipping contract, the sensitivity of tips to service quality should not depend on whether a meal occurs on a weekend rather than a weekday.

As discussed above, we believe the service measure most likely to be influenced by the day of the week is speed of service. To examine empirically how noise affects the sensitivity of tips to speed of service, we interact whether the tip occurred on the weekend with this measure and include this interactive term as an independent variables in the prior specification. Column 4 of Table 3 presents the results. The coefficient associated with the weekend/speed-of-service interactive variable is negative and statistically significant.²⁰ These regression results indicate that a one-point increase in the speed-of-service rating will increase the tip by 1.319 percentage points if the meal occurred on a weekday and only by 0.299 percentage points if the meal occurred on a weekend. While the sensitivity of tips to service should not depend on noise in an efficient tipping contract, these results indicate that increased noise tends to make tips less sensitive to service.²¹ This suggests that tipping behavior is not fully efficient.²²

server is included in the service quality measure.

²⁰ The direct effect of whether the tip occurred on a weekend is accounted for by the interviewer indicator variables.

²¹ One might think a similar story should hold for attentiveness. If we include an attentiveness/weekend interaction term, its coefficient is negative but close to zero; the magnitude of the coefficient on the speed-of-service/weekend interaction term does not change, but is no longer statistically significant at conventional levels.

²² Under some additional and potentially restrictive assumptions, an increase in noise has implications for expected service quality and expected tip. If noise reduces the marginal return to effort (that is, for any $e_2 > e_1$, $E(v(Q,n)|e_2, \theta)$

4.4 Tips Depend on Repeated Interaction

A second feature of the environment that might influence tipping behavior is whether there is repeated interaction. Repetition may permit a more “collusive” tipping contract, analogous to the ability of firms to collude in repeated interactions. In the tipping environment, repeated interaction permits the customer to punish the server for poor service and the server to punish the customer for a small tip in subsequent interactions. If repetition permits a more collusive agreement in this way, then repetition could lead to larger expected tips and higher expected service quality. However, repetition can have these effects only if the one-shot tipping contract is not fully efficient; that is, if one-shot contracts were fully efficient, as they are in our simple theoretical framework, then there is no additional surplus to be gained from collusion.²³

To examine empirically how repeated interaction affects percent tip, we estimate the following model:

$$\text{Percent Tip} = \alpha_i + \beta \mathbf{X} + \chi (\text{Service Quality}) + \zeta (\text{Tipper Frequent Particular Restaurant}) + \varepsilon .$$

In this specification, we use the aggregate measures for food and service quality. The number of times per month the tipper frequents the particular restaurant is also included as an independent variable. The results of this specification are presented in Column 5 of Table 3. The coefficient estimate associated with the number of times the tipper frequents the particular restaurant is positive and statistically significant. As for the marginal effect, tip increases by an average of 0.187 percentage points if the customer frequents the particular restaurant one additional time per month. Hence, we see evidence that expected tips are larger in repeated interactions (see Lynn

– $E(v(Q,n)|e_i, \theta)$ is decreasing in noise) then the efficient effort level e^* is smaller on weekends. If in addition the expected service quality $E(Q|e, \theta)$ is increasing in e and nonincreasing in noise, then expected service quality is lower on weekends than on weekdays. Finally, if the server’s share of the surplus is not larger on weekends than on weekdays, the expected tip should be smaller on weekends. We tested for these implications in our data and found no significant relationship between percent tip and weekends, but a negative and statistically significant relationship between service quality and weekends. We are cautious in making too much of these results because we had to exclude the interviewer indicator variables from these regressions due to collinearity with the weekend indicator variables.

²³ Even when one-shot interactions are fully efficient, repetition could matter if it changes the relative bargaining power and therefore the distribution of surplus. However, a change in relative bargaining power should not change the expected service quality, and we observe a change in service quality in the data.

and McCall for qualitatively similar empirical results).^{24,25}

To examine empirically how repeated interaction affects service quality, we estimate the same model as above with service quality as the dependent variable. The results from this specification are presented in the last column of Table 3. The coefficient associated with the number of times the tipper frequents the particular restaurant is positive and both economically and statistically significant. This result implies that, in addition to its direct effect on percent tip (identified in the column (5) of Table 3), repetition also has an indirect effect on percent tip: Repetition leads to better service quality which in turn results in a higher percent tip.

These empirical results are consistent with repetition leading to more collusive outcomes, providing further evidence that tipping behavior is not fully efficient in one-shot interactions.²⁶

V. Why Do People Tip?

In the previous section, we found that while the norm of tipping seems to have elements of efficiency, it does not appear to be fully efficient. Given that tipping behavior is not determined by an explicit tipping contract, but by the degree to which people adhere to a norm, it is not surprising that tipping behavior is not fully efficient. In this section, we focus more directly on why people tip.

As discussed in Section 2, we suspect that negative feelings associated with violating a norm is the main enforcement mechanism for restaurant tipping. Of course the question remains

²⁴ The implications of repetition for the sensitivity of tips to service quality is unclear because the answer depends on how one-shot interactions are inefficient. We empirically tested for a relationship by including as independent variables the interaction between service quality and the frequency of the tipper's visits to the particular restaurant. This test yielded no evidence that the relationship between tip and service quality depends on how often the tipper frequents the particular restaurant.

²⁵ In principle, patronage frequency should matter more for small restaurants than it does for large restaurants. When we included a patronage-frequency/restaurant-size interaction term in our regressions, we found that restaurant size had no significant effect on the relationship between tip and patronage frequency.

²⁶ Our claim that a single interaction could be fully efficient relies on our assumption that the customer and the server are both risk-neutral over tips. If the server were instead risk-averse, then the optimal one-shot contract may not be fully efficient because creating incentives for high effort may prevent optimal risk sharing. Indeed, the repeated-moral-hazard literature (e.g., Rogerson (1985) and Fudenberg, Holmstrom, and Milgrom (1990)) emphasizes how repeated interaction permits intertemporal risk-sharing and therefore more efficient contracts. But the assumption of a risk-averse server has implications that are not borne out by our data. First, repetition should decrease the sensitivity of tips to service quality, the essence of intertemporal risk-sharing. Second, repetition should increase expected service quality, because less risk is imposed on the server and therefore service is cheaper to buy. Third, if the server's individual rationality constraint binds, repetition should decrease expected tips because the server bears less risk and therefore requires less expected compensation. Our empirical results provide no evidence supporting the first and third implications. This supports our contention that risk aversion does not play a major role in tipping.

why a person might have these negative feelings. Perhaps she dislikes having someone disapprove of her, even someone with whom she will never interact again. Perhaps she has internalized some standard of fairness that leads her to feel guilty if she does not reward the server for his efforts. However, rather than go into detail about the sources of this internal enforcement, we examine its implications.

We posit a simple theoretical model for how a person decides how much to tip. Our model is based on two hypotheses. First, we posit that a person experiences disutility from not adhering to the norm, and in deciding how much to tip, the customer trades off material payoffs against this disutility. To formalize this hypothesis, we suppose that at the end of a meal the customer chooses her tip to maximize the preferences represented by

$$U^C(T, Q, \theta) = v(Q, \theta) - T - \gamma(\theta) h(T^N(Q, \theta) - T).$$

In this formulation, the first two terms represent the customer’s material payoffs — that is, the payoffs we used in Section 4. The new third term reflects the disutility from not adhering to the norm. $T^N(Q, \theta)$ is the “norm”, by which we mean the appropriate tip as a function of service quality and other factors. The customer suffers disutility whenever she deviates from this norm, where we assume the function h is increasing, convex, and satisfies $h(0) = 0$. Finally, the factor $\gamma(\theta)$ captures the degree that the person cares about deviating from the norm (relative to material payoffs). We assume that the magnitude of this concern depends on features of the environment, as we discuss below.²⁷

The question of what the norm $T^N(Q, \theta)$ is remains. One might argue that the norm is merely 15%, but this is clearly not consistent with our data where well over half of those surveyed left a tip greater than 15% and where we have already seen how a number of factors affect percent tip. Because people often debate at the end of a meal how much the server *deserves*, our second hypothesis is that the underlying norm is based on internalized concerns for fairness and, in particular, that the server deserves to be rewarded for his efforts. To formalize this hypothesis, we assume there is a function $\tilde{T}(e)$ that represents the appropriate tip as a

²⁷ Basu (1997,2001) distinguishes between preference-changing norms and rationality-limiting norms. Our simple formulation falls in the former category, which we prefer because it formalizes the sense that a person has discretion over the degree that she adheres to the norm.

function of the server's effort level. Unfortunately, the customer cannot observe the server's effort level and is therefore forced to *infer* it from perceived service quality Q and other factors θ . Let $E_e(e|Q, \theta)$ denote the server's expected effort level conditional on Q and θ . We then assume the appropriate tip is $T^N(Q, \theta) = \tilde{T}(E_e(e|Q, \theta))$.²⁸

Our point is not the specific functional forms, but rather the general intuitions that (i) the customer is likely to trade off paying a smaller tip for material reasons against feeling guilt and shame from violating the norm, (ii) the norm itself is based on the customer's assessment of how much the server deserves for his effort, and (iii) the degree that the person cares about violating the norm may depend on features of the environment. These intuitions have testable predictions for tipping behavior.

While we do not take our functional forms too seriously, it is instructive to use our formal model to derive comparative statics. Because the customer's preferences are concave in T , the customer's optimal tip T^* satisfies the first-order condition $\chi(\theta)h'(T^N(Q, \theta) - T^*) = 1$. This first-order condition permits us to derive comparative statics over features of the environment. In our analysis that follows, we distinguish between features of the environment (elements of θ) that influence $\chi(\theta)$ and features of the environment (other elements of θ) that influence $T^N(Q, \theta)$. For simplicity, we derive our comparative statics assuming that the customer views the server's effort level as independent of θ .²⁹

5.1 Factors that Influence $\chi(\theta)$

Consider first comparative statics over factors that influence the degree that the customer cares about the norm. It is straightforward to derive that $dT^*/d[\chi(\theta)] > 0$ and that $d[dT^*/dQ]/d[\chi(\theta)] = 0$. These formal results reflect two more general intuitions: If some feature

²⁸ This reduced-form formulation abstracts away from the determinants of the server's effort level e . The server's effort level might be stochastic but exogenous (e.g., there are hardworking servers and lazy servers) or the server might choose e taking into account how the tipper rewards service. We merely assume that however e is determined, the customer forms her expectation $E_e(e|Q, \theta)$ taking its determination into account. However, our formulation assumes that the customer does not care directly about the determination of e ; she cares only insofar as it affects her formulation of $E_e(e|Q, \theta)$.

²⁹ There are two ways that this assumption is restrictive. First, θ might have some direct effect on effort level. Second, in situations where we conclude that θ changes the sensitivity of tips to service quality, the server might react to the changed marginal incentives. While these effects might mitigate or enhance our conclusions below, we suspect that they are second-order relative to the effects we identify.

of the environment causes a customer to care more about adhering to the norm of restaurant tipping, then we should expect to see larger tips, while at the same time we should not expect to see much of an effect on the sensitivity of tips to perceived service quality.

What factors influence $\gamma(\theta)$? Under our conceptualization, the norm of restaurant tipping is enforced by guilt and shame from violating the norm. The degree that the customer experiences such feelings will depend on how much she identifies with the server and values the social approval and esteem of the server. In our dataset, there are a number of factors that we suspect affect $\gamma(\theta)$. Perhaps the most obvious is repetition: The more often the tipper interacts with a particular server, the more the tipper identifies with that server and values the social approval and esteem of that server, therefore $\gamma(\theta)$ is larger. If so, then our theoretical framework predicts that repetition should lead to higher tips, but should not affect the sensitivity of tips to service. This is exactly what we found in our earlier analysis (where the latter result is discussed in footnote 24). Hence, our empirical results on the effects of repetition are consistent with our simple theoretical model.

Two other factors in our dataset that might influence $\gamma(\theta)$ are age and the number of times a person frequents restaurants. We have already seen throughout our regressions that these factors have significant effects on percent tip. A natural — but admittedly post hoc — explanation for these results is that younger people and people who frequent restaurants regularly identify more with servers. Our theoretical framework also suggests that these factors should not affect the sensitivity of tips to service quality; we test these predictions below.

A final factor in our dataset that might influence $\gamma(\theta)$ is gender interactions. Our earlier regressions suggest that basic gender effects are small. Our prior specifications do not consider cross-gender effects, however, and it seems reasonable to believe that males might care more about the social approval of females whereas females might care more about the social approval of males. Moreover, we suspect these cross-gender effects will matter more for younger people. We also test these predictions below.³⁰

³⁰ Alcohol consumption is another variable that might influence $\gamma(\theta)$; however, there are many different ways that alcohol consumption might influence tipping behavior, and hence we do not focus on predictions with regard to alcohol consumption.

5.2 Factors that Influence $T^N(Q, \theta)$

Consider next comparative statics over factors that influence the norm itself. According to our model, the norm is given by $T^N(Q, \theta) = \tilde{T}(E_e(e|Q, \theta))$. For simplicity, we further suppose that \tilde{T} is linear with slope $d\tilde{T}/de = t > 0$.³¹ This means that factors that influence the norm are factors that influence $E_e(e|Q, \theta)$. It is straightforward to derive that $dT^*/d\theta = dT^N/d\theta = t*d[E_e(e|Q, \theta)]/d\theta$ and that $d[dT^*/dQ]/d\theta = d[dT^N/dQ]/d\theta = t*d[dE_e(e|Q, \theta)/dQ]/d\theta$. As before, these formal results reflect two more general intuitions. First, if some feature of the environment causes the tipper to infer a larger effort level for any given level of perceived service quality, then we should expect to see larger tips. Second, if some feature of the environment permits the tipper to extract more information about effort from perceived service quality, then we should expect to see tips become more sensitive to perceived service quality.³²

In our dataset, the most obvious factor that influences $E_e(e|Q, \theta)$ is whether a meal occurs on a weekday or weekend. As in Section 4, we expect the relationship between effort and service quality to be weaker on weekends than it is on weekdays. For many natural formalizations of a weaker relationship, it follows that $dE_e(e|Q, \theta)/dQ$ is smaller on weekends and therefore that tips should be less sensitive to service quality on weekends, just as we found in Section 4. Of course one can construct some formalizations of a weaker relationship where $dE_e(e|Q, \theta)/dQ$ is not decreasing. However, our point is that our simple theoretical model suggests a plausible and intuitive explanation for why tips are less sensitive to service quality on weekends: The customer wants to reward the server for his effort, but on weekends it is more difficult for the customer to extract from service quality exactly how much effort the server is exerting; therefore, tips depend less on service quality on weekends.

Other factors that should influence $E_e(e|Q, \theta)$ are factors that require additional effort on the part of the server to maintain a given level of service quality, therefore increasing $E_e(e|Q, \theta)$

³¹ We adopt this simplification so that our results are not driven by the shape of \tilde{T} .

³² The second effect is similar to attribution theory from the psychology literature. In one particularly relevant study, Seligman et al (1985) examined tipping behavior by pizza-delivery customers (who did not know that they were subjects in an experiment). Each subject was originally told that the order would arrive in about 45 minutes. Ten minutes later, after the cook could assess whether the delivery would be early or late, the subject was called back, told that the pizza would be early or late, and randomly given either an explanation that implicated the driver's behavior or an explanation that implicated exogenous forces. The main result is that, for those customers given explanations that implicated the driver's behavior, delivery time had a significant effect on tip, whereas for the other group of customers, delivery time had no effect on tip.

for a given Q . Two such factors in our dataset are group size and the number of courses ordered. In other words, controlling for service quality and bill size among other things, having a bigger group or having more courses should lead the tipper to infer a larger effort level for any given level of perceived service quality.³³ Hence, our theoretical framework predicts that these factors should lead to higher tips. Indeed, we have seen throughout our regressions that group size has a large positive effect on percent tip. The coefficients on number of courses have been consistently positive but not statistically significant. It also seems plausible that these factors should permit the tipper to extract more information about effort from perceived service quality. If so, then our theoretical framework predicts that these factors should increase the sensitivity of tips to service quality. We test these predictions below.

5.3 Testing the Predictions

While we have already discussed how our earlier empirical results relate to our simple theoretical model, we now test some further predictions suggested by our analysis above. We discuss above our belief that age and the frequency of restaurant visits influence the degree that the tipper cares about adhering to the norm and, therefore, should have no effect on the sensitivity of tips to service. We also discuss above our belief that group size and the number of courses permit the tipper to extract more information about effort from perceived service quality, therefore potentially increasing the sensitivity of tips to service quality. To test these predictions, we interact these variables with service quality and include these interaction terms as independent variables. The results are reported in Column 1 of Table 4. As expected, the interaction terms for age and frequency have small coefficients that are not significant. Also as expected, the interaction term for group size is positive and both economically and statistically significant. Increasing group size by one person increases the effect a one-point increase in service quality has on tip by 0.961 percentage points. The interaction term for number of courses is negative, the opposite of our prediction, but it is not significantly different from zero.

We also discuss above our belief that cross-gender interactions will increase the degree

³³ Consider the effort required to serve a large group rather than a small group. If both groups have the same bill size, the effort required by the server to provide a certain service quality is greater for the large group because, for instance, there are more orders to remember and there are more individuals making requests such as filling up a water glass and providing additional condiments. A similar explanation can be used to justify why greater effort is more likely required if a customer orders multiple courses.

that the tipper cares about adhering to the norm, hence leading to larger tips, and that such effects will matter more for younger people. To test the first hypothesis, we include indicator variables constructed based on tipper's and server's genders. A Male-tips-Female variable equals one if the tipper is male and the server is female. Similarly, a Female-tips-Male variable equals one if the tipper is female and the server is male. Because the specification includes an indicator variable for the gender of the tipper, the coefficient associated with the Male-tips-Female variable represents the differential effect on tip size of a male tipper having a female rather than a male server, and the coefficient associated with the Female-tips-Male variable is the differential effect for a female tipper when the server is male rather than female.³⁴ The coefficient estimates of this specification are presented in the second column of Table 4. While the Male-tips-Female coefficient is very close to zero, the coefficient associated with the Female-tips-Male variable is positive and statistically significant. The coefficient suggests that female customers leave tips that are 1.812 percentage points larger if their servers are male compared to female.

To test whether cross-gender effects matter more for younger people, we interact the Male-tips-Female and Female-tips-Male variables with the age of the tipper. We construct a dummy variable that equals one if the tipper is either a teenager or young adult and zero if the tipper is either middle aged or elderly. The coefficients associated with these interactive terms represent the differential effect on percent tip of having a server of the opposite gender for younger and older tippers. The results from including these interactive terms as independent variables are presented in the last column of Table 4. The coefficient estimates for both interactive terms are positive. The magnitudes of the coefficients indicate that the increase in tip size when the server's gender is different than the tipper's gender is substantially more when the tipper is a teenager or young adult compared to middle aged or elderly. For example, the percent tip when the tipper is female and the server is male is 2.265 percentage points greater when the tipper is a teenager or young adult.³⁵

³⁴ There seems to be no systematic relationship between the genders of the tipper and the server. The probability of a male server conditional on a male tipper is .64 while the probability of a male server conditional on a female tipper is .59. Similarly, the probability of a male tipper conditional on a male server is .75 while the probability of a male tipper conditional on a female server is .72.

³⁵ Our focus in this section has been comparative statics over percent tip and not comparative statics over service quality. In part, we do so because comparative statics over service quality are hard to derive without a better notion of how server effort is chosen. But another reason for doing so is that most of the variables in our dataset seem to

VI. CONCLUSION

Our goal in this paper has been to analyze restaurant tipping as a behavioral norm. Our empirical analysis identifies a variety of factors that influence tipping behavior, and we introduce a simple theoretical framework to help to interpret these results. Our empirical results indicate that percent tip depends not only on service quality but also on a variety of other factors, including repetition, age, group size, the frequency of one's visits to restaurants, and cross-gender interactions. Moreover, we find that the sensitivity of tips to service quality depends on noise — proxied by whether the meal occurs on a weekday or a weekend — and group size. Our theoretical framework provides testable implications for what tipping behavior would look like if the tipping norm serves as a substitute for an efficient tipping contract, from which we conclude that our observed tipping behavior does not appear to be fully efficient. We also posit a simple theoretical model for what the norm might be and how the norm might influence behavior, and we show that our observed tipping behavior is consistent with this model.

Our empirical results and accompanying interpretation suggest that the norm of tipping is consistent with the underlying psychology that people apply throughout their social interactions. Perhaps this explains why tipping is a billion-dollar-per-year phenomenon and why tipping has flourished in so many countries. However, it is also interesting to note that the restaurant-tipping norm varies significantly across countries. For example, in Australia, China, Denmark, Japan, and Iceland, restaurant tipping is not used; and in Belgium, France, Italy, the Netherlands, Norway, and Sweden, it is customary merely to round up the bill or to leave small change in addition to an automatic service charge (Starr, 1988). The question arises of why there exist significant cross-country differences. Our analysis suggests two possible answers. First, to the extent that the tipping norm arises for efficiency reasons, the potential efficiency benefits from a tipping norm may differ across countries. For instance, in certain cultures it could be that customers care less about service, that restaurant owners are better able to monitor a server's effort level, or that servers are more prone to exert high effort even in the absence of incentives, resulting in less need for a tipping norm. Second, to the extent that the tipping norm is enforced

have very little effect on service quality (the main exception being repetition, as discussed in Section 4).

by internalized feelings of guilt and shame, perhaps the degree to which people experience such feelings may differ across countries.³⁶

Norms clearly influence human behavior, and a growing literature in economics studies the role of norms. However, unlike much of this previous research that focuses on “social norms” that are enforced through direct social sanctions, our analysis considers a norm that is (at least primarily) enforced by internalized feelings of guilt and shame. We suspect that such internal enforcement is at work for many other norms as well. For instance, people tend to help people in distress and not cheat on their spouses, even in situations where no one will observe their behavior. Decedents usually divide their estates according to the prevailing societal norms (e.g., equally among the children in the United States (Wilhelm 1996) vs. to a single child in many other societies (Chu 1991)) despite the fact that no one (except their lawyers) need know their division until after their deaths. People obey the norm of voting — “doing their civic duty” — even in situations where they know their vote will not matter and where no one will know whether they voted. We hope that, in addition to helping economists better understand restaurant-tipping behavior, our empirical and theoretical analysis will help create a framework for how to analyze these and other internalized norms.

³⁶ An interesting question is how service quality across countries depends on the restaurant-tipping norms.

References

- Akerlof, G., 1980. A theory of social custom, of which unemployment may be one consequence. *Quarterly Journal of Economics* 94, 749-775.
- Akerlof, G., 1982. Labor contracts as partial gift exchange. *Quarterly Journal of Economics* 97, 543-569.
- Arrow, K., 1971. Political and economic evaluation of social effects and externalities. In: M. Intriligator (Ed.). *Frontiers of Quantitative Economics*. North-Holland, Amsterdam, pp. 3-25.
- Axelrod, R., 1986. An evolutionary approach to norms. *American Political Science Review* 80, 1095-1111.
- Basu, K., 1997. The role of norms and the law. Department of Economics Working Paper No. 451, Cornell University.
- Basu, K., 2001. The role of norms and law in economics: an essay in political economy. In: J.W. Scott and D. Keates (Eds.). *Schools of Thought: Twenty-Five Years of Interpretive Social Science*. Princeton University Press, pp. 154-178.
- Ben-Zion, U. and Karni, E., 1977. 'Tip' payments and the quality of service. In: O. Ashenfelter and W. Oates (Eds.). *Essays in Labor Market Analysis*. John Wiley & Sons, New York, pp. 37-44.
- Berg, J., Dickhaut, J., and McCabe, K., 1995. Trust, reciprocity, and social history. *Games and Economic Behavior* 10, 122-142.
- Bernheim, B., 1994. A theory of conformity. *Journal of Political Economy* 102, 841-877.
- Besley, T. and Coate, S., 1992. Understanding welfare stigma: taxpayer resentment and statistical discrimination. *Journal of Public Economics* 48, 165-183.
- Bodvarsson, O. and Gibson, W., 1997. Economics and restaurant gratuities: determining tip rates. *American Journal of Economics and Sociology* 56, 187-203.
- Bolton, G. and Ockenfels, A., 2000. ERC: a theory of equity, reciprocity, and competition. *American Economic Review* 90, 166-193.
- Charness, G. and Rabin, M., 2002. Social preferences: some simple tests and a new model. *Quarterly Journal of Economics* 117, 817-869.
- Chu, C., 1991. Primogeniture. *Journal of Political Economy* 99, 78-99.
- Dixit, A., 1996. *The Making of Economic Policy: A Transaction-Cost Politics Perspective*. MIT

Press, Cambridge.

Elster, J., 1989a. Social norms and economic theory. *Journal of Economic Perspectives* 3(4), 99-117.

Elster, J., 1989b. *The Cement of Society: A study of social order*. Cambridge University Press.

Fehr, E. and Gächter, S., 2000. Fairness and retaliation: the economics of reciprocity. *Journal of Economic Perspectives* 14(3), 159-181.

Fehr, E., Kirchsteiger, G., and Riedl, A., 1993. Does fairness prevent market clearing? An experimental investigation. *Quarterly Journal of Economics* 108, 437-459.

Fehr, E. and Schmidt, K., 1999. A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics* 114, 817-868.

Fudenberg, D., Holmstrom, B., and Milgrom, P., 1990. Short-term contracts and long-term agency relationships. *Journal of Economic Theory* 51, 1-31.

Hemenway, D., 1984. *Prices and Choices: Microeconomic Vignettes*. Ballinger, Cambridge.

Holmstrom, B. and Milgrom, P., 1991. Multitask principal-agent analyses: incentive contracts, asset ownership, and job design. *Journal of Law, Economics, and Organization* 7, 24-52.

Homans, G., 1953. Status among clerical workers. *Human Organization* 12, 5-10.

Homans, G., 1954. The cash posters. *American Sociological Review* 19, 724-733.

Jacob, N. and Page, A., 1980. Production, information costs, and economic organization: the buyer monitoring case. *American Economic Review* 70, 476-478.

Kahneman, D., Knetsch, J., and Thaler, R., 1986. Fairness as a constraint on profit seeking: entitlements in the market. *American Economic Review* 76, 728-741.

Lindbeck, A., Nyberg, S., and Weibull, J., 1999. Social norms and economic incentives in welfare state. *Quarterly Journal of Economics* 114, 1-35.

Lynn, M. and McCall, M., 2000. Gratitude and gratuity: a meta-analysis of research on the service-tipping relationship. *Journal of Socio-Economics* 29, 203-214.

Lynn, M., Zinkhan, G., and Harris, J., 1993. Consumer tipping: a cross-country study. *Journal of Consumer Research* 20, 478-485.

Mills, S. and Riehle, H., 1987. What customers think about tips vs. service charges? *Restaurants USA*, (Oct.), 20-22.

- Posner, R. and Rasmusen, E., 1999. Creating and enforcing norms, with special reference to sanctions. *International Review of Law and Economics* 19, 369-382.
- Rabin, M., 1993. Incorporating fairness into game theory and economics. *American Economic Review* 83, 1281-1302.
- Rabin, M., 2000. Risk aversion and expected-utility theory: a calibration theorem. *Econometrica* 68, 1281-1292.
- Rogerson, W., 1985. Repeated moral hazard. *Econometrica* 53, 69-76.
- Seligman, C., Finegan, J., Hazelwood, J., and Wilkinson, M., 1985. Manipulating attributions for profit: a field test of the effects of attributions on behavior. *Social Cognition* 3, 313-321.
- Starr, N., 1988. *The International Guide to Tipping*. Berkeley, New York.
- Sudgen, R., 1989. Spontaneous order. *Journal of Economic Perspectives* 3(4), 85-97.
- Wilhelm, M., 1996. Primogeniture. *American Economic Review* 86, 78-99.

TABLE 1
Summary Statistics
(Based on 1393 Observations)

| Variables | Mean | Standard Deviation |
|--|--------|--------------------|
| Size of Tip (dollars) | 4.30 | 3.12 |
| Size of Bill (dollars) | 26.42 | 18.24 |
| Percent Tip (Tip/Bill) | 17.56% | 11.26 |
| Group Size | 2.37 | 0.91 |
| Number of Courses in Meal | 1.96 | 0.97 |
| Alcohol Consumed (=1 if yes, 0 otherwise) | 0.53 | 0.50 |
| FOOD MEASURES: | | |
| Appearance of Food | 4.35 | 0.71 |
| Size of Food Portion | 4.55 | 0.67 |
| Taste of Food | 4.45 | 0.73 |
| Temperature of Food | 4.39 | 0.81 |
| Price of Food | 4.31 | 0.78 |
| SERVER MEASURES: | | |
| Appearance of Server | 4.40 | 0.77 |
| Knowledge of Server | 4.46 | 0.81 |
| Friendliness of Server | 4.61 | 0.71 |
| Speed of Service | 4.36 | 0.91 |
| Attentiveness of Server | 4.33 | 0.94 |
| Gender of Server (=1 if female, =0 if male) | 0.38 | 0.48 |
| Gender of Tipper (=1 if female, =0 if male) | 0.26 | 0.44 |
| Age of Tipper | 2.68 | 0.67 |
| Times Tipper Frequents Particular Restaurant (monthly) | 2.09 | 3.19 |
| Times Tipper Frequents Any Restaurant (monthly) | 11.55 | 9.57 |
| Interview Occurred on Weekend | 0.08 | 0.27 |
| Number of Seats in Restaurant | 212 | 121 |

TABLE 2

| Independent Variable | Dependent Variable | Percent Tip | Percent Tip | Service Quality |
|---|--------------------|---------------------|---------------------|--------------------|
| Size of Bill (dollars) | | -0.543** (0.156) | -0.548** (0.154) | 0.001 (0.003) |
| Size of Bill Squared | | 0.003** (0.001) | 0.003** (0.001) | 0.0002 (0.0002) |
| Group Size | | 1.751* (0.992) | 1.763* (0.967) | -0.050* (0.028) |
| Number of Courses in Meal | | 0.241 (0.543) | 0.256 (0.545) | -0.015 (0.020) |
| Alcohol Consumed | | 2.077** (0.776) | 2.106** (0.750) | 0.001 (0.042) |
| FOOD ATTRIBUTES: | | | | |
| Appearance of Food | | 0.346 (0.410) | | |
| Size of Food Portion | | 0.315 (0.596) | | |
| Taste of Food | | -0.349 (0.709) | | |
| Temperature of Food | | 0.413 (0.569) | | |
| Price of Food | | 0.747 (0.455) | | |
| Food Measure | | | 1.481** (0.599) | 0.529** (0.043) |
| Gender of Server | | -0.615 (0.645) | -0.583 (0.656) | 0.025 (0.041) |
| Gender of Tipper | | 0.215 (0.698) | 0.209 (0.694) | 0.034 (0.038) |
| Age of Tipper | | -0.645* (0.365) | -0.620* (0.354) | 0.027 (0.025) |
| Times Tipper Frequents Any Restaurant (monthly) | | 0.078* (0.041) | 0.079* (0.041) | -0.0004 (0.002) |
| Indicator Variables for Interviewers | | YES | YES | YES |
| R-squared | | 0.1694 | 0.1680 | 0.3440 |
| Observations | | 1393 | 1393 | 1393 |

Notes: Standard error is in parentheses, (*) represents statistically significant at ten percent level, and (**) represents statistically significant at five percent level. The standard errors are robust to arbitrary heteroskedasticity and correlation within interviewer.

TABLE 3

| Dependent Variable | PERCENT TIP | | | | | SERVICE QUALITY |
|---|---------------------|---------------------|---------------------|---------------------|---------------------|--------------------|
| Independent Variable | (1) | (2) | (3) | (4) | (5) | (6) |
| Size of Bill (dollars) | -0.549** (0.156) | -0.545** (0.156) | -0.510** (0.184) | -0.550** (0.156) | -0.547** (0.153) | 0.001 (0.003) |
| Size of Bill Squared | 0.003** (0.001) | 0.003** (0.001) | 0.003** (0.001) | 0.003** (0.001) | 0.003** (0.001) | 0.0001 (0.0002) |
| Group Size | 1.852* (1.010) | 1.827* (1.005) | 1.758 (1.188) | 1.857* (1.010) | 1.885* (0.985) | -0.047* (0.028) |
| Number of Courses in Meal | 0.294 (0.549) | 0.265 (0.543) | 0.025 (0.799) | 0.291 (0.549) | 0.271 (0.546) | -0.016 (0.021) |
| Alcohol Consumed | 2.066** (0.784) | 2.080** (0.786) | 2.496** (1.184) | 2.053** (0.786) | 2.060** (0.759) | -0.002 (0.042) |
| FOOD ATTRIBUTES: | | | | | | |
| Appearance of Food | 0.249 (0.455) | 0.185 (0.454) | 0.178 (0.588) | 0.261 (0.454) | | |
| Size of Food Portion | 0.140 (0.567) | 0.198 (0.600) | 0.821 (0.854) | 0.138 (0.569) | | |
| Taste of Food | -0.435 (0.672) | -0.478 (0.707) | -0.477 (0.924) | -0.443 (0.672) | | |
| Temperature of Food | 0.284 (0.521) | 0.258 (0.515) | 0.296 (0.457) | 0.270 (0.519) | | |
| Price of Food | 0.628 (0.435) | 0.542 (0.461) | -0.044 (0.557) | 0.620 (0.434) | | |
| Food Quality | | | | | 0.585 (0.660) | 0.520* (0.043) |
| SERVER ATTRIBUTES: | | | | | | |
| Appearance of Server | 0.064 (0.469) | | 0.070 (0.785) | 0.068 (0.464) | | |
| Knowledge of Server | -0.781* (0.462) | | -1.301* (0.759) | -0.787* (0.462) | | |
| Friendliness of Server | 1.336* (0.716) | | 1.925* (0.994) | 1.321* (0.710) | | |
| Speed of Service | 1.196** (0.462) | | 1.223* (0.708) | 1.319** (0.486) | | |
| Attentiveness of Server | -0.475 (0.687) | | -0.758 (0.973) | -0.480 (0.686) | | |
| Service Quality | | 1.464** (0.551) | | | 1.430** (0.537) | |
| Gender of Server | -0.776 (0.638) | -0.651 (0.654) | -1.114 (0.987) | -0.787 (0.638) | -0.510 (0.677) | 0.032 (0.041) |
| Gender of Tipper | 0.186 (0.711) | 0.164 (0.701) | -0.012 (1.061) | 0.175 (0.711) | 0.191 (0.690) | 0.036 (0.038) |
| Age of Tipper | -0.640* (0.357) | -0.683* (0.362) | -0.866* (0.449) | -0.636* (0.357) | -0.678* (0.345) | 0.025 (0.025) |
| Times Tipper Frequents this Particular Restaurant (monthly) | | | | | 0.187** (0.084) | 0.011** (0.005) |
| Times Tipper Frequents Any Restaurant (monthly) | 0.077** (0.039) | 0.078** (0.040) | 0.064 (0.039) | 0.075* (0.039) | 0.070* (0.038) | -0.001 (0.002) |
| (Weekend Day) * (Speed of Service) | | | | -1.020* (0.536) | | |
| Indicator Variables for Interviewers | YES | YES | YES | YES | YES | YES |
| R-squared | 0.1798 | 0.1746 | 0.2001 | 0.1803 | 0.1757 | 0.3462 |
| Observations | 1393 | 1393 | 874 | 1393 | 1393 | 1393 |

Notes: Standard error is in parentheses, (*) represents statistically significant at 10 percent level, and (**) represents statistically significant at 5 percent level. The standard errors are robust to arbitrary heteroskedasticity and correlation within interviewer.

TABLE 4

| Independent Variables | Dependent Variable | PERCENT TIP | | |
|---|--------------------|---------------------|---------------------|---------------------|
| | | (1) | (2) | (3) |
| Size of Bill (dollars) | | -0.541** (0.151) | -0.543** (0.152) | -0.538** (0.152) |
| Size of Bill Squared | | 0.003** (0.001) | 0.003** (0.001) | 0.003** (0.001) |
| Group Size | | -2.344 (1.666) | 1.852* (0.983) | 1.826* (0.993) |
| Number of Courses in Meal | | 2.465 (2.651) | 0.255 (0.544) | 0.254 (0.545) |
| Alcohol Consumed | | 2.078** (0.754) | 2.052** (0.755) | 2.053** (0.757) |
| Food Quality | | 0.672 (0.652) | 0.602 (0.661) | 0.598 (0.643) |
| Service Quality | | 0.496 (1.711) | 1.410** (0.542) | 1.395** (0.540) |
| Gender of Server | | -0.477 (0.661) | | |
| Male tips Female | | | 0.005 (0.776) | -0.458 (0.860) |
| Female tips Male | | | 1.812* (1.061) | 0.697 (0.795) |
| Gender of Tipper | | 0.223 (0.675) | -0.888 (0.798) | -0.814 (0.777) |
| Age of Tipper | | 0.030 (2.568) | -0.655* (0.343) | |
| Tipper is Teenager or Young Adult | | | | 0.161 (0.644) |
| Times Tipper Frequents this Particular Restaurant (monthly) | | 0.198** (0.088) | 0.188** (0.083) | 0.186** (0.084) |
| Times Tipper Frequents Any Restaurant (monthly) | | 0.092 (0.155) | 0.069* (0.038) | 0.069* (0.036) |
| (Service Measure)*(Group Size) | | 0.961** (0.489) | | |
| (Service Measure)*(Number of Courses) | | -0.498 (0.610) | | |
| (Service Measure)*(Age of Tipper) | | -0.157 (0.578) | | |
| (Service Measure)*(Times Tipper Frequents Any Restaurant) | | -0.006 (0.034) | | |
| (Male tips Female) * (Tipper is Teenager or Young Adult) | | | | 1.206 (1.529) |
| (Female tips Male) * (Tipper is Teenager or Young Adult) | | | | 2.265 (1.591) |
| Indicator Variables for Interviewers | | YES | YES | YES |
| R-squared | | 0.1786 | 0.1767 | 0.1779 |
| Observations | | 1393 | 1393 | 1393 |

Notes: Standard error is in parentheses, (*) represents statistically significant at 10 percent level, and (**) represents statistically significant at 5 percent level. The standard errors are robust to arbitrary heteroskedasticity and correlation within interviewer.

APPENDIX

The survey was conducted in the Fall of 1990 and the Spring of 1991. One hundred and twelve students in two undergraduate consumer behavior courses conducted 112 survey sessions outside 39 restaurants in Houston, Texas. The restaurants consisted of Rosa Mexicano, Black Labrador, Ritas (two locations), Cyclone Anayas (two locations), India's, Rusty Pelican, Steak & Ale, Pappasito's (two locations), Black Eyed Pea, Thai Café, Golden Room, Churrasco's, On The Border, 59 Diner, Mingerellis, Doneraki, Saigon Pagolac, Adrian's, Atchafalaya, El Torito, Pizzeria Uno, Baba Yega, Barron's, Boca Del Rio, Bombay Grill, Casa Ole, Fandangos, Fornos of Italy, Kim Son, Landrys, La Strada, Little Papisittos, Los Tios, Mamasita's, Mason Jar, and Taste of Texas. Each survey session was conducted between 6pm and 10pm at a single restaurant. The interviewer approached all individuals leaving the restaurant (except those who left while the student was busy interviewing another individual) and asked if they would participate in a survey. The survey questions and the approach script are provided below.³⁷

The students were not told the purpose of the survey. When returning the completed survey sheets, each student was questioned to ascertain the extent to which he/she followed the interviewing procedure. In addition, an employee at each restaurant was asked to verify that the student did in fact conduct the interviews. An interview was not included in the analysis contained in this paper if (i) the restaurant's staff could not verify that the student conducted the interviews, (ii) the student significantly departed from the prescribed procedure, (iii) the respondent's dining party did not order food, (iv) the respondent's dining party had more than 5 members, and (v) if the respondent's dining party had more than one person paying the bill.

³⁷ The survey on the subsequent page was used for the Spring of 1991. The survey for the Fall of 1990 was similar except for minor wording differences, a reversal of the order of the questions about tip size and bill size, and a question about ethnicity.

Interviewer: _____
Restaurant: _____

Date: _____
Refusals: _____

APPROACH SCRIPT:

“EXCUSE ME, MY NAME IS _____. I’M A STUDENT AT THE UNIVERSITY OF HOUSTON AND I’M DOING A SURVEY OF RESTAURANT CUSTOMERS FOR A MARKETING CLASS. WOULD YOU BE WILLING TO ANSWER SEVERAL QUESTIONS FOR ME? THANK YOU FOR AGREEING TO HELP ME. IF YOU DON’T MIND, I WOULD LIKE ONLY THE PERSON PAYING THE BILL TO ANSWER THESE QUESTIONS.”

(If there were separate checks or if more than one person contributed to a single check, then interview each payee separately and indicate that these interviews came from the same table.)

INTERVIEW QUESTIONS:

1. How many people were:
at your table? _____
On your check (or portion of the check)? _____

2. How many checks did your table have? _____

3. Did anyone at your table have:
appetizers? Y / N
soups? Y / N
salads? Y / N
entrees? Y / N
desserts? Y / N
alcohol? Y / N

4. Given a 5 point scale with 1 being poor and 5 being excellent, how would you rate the food on:
appearance 1 2 3 4 5
portion size 1 2 3 4 5
taste 1 2 3 4 5
temperature 1 2 3 4 5
price 1 2 3 4 5

5. Given a 5 point scale with 1 being poor and 5 being excellent, how would you rate the server on:
appearance 1 2 3 4 5
knowledge of menu 1 2 3 4 5
friendliness 1 2 3 4 5
speed of service 1 2 3 4 5
attentiveness 1 2 3 4 5

6. Was your server a male or a female? M / F

7. How often do you eat dinner at:
this particular restaurant? _____
all other full-service restaurants? _____

8. How much money, in dollars and cents, did you tip the server? _____

9. How much was your bill not including tip? _____

“THANK YOU VERY MUCH FOR YOUR TIME AND COOPERATION.”

Payee’s Gender: M / F
Payee’s Age: Teenager / Young Adult / Middle Aged / Elderly