Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

The Effects of Tip Recommendations on Customer Tipping, Satisfaction, Repatronage, and Spending

Damon Alexander, Christopher Boone, Michael Lynn

http://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Articles in Advance, pp. 1–20
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

The Effects of Tip Recommendations on Customer Tipping, Satisfaction, Repatronage, and Spending

Damon Alexander,[a] Christopher Boone,[b] Michael Lynn[b]

[a] Washio, Santa Monica, California 90402; [b] School of Hotel Administration, Cornell University, Ithaca, New York 14853

Contact: d.alexander6@gmail.com (DA); boone@cornell.edu (CB); wml3@cornell.edu, https://orcid.org/0000-0001-9234-1248 (ML)

Received: December 15, 2017
Revised: March 20, 2019; October 24, 2019
Accepted: October 28, 2019
Published Online in Articles in Advance: June 1, 2020

https://doi.org/10.1287/mnsc.2019.3541

Abstract. A field experiment involving 94,571 orders from 24,637 customers of an app-based laundry pick-up, cleaning, and delivery service examined the effects of various randomly assigned tip recommendations on consumers' tip amounts, satisfaction ratings, frequency of return, and bill size. We find that tip recommendations affect tip amounts, but not customer satisfaction, patronage frequency, or bill size, which implies that neither the processes underlying the tip-recommendation effects on tipping nor consumer tipping itself affect these other consumer outcomes. From a practical perspective, these results and conclusions inform efforts to increase or decrease tipping. Recommending larger tip amounts, at least within the \$2–\$10 or 5%–25% ranges studied here, appears to be a safe means of increasing the amounts customers leave. More generally, altering customers' tipping behavior will not itself adversely affect those customers' subsequent satisfaction, repatronage, or spending, as long as the means used to alter tipping do not directly affect these other outcomes.

History: Accepted by John List, behavioral economics.

Supplemental Material: Data are available at https://doi.org/10.1287/mnsc.2019.3541.

Keywords: tipping • customer satisfaction • customer retention • customer spending • tip recommendations

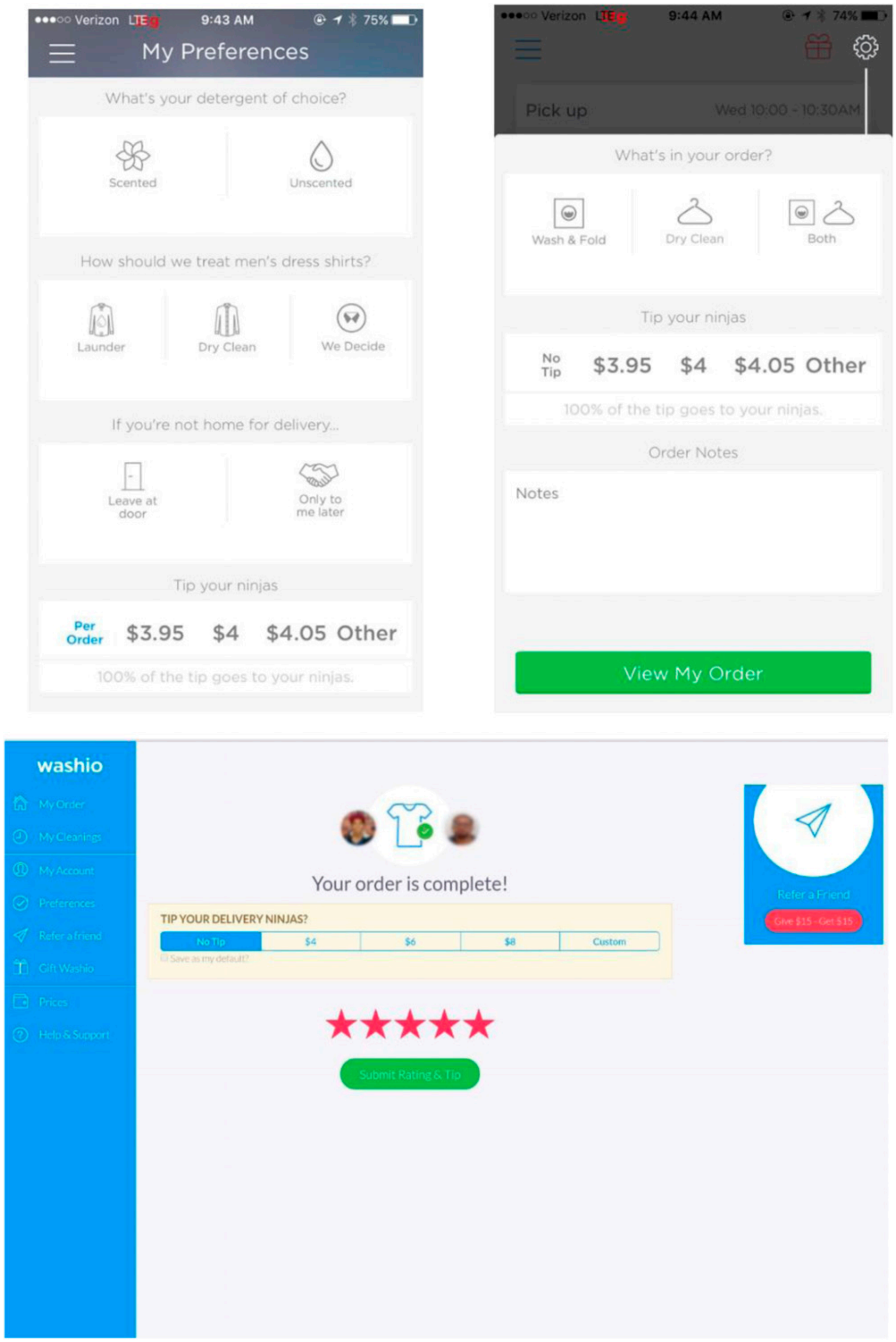
1. Introduction

Every day, millions of consumers around the world give voluntary payments of money (called "tips," "propinas," and "trinkgelds," among other names) to service providers who have already served them. Among those service workers commonly receiving tips are baristas, bartenders, casino dealers, concierges, deliverymen, doormen, hair cutters, hotel maids, masseurs, porters, sommeliers, street musicians, taxi drivers, tour guides, and waiters (Lynn 2016). Precise measures of the total amount tipped worldwide do not exist, but estimates place the amount tipped to restaurant workers in the United States alone at over \$45 billion a year, so the former number must be enormous (Azar 2011). To service workers, these tips represent a source of income, job motivation, and subordination to consumers (Shamir 1983, Namasivayam and Upneja 2007, Shy 2015). To consumers, these voluntary gifts represent an avoidable cost, a normative obligation, an expression of satisfaction, an incentive for future service, and an attempt to help service workers (Becker et al. 2012, Lynn 2015b). To service firms permitting tipping of their employees, these consumer payments represent a form of buyer monitoring, employee compensation, pay-what-you-want pricing, price discrimination, and price partitioning (Schwartz 1997, Lynn and Withiam 2008, Azar 2011).

The complexity and theoretical richness of this behavior has attracted the interest of scholars in anthropology, economics, hospitality management, human resources management, marketing, social psychology, sociology, and tourism management (see Lynn 2006, Azar 2007, and Lynn 2015b for reviews). Research on this topic has examined both ways to increase the tips consumers leave (Seiter et al. 2011, Lynn, 2018) and the organizational consequences of permitting and encouraging tipping (Azar 2011, Lynn 2017). We contribute to both streams of research by investigating the effects of tip recommendations on customers' tipping, satisfaction, repatronage, and spending.

Our data come from a now-defunct app-based laundry pick-up, cleaning, and delivery service. The app that customers used to place and pay for orders also allowed them to add a tip for the delivery drivers to their bills. Specifically, the app presented customers with three suggested tip amounts (each of which could be selected with one click), an option to leave a custom

Figure 1. (Color online) Examples of How the Tip Recommendations/Options Were Presented to Customers



Notes. The image on the top left is a screenshot of the app when saving preferences via a mobile device. The image on the top right is a screenshot of the app when placing an order via mobile device. The bottom image is a screenshot of a completed order as presented on a computer.

tip amount, and a one-click option not to tip (see Figure 1). The company presented different customers with different, randomly selected sets of suggested tip amounts, but presented the same customer with the same suggestions each time he or she used the app. The different sets of suggested tip amounts presented to different customers allowed us to examine the effects of recommending (i) large versus small tip amounts, (ii) percentage versus dollar tip amounts, (iii) a broad versus narrow range of tip amounts (holding the average suggested tip constant), and (iv) round versus nonround dollar tip amounts ($x.00 versus $x.99).

We found that, in the context of providing several different tip recommendations, the provision of larger tip recommendations increased the size of tips that were left as well as overall tip revenues—even while decreasing the likelihood of tipping when the recommendations exceeded familiar and normative amounts. Larger tip recommendations had no corresponding effects on customer satisfaction, repatronage frequency, or bill sizes. These findings suggest that larger tip recommendations do not evoke a negative affective reaction that jeopardizes future patronage—even from those people who are dissuaded from tipping by the larger tip recommendations. Thus, firms can increase employee income by recommending larger tips (within the range studied here) without fear of negative customer reactions. The findings also suggest that tipping behavior itself has little positive or negative effect on the tippers' subsequent attitudes or behaviors, so service firms can set tipping policies with little regard for such effects. Additional findings indicated that (i) recommending equivalent tips in percentage instead of dollar formats increased tip amounts when some of the recommended percentages were subnormative; (ii) increasing the range of suggested tips increased the likelihood of choosing one of the recommended amounts, but did not affect the amount tipped on average; and (iii) recommending equivalent round and nonround tip amounts (differing in size by only a penny, but with different left-most as well as right-most digits) did not produce different levels of tipping.

2. Related Literature

Our study contributes to knowledge about ways to increase the tips consumers leave as well as about the organizational consequences of permitting and encouraging tipping. We also draw from, and build on, a related literature concerned with the effects of different appeals for charitable donations.

2.1. Effects of Increasing Recommended Tip Sizes

The tip recommendations in this study included several sets of recommendations that differed in magnitude of the recommended amounts, but were otherwise similar to one another. Our examination of the effects of these differences in recommended tip size most directly builds on Haggag and Paci's (2014) study of default tips in taxicabs. Using a large database of New York City (NYC) taxicab rides, they tested (i) the effects of one firm's use of different default tip amounts and formats ($ versus %) for fares above and below $15 and (ii) the effects of differences between two firms' percentage default tip amounts on fares above $15 to assess the impact of different default sizes. Across both specifications, they found that larger default tip options were associated with a lower likelihood of tipping, but with larger tip amounts from those who did tip. The empirical setting presents some challenges: Namely, the impact of default tip size is confounded with tip format ($ versus %) in the first specification, and the effects in the second specification may be confounded by other differences between the companies. However, two recent working papers conceptually replicate these findings—one using a difference-in-difference analysis of changes in one of two NYC taxicab companies' default tip options (Hoover 2019) and the second using a natural field experiment manipulating default tipping options on the Uber app (Chandar et al. 2019)[1]. The consistency of directional effects across different studies provides compelling evidence that increasing default tipping options reduces the likelihood of tipping, increases the size of nonzero tips, and increases drivers' tip incomes.[2] Such effects are plausible because suggested or default options convey information about the expected or typical contribution, which is likely to guide the behavior of those willing and able to conform with the implied norm, but may offend or discourage those unwilling or unable to do so (De Bruyn and Prokopec 2013, Haggag and Paci 2014). Our study conceptually replicates these findings using a natural field experiment with naïve subjects randomly assigned to unconfounded tip-recommendation conditions in a different service context, so it provides strong causal inferences about the effects of recommending larger tips and assesses the generalizability of the effects beyond personal transportation services (Al-Ubaydli and List 2013).

Moreover, our data set follows individual customers over time and includes their satisfaction ratings, which permits us to test the effects of the tip recommendations on several new outcomes of interest—namely, customer satisfaction, repatronage frequency, and subsequent spending. Recommending larger tips might affect these other outcomes either directly or indirectly through their effects on tipping. The direct effects (if any) are likely to be negative because large tip recommendations are more likely than small ones to be considered unreasonably high or construed as threats to freedom of choice, either of which could generate negative affect (Clee and Wicklund 1980, Fitzsimons and Lehmann 2004, Carr 2007). As noted by Haggag and Paci (2014), such negative cognitive and affective reactions might also explain why larger asks decrease the likelihood of tipping.

The indirect effects of larger tip recommendations on customer satisfaction, repatronage, and spending through their effects on tipping are potentially more complex and could be negative, positive, or both. First, the larger tips evoked by larger tip recommendations may be viewed or felt by consumers as cost increases, which could lower customers' perceptions of value, satisfaction, repatronage frequency, and subsequent spending consistent with the law of demand (Varki and Colgate 2001).

Second, tip recommendations that evoke larger tips could increase customers' satisfaction, repatronage, and spending because tipping activates self-perception, self-justification, or warm-glow processes. Tipping is a voluntary payment for services rendered as well as a prosocial behavior. Social psychological theory and research suggest that voluntary behavior often affects attitudes either because people infer their attitudes from their behavior (Fazio 1987, Robins and John 1997) or change their attitudes to justify, or be consistent with, their behavior (Aronson 1969, Gawronski 2012). Furthermore, marketing researchers have found that inducing customers to engage in prosocial behavior evokes positive feelings (called a "warm glow") that enhance customers' evaluations of service and retail experiences, as well as their intentions to repatronize the firms that provide those experiences (Giebelhausen et al. 2016, Giebelhausen and Chun 2017, Giebelhausen et al. 2017). Thus, self-perception, self-justification, and warm-glow processes may lead consumers to like, value, and use a service more the more they tip for it.

Third, the opposing effects of larger tip recommendations on tipping likelihood and tip size could decrease nontippers' satisfaction, repatronage, and spending, while simultaneously increasing tippers' satisfaction, repatronage, and spending. The positive effect on tippers' satisfaction, repatronage, and spending of increasing the amounts they tip has already been discussed. Those same self-perception/justification and warm-glow processes could decrease nontippers' satisfaction, repatronage, and spending. The opposing nature of these effects on tippers and nontippers makes it hard to predict the direction of the average effect on the sample as a whole. However, if larger tip recommendations do have opposing effects on the satisfaction, repatronage, and spending of tippers and nontippers, then they should increase the variance in these other outcome variables. We test these competing expectations about the effects of larger tip recommendations, and of tipping itself, on customer satisfaction, repatronage, and spending for the first time.

2.2. Effects of Recommending Percentage vs. Dollar Tip Amounts

The tip recommendations in this study included a set of small percentage tip amounts (5%, 10%, and 15%) and a set of small dollar tip amounts ($2, $4, and $6) that were arithmetically identical when the bill size was $40. It also included a set of larger percentage tip amounts (10%, 15%, and 20%) and a set of larger dollar tip amounts ($4, $6, and $8) that were arithmetically identical when the bill size was $40. Because the vast majority of bill sizes in our data set exceeded $40, the percentage recommendations were generally larger than the dollar recommendations and, given the effects of request magnitude discussed previously, should lead to larger average tips but a lower likelihood of tipping than do the dollar recommendations. However, when bill size was less than $40, the dollar recommendations were larger than the percentage ones, so their effects on tipping should be reversed. More uncertain, but potentially interesting, is what happens when bill size is around $40 and the request magnitudes of these two types of recommendations are comparable. We tested all of these effects below.

2.3. Effects of Recommending a Larger Range of Tips

Among the tip recommendations in our study are two sets of dollar tip recommendations ($2, $4, and $6; and $3, $4, and $5) and two sets of percentage tip recommendations (10%, 15%, and 20%; and 12%, 15%, and 18%) that had the same means but different ranges of recommended amounts. Ours is the first study that we know of to manipulate the range of several recommended tips or donations independently of the mean of those recommendations, so there are no existing findings on which to base expectations about the effects of this manipulation. However, De Bruyn and Prokopec (2013) found that when charitable appeals recommended several donation amounts, the size of the lowest recommendation had a larger effect on donations than did the sizes of the other recommendations. Furthermore, other researchers have found that legitimizing small contributions to charity increases the likelihood of giving without reducing the average gift size (Cialdini and Schroeder 1976, Weyant and Smith 1987) and that increasing the size of requested donations increases the amounts given (Doob and McLaughlin 1989). These findings suggest that increasing the range of recommended tip amounts may increase the proportion of people leaving tips by legitimating smaller tip amounts as well as increase the size of those tips left by simultaneously asking for more. We empirically test this possibility for the first time.

2.4. Effects of Recommending Round vs. Nonround Tip Amounts

The tip recommendations in this study also include one set of round-dollar recommendations ($4, $6, and $8) and another set of nonround recommendations that were a penny less than the round ones ($3.99, $5.99, and $7.99). Round numbers are easier to process than nonround ones (Estelami 1999), and round versus nonround pricing has been shown to have numerous effects on attitudes, beliefs, and behavior. For example, consumers perceive round prices as more convenient than nonround ones (Wieseke et al. 2016), like round prices more than nonround ones (Guido and Peluso 2004), choose round prices at above-chance levels in pay-what-you-want situations (Lynn et al. 2013), and are more likely to buy products priced with round numbers in field studies (e.g., Bray and Harris 2006 and Wieseke et al. 2016). Thus, it is possible that recommending round tip amounts will increase the attractiveness of the recommended tips and, therefore, the likelihood that people will select one of them and leave a tip.

In the closest test of this possibility of which we are aware, Edwards and List (2014) found no support for it; they found that suggesting a charitable donation of $20 (versus a donation of $20.01, $20.02, . . ., $20.08, or $20.09) had no reliable effects on the proportion of subjects donating or on the average size of donations made in their study. However, their nonround suggestions were all larger than their round suggestion, so subjects (who tend to read left to right and to ignore the rightmost digits of prices) may have perceived the nonround suggestions of $20.01 to $20.09 no differently than the round suggestion of $20 (see Thomas and Morwitz 2005). Nonround suggestions that have a different left-most digit (e.g., $2.99 versus $3.00) may be more perceptually distinctive from round suggestions and, therefore, may produce larger round versus nonround suggestion effects on voluntary payments. Our data permitted us to assess this possibility and test the generalizability of Edwards and List's (2014) findings.

3. Setting, Study Design, and Data Description

Our data come from a laundry pick-up, cleaning, and delivery service (called "Washio") that operated in a total of six cities across several regions of the United States during the years 2014–2016. Communication between the company and its customers occurred via an app, which customers downloaded and used to register for the service, place and pay for orders, receive information about pick-up and delivery times and drivers, receive invoices for completed orders, and rate completed transactions. On September 15, 2015, the firm added a function to its app that allowed consumers to electronically tip their delivery drivers using their accounts' credit card numbers. Prior to that date, any tipping that occurred was in cash. When tipping with the app, customers could choose one of three suggested tip amounts, could generate a custom tip amount, or could opt not to tip.

The company used each customer's randomly generated, permanent ID number to assign him or her to one of the following 11 different sets of tip suggestions: (a) $2, $4, and $6; (b) $3, $4, and $5; (c) $3.95, $4, and $4.05; (d) $3.99, $5.99, and $7.99; (e) $4, $6, and $8; (f) $5, $8, and $10; g) 5%, 10%, and 15%; (h) 10%, 15%, and 20%; (i) 12%, 15%, and 18%; (j) 15%, 18%, and 20%; or (k) 15%, 20%, and 25%. None of the tip suggestions were verbally labeled or described, so there were no differences within or across conditions in the explicit meanings of the different suggestion amounts (see Figure 1). The assignment to tip-suggestion condition was permanent, so customers saw the same recommendations each time they patronized the service and were unlikely to be aware of their participation in an experiment. Customers were provided these tipping options when setting up or editing their account preferences (customers could save their tipping preferences and have them automatically applied to future transactions), placing or checking the status of an order (if there was no saved tip preference), and after receiving the electronic invoice for completed orders following delivery of the cleaned clothes (if no tip decision had been previously specified). Customers could also rate the service when they received their invoice (see Figure 1).

Because every customer was assigned to a (nonzero) tip-suggestion condition, this natural field experiment does not allow us to study the impact of *introducing* tipping recommendations; instead, we examine the effects of variation in the tip recommendations by comparing customers assigned to different groups.[3] The various tip-recommendation conditions were selected by the firm for its own reasons, which were not disclosed to us and are not entirely obvious. Nevertheless, they provided an opportunity to test the effects of several interesting variations in suggested tip amounts. In particular, we use these randomly assigned manipulations to examine the effects of providing (i) large versus small suggested tips (holding the asking range constant), (ii) dollar versus percentage amounts of suggested tips (at various bill sizes), (iii) a broad versus narrow range of suggested tips (holding the average suggested tip/contribution constant), and (iv) round versus nonround ($x.00 versus $x.99) suggested tips.

Washio provided data on all of its orders from customers not associated with corporate accounts over its entire period of operation—on 246,132 orders from 50,773 customers. However, 149,416 orders (from 39,504 customers) that were first invoiced before September 15, 2015, were dropped from the main analyses because those orders preceded the app's support of tipping and the customers' receipt of tip recommendations. These pre-September 15, 2015, orders were used only to test the randomness of assignment to tip-suggestion conditions, as explained later. The post-September 15, 2015, data included 2,145 orders from customers with corporate accounts. All of these observations were also dropped from analysis to ensure that customers were personally responsible for paying the bills and tips from all their orders. This left a total of 94,571 orders from 24,637 customers that were retained for the main analyses.

However, the data provided by the company contained numerous missing values, so the number of observations varies across analyses depending on the variables involved.[4] Descriptions of the variables examined in this study are presented in Table 1, and descriptive statistics for the key quantitative variables are presented in Table 2.

4. Analyses and Results

4.1. Analytic Approach

Our experiment had 11 conditions and 13 outcome measures, which created 715 possible paired comparisons and a resulting problem with excessively high experiment-wide error rates. We addressed this problem in three ways. First, we began our analyses of each outcome with an omnibus test of the null hypothesis that the means for all the experimental conditions were equal to one another.[5] Only if this omnibus test indicated that one or more of the means significantly differed from the others did we give much credence to subsequent significant paired comparisons involving that outcome. If the omnibus test for an outcome was not significant, we still report the paired comparisons, but regard those that were significant with some skepticism.

Second, we reduced the likelihood of making type 1 errors by limiting our examination of paired comparisons to only the nine comparisons that differed in a single meaningful way. Specifically, we focused on four paired comparisons that differed only in recommended tip size (E versus A, H versus G, K versus H, and K versus G); two paired comparisons that differed only in the range of recommended tip sizes (A versus B and H versus I); one pair comparison that differed only in recommending even versus odd tip amounts—that is, tip sizes with different left- and right-most digits that differed by only a penny (E versus D); and two paired comparisons that differed only in expression of recommended tips as percentages versus dollars conditional on a bill size of $40 (G versus A and H versus E).

Finally, we conducted a second set of paired comparisons that used the method of List et al. (2016) and code for adjusting analyses for multiple comparisons.

Table 1. Labels and Definitions of Variables Used in This Study

Variable	Definition
Customer ID	A unique random identifier given to each customer and used to assign customers to tip-suggestion conditions.
Tip-Suggestion Condition	A categorical variable indicating which set of tip suggestions the invoiced customer received—A = $2, $4, and $6; B = $3, $4, and $5; C = $3.95, $4, and $4.05; D = $3.99, $5.99, and $7.99; E = $4, $6, and $8; F = $5, $8, and $10; G = 5%, 10%, and 15%; H = 10%, 15%, and 20%; I = 12%, 15%, and 18%; J = 15%, 18%, and 20%; or K = 15%, 20%, and 25%.
Ask-Amount	A variable reflecting the average percentage tip amounts recommended in three otherwise comparable tip-recommendation conditions—condition G (coded as 10), condition H (coded as 15), and condition K (coded as 20). Other conditions were coded as a missing value on this variable.
G%vsA$	A measure of dollar versus percentage recommendation coded as 0 when condition was A ($2, $4, and $6) and coded as 1 when condition was G (5%, 10%, and 15%). Other conditions were coded as a missing value on this variable.
H%vsE$	A measure of dollar versus percentage recommendation coded as 0 when condition was E ($4, $6, and $8) and coded as 1 when condition was H (10%, 15%, and 20%). Other conditions were coded as a missing value on this variable.
Invoice Date	The date the customer was first invoiced for a completed order—typically right after delivery of the cleaned clothes. Recoded as days before and after (centered on) September 15, 2015.
Bill Size	The dollar and cent amount (before coupons or other discounts) of the order.
Bill Size Variability	Absolute value of deviation of bill size from the mean in that customers' tip-suggestion condition.
Tip Likelihood	A binomial variable indicating whether the customer left a tip for the order or not.
Tip Amount	Dollar and cent amount left as a tip (including zero) for the order.
Size of Non-Zero Tip	Dollar and cent amount left as a tip when a tip was left (excluding tips of zero) for the order.
Tip Variability	Absolute value of deviation of tip amount from the mean in that customers' tip-suggestion condition.
Satisfaction Rating Provided	A binomial variable indicating whether the customer left a rating for the order or not.
Satisfaction Rating	Number of stars (on a 5-star scale) the customer gave the invoiced service transaction.
Satisfaction Variability	Absolute value of deviation of satisfaction rating from the mean in that customers' tip-suggestion condition.
Customer Patronage Frequency	Number of times the customer patronized Washio after the app change.
Patronage Frequency Variability	Absolute value of deviation of *Customer Patronage Frequency* from the mean in that customer tip-suggestion condition.
Saved Tip Preference	A binomial variable indicating whether the customer had a saved tip preference or not.
Default Tip	A binomial variable indicating whether the customer left a default tip or not.
Order Number	A variable reflecting the ordinal position of the order in the set of all orders by that customer made after the app change and arranged by date—for example, 1 = customer's first order, 2 = customer's second order, etc.
Median Income	Median income of the ZIP code where the order was picked up and delivered.

Table 2. Descriptive Statistics for the Key Quantitative Variables Used in This Study

Variable	*N*	Minimum	Maximum	Mean	Standard deviation
Ask Amount	26,149	10	20	14.63	4.04
G%vsA$	17,767	0 = A$	1 = G%	0.54	0.50
H%vsE$	16,225	0 = E$	1 = H%	0.55	0.50
Invoice Date (days after 9/15/15)	94,571	0	359	146.16	99.29
Bill Size	94,571	5.99	1365.60	70.58	42.67
Bill Size Variability	94,571	0.00	1295.00	28.78	31.50
Percent Tip	94,571	0.00	133.56	5.46	5.95
Tip Likelihood	94,571	0 = no	1 = yes	0.59	0.49
Size of Non-Zero Tip	55,495	0.01	121.52	6.14	4.94
Tip Amount	94,571	0.00	121.52	3.61	4.85
Tip Variability	94,571	0.00	116.22	3.24	3.39
Satisfaction Rating	74,348	1	5	4.53	0.97
Satisfaction Variability	74,348	0.41	3.59	0.69	0.68
Customer Patronage Frequency	24,637	1	69	3.84	5.11
Patronage Frequency Variability	24,637	0.04	65.16	3.28	3.92
Saved Tip Preference	94,571	0 = no	1 = yes	0.49	0.50
Default Tip	94,571	0 = no	1 = yes	0.49	0.50
Order Number	94,571	1	69	5.82	6.55
Median Income	89,513	$19,887	$172,570	$81,429.13	$28,497.09

This method adjusts for multiple comparisons involving both multiple treatments and multiple outcomes, while taking into account interdependencies among the different treatments and among the different outcomes. The resulting tests control experiment-wide error rates while retaining more statistical power than traditional Bonferroni adjustments. However, this code does not accommodate multiple observations per customer or allow for clustering of error terms, so we performed these adjusted paired comparisons using one randomly selected observation per customer. Only significant adjusted paired comparisons involving an outcome measure with a significant omnibus test are regarded as compelling in our description and discussion of results below.[6]

4.2. Check on Randomization of Tip Suggestions

The 11 tip-suggestion conditions were assigned to customers based on the customers' pseudorandom ID numbers. The algorithm used did not ensure equal sample sizes across conditions, but its use of computer-generated pseudorandom ID numbers did give every customer the same chance as other customers to be in each of the conditions. Because all existing and new customers were assigned to tip-suggestion conditions, we were able to compare pre-September 15, 2015, orders across the conditions to which the customers placing those orders had been assigned with the expectation that no differences in customer satisfaction, patronage frequency, and bill size across tip-suggestion conditions should be evident before customers were exposed to those suggestions. As expected, omnibus tests using errors clustered within customer indicated that differences across assigned conditions in the mean presuggestion bill sizes, patronage frequency, and satisfaction ratings were all within chance levels (see Table 3).

4.3. Effects of Tip Suggestions on Tipping

Descriptive statistics for the measures of tipping behavior in this study are presented by tip-suggestion condition in Table 4. Omnibus tests using errors clustered within customer, which are also reported in Table 4, indicated that the different tip-suggestion conditions did result in reliable differences in the likelihood of tipping, size of nonzero tips, tip amount, and the variability in tip amounts (all $p < 0.0001$). Paired comparisons of selected conditions are also presented in Table 4 and are summarized in Table 8. We tested for mean differences in the outcome variables for sets of pairs for which the interpretations were most clear. For example, comparing condition A ($2, $4, and $6) to condition B ($3, $4, and $5) allows us to study the effect of the *range* of suggestions holding the average amount constant, whereas comparing condition A to condition E ($4, $6, and $8) allows us to study the effect of suggesting larger tip amounts while holding the range constant. The table indicates whether each pairwise difference in means was statistically significant at the 5% level when adjusting for multiple comparisons and multiple outcomes as well as using the naïve (unadjusted) test. In general, these comparisons indicated that both suggesting larger tip amounts (also known as (aka) ask size) and suggesting percentage versus dollar tip amounts (aka percentage ask) consistently increased the size of nonzero tips, average tip amounts (including tips of zero), and the variability in tip amounts. At the same time, a larger ask size decreased the likelihood of leaving a tip, though the positive impact on nonzero

Table 3. Pretreatment Means (and Standard Deviations) of Customer Satisfaction, Repatronage, and Spending for Orders Before September 15, 2015, by Tip-Suggestion Condition

Tip-suggestion condition	Satisfaction rating	Customer patronage frequency[a]	Bill size
A: \$2, \$4, \$6	4.48 (0.98)	3.76 (5.55)	54.13 (34.24)
B: \$3, \$4, \$5	4.51 (0.97)	3.74 (5.28)	54.40 (47.66)
C: \$3.95, \$4, \$4.05	4.48 (0.99)	3.84 (5.64)	54.72 (36.38)
D: \$3.99, \$5.99, \$7.99	4.51 (0.96)	3.69 (5.39)	54.63 (38.29)
E: \$4, \$6, \$8	4.50 (0.94)	3.59 (5.13)	54.33 (35.56)
F: \$5, \$8, \$10	4.51 (0.95)	3.63 (5.14)	55.26 (37.30)
G: 5%, 10%, 15%	4.49 (0.97)	3.75 (5.51)	54.75 (38.35)
H: 10%, 15%, 20%	4.52 (0.95)	3.78 (5.61)	54.96 (50.44)
I: 12%, 15%, 18%	4.46 (0.98)	3.75 (5.40)	54.22 (36.23)
J: 15%, 18%, 20%	4.50 (0.96)	3.57 (5.02)	54.24 (36.61)
K: 15%, 20%, 25%	4.50 (0.95)	3.71 (5.58)	56.35 (39.67)
Omnibus test (df)	F(10, 27,750)[b]	F(10, 39,112)[c]	F(10, 39,112)[b]
Value of test statistic	1.00	0.93	0.58
R^2	0.0003	0.0002	0.0002
N orders/customers	97,078/27,751	NA/39,113	145,324/39,113

Note. NA, not applicable.
[a]Number of times the customer patronized Washio before the app change.
[b]Robust error terms clustered within customer.
[c]Robust error terms.

tip size was large enough to outweigh the negative impact on tip likelihood, resulting in higher tip amounts overall. We also found that suggesting a larger range of tip amounts and suggesting round versus nonround tip amounts had no consistent or reliable effects on any of the tipping measures.

4.3.1. Additional Analyses of Ask-Size Effects. The effects of ask size on tipping replicate similar effects reported by Haggag and Paci (2014). However, recommending larger tips reliably reduced tipping likelihood in this study only when the top recommendations were 15%, 20%, or 25% (condition K), and a posthoc test comparing these significant ask-size effects with the others involving different conditions produced a $\chi^2(1)$ of 9.62, $p < 0.002$. Thus, the negative ask-size effect on tipping likelihood appears to be limited to cases where one set of the recommended amounts exceeds familiar and normative levels.

To explore the robustness of the ask-size effect on tip amounts with as much statistical power as possible, a new variable called "*ask-amount*" was created to reflect the mean percentage tip requested in three otherwise comparable tip-recommendation conditions—condition G (coded as 10), condition H (coded as 15), and condition K (coded as 20). Restricting the analysis to these three conditions, we regressed the tip amount on *ask-amount* and its interaction with several other variables; the results are presented in Table 5. As in the main analysis, a larger *ask-amount* produced larger tips. On average, an increase in the mean tip suggestion of 1 percentage point led to a \$0.17 increase in tips. This effect was moderated by some, but not all, of the variables we examined. It was reliably stronger the larger the bill size, which reinforces the idea that you get what you ask for because the dollar size of a given percentage increases with bill size.

The effect of *ask-amount* on tip amount was substantially larger for those with a saved tip preference, though it remained positive and statistically significant even for those without a saved tip preference.

Table 4. Means (and Standard Deviations) of Measures of Tipping Behavior by Tip-Suggestion Condition

Tip-suggestion condition	Size of nonzero tip	Tip likelihood	Tip amount	Tip variability	Has saved tip preference	Left default tip
A: \$2, \$4, \$6	\$3.57$^{\underline{BEG}}$	0.64^{beG}	\$2.28$^{b\underline{EG}}$	\$1.82$^{b\underline{EG}}$	0.52^{bEg}	0.61$^{B\underline{EG}}$
	(1.81)		(2.24)	(1.32)		
B: \$3, \$4, \$5	\$3.79$^{\underline{A}}$	0.61^{a}	\$2.30^{a}	\$1.84^{a}	0.49^{a}	0.56^{A}
	(1.33)		(2.12)	(1.05)		
C: \$3.95, \$4, \$4.05	\$4.04	0.60	\$2.43	\$1.98	0.50	0.49
	(1.38)		(2.25)	(1.07)		
D: \$3.99, \$5.99, \$7.99	\$4.38^{E}	0.60^{e}	\$2.62^{e}	\$2.22^{E}	0.49^{e}	0.47^{E}
	(1.49)		(2.44)	(1.01)		
E: \$4, \$6, \$8	\$4.62$^{\underline{A}D\underline{H}}$	0.60^{adh}	\$2.79$^{\underline{A}d\underline{H}}$	\$2.31 $^{\underline{A}D\underline{H}}$	0.47^{Adh}	0.52$^{\underline{A}DH}$
	(1.61)		(2.58)	(1.16)		
F: \$5, \$8, \$10	\$5.43	0.57	\$3.12	\$2.75	0.45	0.44
	(2.23)		(3.17)	(1.58)		
G: 5%, 10%, 15%	\$6.66$^{\underline{AHK}}$	0.59$^{Aeh\underline{K}}$	\$3.91$^{\underline{AHK}}$	\$3.64 $^{\underline{AHK}}$	0.51$^{ah\underline{K}}$	0.52$^{\underline{A}h\underline{K}}$
	(5.03)		(5.06)	(3.51)		
H: 10%, 15%, 20%	\$8.07$^{\underline{EGiK}}$	0.58$^{egi\underline{K}}$	\$4.70$^{\underline{EG}i\underline{K}}$	\$4.27$^{\underline{EGiK}}$	0.50$^{egi\underline{K}}$	0.48$^{Eg\underline{IK}}$
	(6.00)		(6.07)	(4.31)		
I: 12%, 15%, 18%	\$8.52$^{\underline{h}}$	0.59^{h}	\$5.05^{h}	\$4.46$^{\underline{h}}$	0.51^{h}	0.43$^{\underline{H}}$
	(5.80)		(6.12)	(4.20)		
J: 15%, 18%, 20%	\$9.62	0.56	\$5.36	\$5.14	0.47	0.39
	(7.40)		(7.30)	(5.19)		
K: 15%, 20%, 25%	\$11.23$^{\underline{GH}}$	0.50$^{\underline{GH}}$	\$5.63$^{\underline{GH}}$	\$5.82$^{\underline{GH}}$	0.44$^{\underline{GH}}$	0.43$^{\underline{GH}}$
	(7.62)		(7.79)	(5.18)		
Omnibus test (df)	$F(10, 16{,}530)$	Wald $\chi^2(10)$	$F(10, 24{,}636)$	$F(10, 24{,}636)$	Wald $\chi^2(10)$	Wald $\chi^2(10)$
Value of test statistic	271.72****	56.41****	99.27****	311.29****	23.95**	192.20****
R^2/Pseudo R^2	0.24	0.003	0.06	0.16	0.0016	0.0098
N orders/customers	55,495/16,531	94,571/24,637	94,571/24,637	94,571/24,637	94,571/24,637	94,571/24,637

Notes. Alphabetic superscripts indicate the conditions each statistic was compared with; capitalized superscripts mark comparisons reliable at the 0.05 level unadjusted for multiple comparisons; and underlined superscripts mark comparisons reliable at the 0.05 level adjusted for multiple comparisons using the method of List et al. (2016). Analyses adjusting for multiple comparisons were based on one randomly selected observation per customer and included all paired comparisons, but tests were performed separately for the following groups of outcomes: {*size of nonzero tip*}, {*tip likelihood*, *tip amount*, *tip variability*}, and {*saved preference*, *default tip*}. For the omnibus tests, we report F statistics for regressions estimated using ordinary least squares, and χ^2 statistics from Wald tests for regressions estimated using binomial logistic regression. Robust standard errors for omnibus regressions are adjusted for clustering at the customer level.

$^{**}p < 0.01$; $^{****}p < 0.0001$.

Unfortunately, it is difficult to precisely interpret these effects because tip recommendations also affected the tendency to save a tip preference (see Table 4), and therefore the *ask-amount-by-saved-tip-preference* interaction is likely to be endogenous. This interaction may result from a desire to minimize the cognitive effort of tipping, which might incline people both to go along with tip recommendations and to save their tip preference.[7]

We found no significant interaction between *ask-amount* and the median income of the customer's ZIP code or between *ask-amount* and customer satisfaction with the service. We also investigated whether the impact of the tip suggestion varied with the number of times the customer was exposed to the treatment. To do so, we interacted *ask-amount* with posttreatment order number, where, for example, order number 1 is the first order the customer placed after September 15, 2015. In order to control for compositional effects, we restrict the sample to only those customers who placed at least four orders, and we look at the effects across only those first four orders.[8] There was a positive and statistically significant interaction between *ask-amount* and *order number* over the first four orders, which provides suggestive evidence that the magnitude of the treatment effect may have increased over time (column (6) of Table 5). However, one limitation of this analysis is that an app update was required in order to enable the tipping functionality and tip suggestions for customers using mobile devices. Unfortunately, we do not know precisely when each customer updated the app, so these results are confounded with the fact that customers were less likely to have been exposed to the treatment during earlier orders.[9] To attempt to disentangle these effects, we ran two additional specifications. First, we repeated the specification from Table 5 using the size of nonzero tips as the outcome variable, which ensures that the sample is restricted to customers who have been exposed to the tip suggestions; and, second, we restricted the analysis to customers who placed their first posttreatment order

Table 5. Tip Amount vs. Ask Amount and Interactions with Other Variables

	(1)	(2)	(3)	(4)	(5)	(6)
Variable	Tip amount	Tip amount	Tip amount	Tip amount	Tip amount	Tip amount
Ask amount	0.172***	–0.114	0.0529*	0.0486	0.154*	0.0905*
	(0.0261)	(0.0649)	(0.0208)	(0.0734)	(0.0744)	(0.0396)
Bill size		0.0106				
		(0.0153)				
Ask × Bill size		0.00394***				
		(0.00104)				
Saved preference			0.422			
			(0.584)			
Ask × Saved preference			0.323***			
			(0.0422)			
Median income ($000)				–0.0263*		
				(0.0123)		
Ask × Median income				0.00159		
				(0.000876)		
Satisfaction rating					0.303	
					(0.224)	
Ask × Satisfaction rating					0.00617	
					(0.0161)	
Order number						–0.239
						(0.176)
Ask × Order number						0.0481***
						(0.0133)
R^2	0.012	0.233	0.187	0.014	0.016	0.028
Customers	6,714	6,714	6,714	6,509	4,944	2,054
Observations	26,149	26,149	26,149	24,793	20,537	8,216

Notes. The table reports coefficients from ordinary least squares regressions of tip amount on ask amount and a number of interactions, with the sample restricted to only those customers in treatment conditions G (5%, 10%, and 15%), H (10%, 15%, and 20%), and K (15%, 20%, and 25%). *Ask amount* is defined as the suggested percentage of the middle option, so it is equal to 10 or 15 or 20. *Median income* represents the median income of the customer's ZIP code; and *saved preference* is an indicator for whether the customer chose to save their tip preference. In order to control for compositional effects, the sample in column (6) is restricted to only those customers with four or more orders *and* only the first four orders of those customers; thus, order number is an integer ranging from 1 to 4. Robust standard errors are in parentheses, adjusted for clustering at the customer level.

$^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

after November 1, 2015, increasing the likelihood that these customers would have the tipping option available at their first order. In both cases, we again found positive and significant effects, suggesting that the observed differences by order number may not be fully explained by delays in updating the app.[10]

4.3.2. Additional Analyses of Percentage-Ask Effects. To explore the processes underlying the percentage-ask effect, two new indicator variables—*G%vsA$* and *H%vsE$*—were created to contrast the two types of requests, holding other ask characteristics constant. *G%vsA$* was coded as 0 for condition A ($2, $4, and $6) and 1 for condition G (5%, 10%, and 15%). Similarly, *H%vsE$* was coded as 0 for condition E ($4, $6, and $8) and 1 for condition H (10%, 15%, and 20%). For both pairs of conditions, the dollar and percentage recommendations amounted to the same ask size when bill size was exactly $40, whereas the dollar recommendations amounted to a larger ask size below $40, and the percentage recommendations amounted to a larger ask size above $40.

Consistent with the main analyses, both percentage-ask indicators had positive effects on tip amount (see Table 6). Because roughly 80% of the observations involved bill amounts over $40, these percentage-ask effects could be due to differences in the recommended tip amounts. In other words, they could be disguised ask-size effects. Consistent with this possibility, the effects of both variables grew reliably stronger as bill size, and the resulting difference in the ask size of the percentage versus dollar recommendations, increased. However, the effects of *G%vsA$* and *H%vsE$* were not reliably reversed at bill sizes below $40, as would be expected if percentage versus dollar asks had only ask-size effects.

To identify potential percentage versus dollar ask effects that were independent of ask size, we examined the effects of *G%vsA$* and *H%vsE$* in a subsample with bill sizes between $39.50 and $40.50, where the difference in ask size across percentage versus dollar asks was minimal (see Table 6). *H%vsE$* had no reliable effects on tip amount in this subsample analysis, but *G%vsA$* had a reliably positive effect (see Table 6). Perhaps subjects' familiarity with tipping 15%–20% in restaurants made the 5% and 10% options in condition G seem too small and encouraged selection of the 15% option, whereas subjects' familiarity with all the options of \$2, \$4, and \$6 in condition A created no comparable encouragement to select the \$6 option. This would explain why $H\%vsE\$$ had no reliable effects because it involved fewer subnormative tip percentage options.

4.4. Effects of Tip Suggestions on Customer Satisfaction, Patronage Frequency, and Spending

Descriptive statistics for the measures of customer satisfaction, patronage frequency, and spending in this study are presented by tip-suggestion condition in Table 7. Omnibus tests using errors clustered within customer, which are also reported in Table 7, indicated that the different tip-suggestion conditions did not reliably affect average customer satisfaction, patronage frequency, or spending. Nor did the different tip-suggestion conditions affect variability in customer spending. They did affect variability in customer satisfaction and patronage frequency—variability in customer satisfaction was slightly lower for tip-condition G than for the other conditions, and variability in patronage frequency was slightly lower in conditions E and J than the other conditions. However, those effects were idiosyncratic and neither theoretically nor economically important.

Simple paired comparisons did find that satisfaction ratings were higher in condition G (5%, 10%, and 15%) than in conditions H (10%, 15%, and 20%) and K (15%, 20%, and 25%), but the difference in satisfaction ratings was small in magnitude (less than 0.1 on a five-point scale) and did not remain significant in analyses adjusting for multiple comparisons. Furthermore, satisfaction ratings did not differ between conditions H and K or between conditions A and E. Thus, the paired comparisons of tip recommendations that affected tipping did not reliably affect central tendencies or variabilities in customer satisfaction, patronage frequency, or spending (see Tables 7 and 8). This means that increasing the amounts customers tipped did not affect these outcomes either, which suggests that none of the self-perception/justification, warm-glow, or value-perception processes observed in other contexts operated here.

We next examine whether our failure to find evidence that tip recommendations affected the tipper's satisfaction could be attributable to problems with our measurement of satisfaction. First, customers did not have to supply satisfaction ratings, and it is possible that only those who were satisfied chose to provide the ratings. This self-selection process could have hidden any negative effects of tip recommendations on satisfaction ratings. However, tip recommendations did not affect the likelihood of providing satisfaction ratings (see Table 7), so if completion of the ratings is itself a sign of satisfaction, as this argument assumes, then the null results involving mean satisfaction ratings were replicated using this new measure.

Second, we consider the possibility that satisfaction ratings were so high that the resulting restriction of range may have hidden the true effects of tip recommendations on those ratings. However, those satisfaction ratings were not too high or restricted to prevent them from being reliably related to tip amount ($B = 0.28$, standard error (S.E.) $= 0.03$, $p < 0.001$), patronage frequency ($B = 0.84$, S.E. $= 0.09$, $p < 0.001$), or bill size ($B = -1.40$, S.E. $= 0.23$, $p < 0.001$) in separate regressions of these variables on satisfaction using robust standard errors clustered with customer. Thus, if restriction of range in the satisfaction ratings hid the true effects of tip recommendations on satisfaction, those true effects must have been very small. Also arguing for, at best, small and inconsequential true effects of tip recommendations on satisfaction is the fact that those recommendations did not affect repatronage frequency or spending either.[11]

4.5. Effects of Tip Suggestions on Consumer Use of Default Tip Amounts

The proportion of orders on which customers tipped one of the recommended or default options is presented by tip-recommendation condition in the final column of Table 4. Omnibus tests using errors clustered within customer, which are also reported in Table 4, indicated that the different tip-suggestion conditions did affect use of the suggested tip amounts. Paired comparisons of selected conditions are also presented in Table 4 and are summarized in Table 8. As might be expected, these tests indicated that consumers were more likely to use one of the recommended or default tip options when the suggested tips were smaller and had a larger range.

5. Discussion

5.1. Summary of Key Findings

Among the noteworthy findings from our analysis of this natural field experiment on the effects of various tip recommendations are the following: (i) Arithmetically equivalent percentage versus dollar tip recommendations increased tip amounts in the case where some of the percentages recommended were subnormative; (ii) increasing the range of recommended tips increased selection/use of one of the recommended amounts, but did not affect the amount tipped; and (iii) round and nonround tip recommendations (with different left-most as well as right-most digits) did not produce different levels of tipping. However, the most interesting and important findings are that

Table 6. The Effect of Recommendation Format (Percentage vs. Dollar) on Tip Amount

	All bill sizes		All bill sizes		Bill size < \$40		\$39.50 < Bill size < \$40.50	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variable	Tip amount	Tip amount	Tip amount	Tip amount	Tip amount	Tip amount	Tip amount	Tip amount
G%vsA$	1.635***		−1.359**		0.116		0.846**	
	(0.151)		(0.414)		(0.133)		(0.277)	
H%vsE$		1.912***		−2.171***		0.255		0.264
		(0.164)		(0.384)		(0.152)		(0.356)
Bill size			0.00970***	0.00785***				
			(0.00114)	(0.00140)				
G%vsA$ × Bill size			0.0419***					
			(0.00661)					
H%vsE$ × Bill size				0.0582***				
				(0.00611)				
R^2	0.040	0.038	0.201	0.217	0.001	0.003	0.047	0.003
Customers	4,423	4,307	4,423	4,307	1,692	1,705	279	234
Observations	17,767	16,225	17,767	16,225	3,473	3,480	369	299

Notes. The table reports the coefficients from ordinary least squares regressions of tip amount on an indicator variable for whether the tip recommendation was provided in percentage terms. *G%vsA$* is equal to 1 when the suggestion condition is G (percentage) and 0 when it is A (dollar amounts), and *H%vsE$* is defined analogously. Columns (3) and (4) include interactions between the treatment indicators and bill size. Each sample is restricted to only the two suggestion conditions being compared (i.e., G and A or H and E). Robust standard errors are in parentheses, adjusted for clustering at the customer level.

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

recommending larger tip amounts decreased the likelihood of tipping (when the recommendations exceeded familiar and normative amounts) and increased the size of tips that were left as well as overall tip revenues, but had no comparable or opposite effects on customer satisfaction, repatronage frequency, or spending.[12] These latter results have implications for (i) our confidence in the existence, or not, of the effects; (ii) possible boundary conditions for the processes hypothesized to underlie the effects; and (iii) the advisability of using larger asks to boost voluntary payments. These issues are discussed below.

5.2. Confidence in Ask-Size Effects

It has been noted that replication rates in the psychological sciences are disappointingly low (Open Science Collaboration 2015) and that published effects are more likely to be false than true (Ioannidis 2005), so readers may wonder about the confidence they should place in our conclusions. Fortunately, the large sample size in this study—involving 94,571 orders from 24,637 customers—provides high levels of power and, therefore, tight confidence intervals. For example, the 95% confidence intervals for our nine unadjusted pair-wise comparisons of mean tip amounts ranged from ±\$0.18 to ±\$0.54 (on a measure with an overall standard deviation of \$4.85), and those for our nine unadjusted pair-wise comparisons of mean satisfaction ratings ranged from ±0.05 to ±0.07 (on a five-point scale with an overall standard deviation of 0.97).

For a more complete perspective on the level of confidence that our findings warrant, we use a formula described by Maniadas et al. (2014) to calculate the poststudy probability (PSP) that our positive conclusions about ask-size effects are true, given various priors regarding those probabilities. For positive effects, Maniadas et al. (2014) argue that "PSP is equal to the number of true associations which are declared true divided by the number of all associations which are declared true: $PSP = (1-\beta)\pi/((1-\beta)\pi + \alpha(1-\pi))$," where π is the fraction of associations that are true (or are assumed to be true a priori), $1-\beta$ is the statistical power of the study, and α is the alpha level used in the study. We calculated the posthoc powers of the pairwise comparisons of ask size in our study using the standard formula for the power of an inference about the means from two independent samples. In these calculations, we used the number of customers in each condition, rather than the number of observations, as the sample sizes, so our power estimates are likely to be conservative.[13] Nevertheless, all of the comparisons had a posthoc power of 0.99, so that value was used to calculate the poststudy probabilities of ask-size effects on tip amounts of the magnitudes we observed given various priors (see Table 9). The poststudy confidence that should be placed on these effects depends on the level of confidence in the effects held prior to the study, but is substantially greater than those priors. For a prior of 0.01, the poststudy probabilities of the effects increase 1,600%, and for a prior of 0.5, they increase 90%. The priors would have

Table 7. Means (and Standard Deviations) of Various Measures by Tip-Suggestion Condition

Tip-suggestion condition	Satisfaction rating	Satisfaction variability	Customer patronage frequency	Patronage frequency variability	Bill size	Bill size variability	Satisfaction rating provided
A: \$2, \$4, \$6	4.53[beg]	0.69[beG]	3.96[beg]	3.36[bEg]	71.36[beg]	28.86[beg]	0.78[beg]
	(0.96)	(0.67)	(5.21)	(3.98)	(42.09)	(30.63)	
B: \$3, \$4, \$5	4.52[a]	0.70[a]	3.82[a]	3.23[a]	69.44[a]	27.51[a]	0.79[a]
	(0.97)	(0.67)	(5.06)	(3.89)	(39.50)	(28.34)	
C: \$3.95, \$4, \$4.05	4.52	0.70	3.84	3.28	69.96	28.29	0.78
	(0.97)	(0.67)	(5.22)	(4.06)	(40.93)	(29.58)	
D: \$3.99, \$5.99, \$7.99	4.52[e]	0.71[e]	3.83[e]	3.29[e]	71.13[e]	28.66[e]	0.80[e]
	(1.01)	(0.72)	(5.10)	(3.90)	(42.42)	(31.28)	
E: \$4, \$6, \$8	4.54[adh]	0.69[adh]	3.72[adh]	3.12[Adh]	69.42[adh]	28.74[adh]	0.78[adh]
	(0.97)	(0.68)	(4.76)	(3.60)	(42.94)	(31.90)	
F: \$5, \$8, \$10	4.54	0.68	3.82	3.25	71.36	30.00	0.79
	(0.94)	(0.65)	(5.20)	(4.05)	(44.61)	(33.02)	
G: 5%, 10%, 15%	4.59[aHK]	0.64[AHK]	4.07[ahk]	3.57[aHK]	71.47[ahk]	29.52[ahk]	0.78[ahk]
	(0.93)	(0.67)	(5.65)	(4.37)	(42.91)	(31.15)	
H: 10%, 15%, 20%	4.53[eGik]	0.70[eGik]	3.81[egik]	3.29[eGik]	70.07[egik]	28.54[egik]	0.80[egik]
	(0.98)	(0.68)	(4.99)	(3.75)	(42.31)	(31.24)	
I: 12%, 15%, 18%	4.53[h]	0.70[h]	3.88[h]	3.27[h]	69.78[h]	27.81[h]	0.80[h]
	(0.98)	(0.68)	(4.96)	(3.74)	(42.01)	(31.48)	
J: 15%, 18%, 20%	4.52	0.70	3.63	3.02	70.60	29.56	0.77
	(0.95)	(0.65)	(4.67)	(3.56)	(48.05)	(37.88)	
K: 15%, 20%, 25%	4.51[Gh]	0.71[Gh]	3.80[gh]	3.31[Gh]	72.05[gh]	29.39[gh]	0.78[gh]
	(0.98)	(0.67)	(5.18)	(3.98)	(42.55)	(30.78)	
Omnibus test (df)	$F(10, 18{,}135)$	$F(10, 18{,}135)$	Wald $\chi^2(10)$	Wald $\chi^2(10)$	$F(10, 24{,}636)$	$F(10, 24{,}636)$	Wald $\chi^2(10)$
Value of Test Statistic	1.21	3.04***	10.62	26.27**	0.72	1.28	10.75
R^2/Pseudo R^2	0.0004	0.0008	0.0002	0.0005	0.0004	0.0006	0.0006
N orders/customers	74,348/18,136	74,348/18,136	NA/24,637	NA/24,637	94,571/24,637	94,571/24,637	94,571/24,637

Notes. Alphabetic superscripts indicate the conditions each statistic was compared with; capitalized superscripts mark comparisons reliable at the 0.05 level unadjusted for multiple comparisons; and underlined superscripts mark comparisons reliable at the 0.05 level adjusted for multiple comparisons using the method of List et al. (2016). Analyses adjusting for multiple comparisons were based on one randomly selected observation per customer and included all paired-comparisons, and tests were performed separately for the following groups of outcomes: {*satisfaction rating*, *satisfaction variability*}, {*patronage frequency*, *bill size*, *rating provided*}, and {*frequency variability*, *bill size variability*}. *F* statistics are reported for ordinary least-squares regressions, and Wald χ^2 statistics for regressions estimated using negative binomial regression (*patronage frequency* and *variability*) or logistic regression (*rating provided*). Robust standard errors for omnibus regressions are adjusted for clustering at the customer level.

$^{**}p < 0.01$; $^{***}p < 0.001$; $^{****}p < 0.0001$.

to be lower than 0.05 to bring the poststudy probabilities below 0.5.

Maniadas et al. (2014) do not provide a formula to calculate the PSP of a null result (PSPN). Nor could we find one that suited our needs elsewhere. However, using the same logic as previously, we argue that PSPN is equal to the number of true null effects, which are declared null divided by the number of all effects which are declared null: $PSPN = (\pi'(1-\alpha))/(\pi'(1-\alpha) + \beta\ (1-\pi'))$, where true null effects are defined as absolute differences between condition means of less than the smallest amount considered meaningful, π' is prior probability of a true null effect, β is the chance of getting a nonsignificant result if the smallest effect considered meaningful is true, and α is alpha level. We used this formula along with our study sample sizes (numbers of customers per condition) and standard deviations to calculate the poststudy probabilities that our ask-size null effects on satisfaction ratings were true, where true null effects are defined as those whose absolute value is less than 0.1 (see Table 9).[14] Again, the poststudy confidence that should be placed on these null effects depends on the level of confidence in the null effects held prior to the study, but is substantially greater than those priors. For a prior of 0.01, the poststudy probabilities of these null effects increase 300%–400%, and for a prior of 0.5, they increase 46%–56%. Priors as low as 0.25 result in poststudy probabilities of at least some of these null effects to exceed 0.5.

5.3. Potential Boundary Conditions for Hypothesized Underlying Processes

The finding that larger tip recommendations affected customers' tipping behavior, but not their satisfaction, repatronage frequency, or spending, suggests that

Table 8. Reliability (With and Without Adjustment for Multiple Comparisons) and Direction of Effect for Key Paired Comparisons

Paired comparison	Tip likelihood	Size of nonzero tip	Tip amount	Tip variability	Satisfaction	Satisfaction variability	Patronage frequency	Patronage frequency variability	Bill size	Bill size variability	Satisfaction rating provided	Has saved tip preference	Left default tip
Suggest large vs. small amounts													
E vs. A	n.s.	+	+	+	n.s.	n.s.	n.s.	(–)	n.s.	n.s.	n.s.	(–)	—
H vs. G	n.s.	+	+	+	(–)	(+)	n.s.	(–)	n.s.	n.s.	n.s.	n.s.	n.s.
K vs. H	—	+	+	+	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	—	—
K vs. G	—	+	+	+	(–)	(+)	n.s.	(–)	n.s.	n.s.	n.s.	—	—
Suggest wide vs. narrow range of amounts													
A vs. B	n.s.	—	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	(+)
H vs. I	n.s.	n.s.	n.s.	—*	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	+
Suggest round vs. nonround amounts													
E vs. D	n.s.	(+)	n.s.	(+)	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	(+)
Suggest percentage vs. dollar amounts													
G vs. A	(–)	+	+	+	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	—
H vs. E	n.s.	+	+	+	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	(–)

Notes. Conditions used in comparisons were: A = \$2, \$4, and \$6; B = \$3, \$4, and \$5; E = \$4, \$6, and \$8; D = \$3.99, \$5.99, and \$7.99; G = 5%, 10%, and 15%; H = 10%, 15%, and 20%; I = 12%, 15%, and 18%; and K = 15%, 20%, and 25%. Signs mark larger (+), smaller (–), or nonsignificantly different (n.s.) means in the condition on the left than in the condition on the right. All alphas were 0.05. Effects in parentheses were reliable in nonadjusted analysis of the entre sample, but were unreliable in analysis of a reduced sample (one randomly selected observation per customer) that adjusted for multiple comparisons using the method and code of List et al. (2016). Effects with asterisks were not significant in nonadjusted analysis of the entire sample, but were significant in the analysis of a reduced sample that adjusted for multiple comparisons. Effects with neither parentheses nor asterisks were reliable in both adjusted and nonadjusted analyses.

Table 9. Poststudy Probabilities of Ask-Size Effects and Ask-Size Null Effects Given Various Prior Probabilities of Those Effects

	Ask-size effects on tip amount				Ask-size null effects on satisfaction ratings			
Prior probability	E – A ≥ \$0.51	H – G ≥ = \$0.79	K – H ≥ \$0.93	K – G ≥ \$1.72	|E – A| < 0.1	|H – G| < 0.1	|K – H| < 0.1	|K-G| < 0.1
Prior probability of effect (π)								
0.01	0.17	0.17	0.17	0.17				
0.05	0.51	0.51	0.51	0.51				
0.10	0.69	0.69	0.69	0.69				
0.25	0.87	0.87	0.87	0.87				
0.50	0.95	0.95	0.95	0.95				
Prior probability of null effect (π′)								
0.01					0.04	0.03	0.03	0.03
0.05					0.16	0.14	0.12	0.13
0.10					0.29	0.26	0.23	0.24
0.25					0.55	0.51	0.47	0.48
0.50					0.78	0.76	0.73	0.74

Notes. Conditions used in comparisons of ask size were: A = \$2, \$4, and \$6; E = \$4, \$6, and \$8; G = 5%, 10%, and 15%; H = 10%, 15%, and 20%; and K = 15%, 20%, and 25%. The poststudy probabilities of our observed ask-size effects being true is $(1-\beta)\pi/((1-\beta)\pi+\alpha(1-\pi))$, where $1-\beta$ is power (posthoc power from current study contrast in this case), π is the prior probability of effect truth, and α is the alpha level (0.05 in this case). The poststudy probabilities of our observed ask-size null effects being true is $(\pi'(1-\alpha))/((\pi'(1-\alpha))+\beta(1-\pi'))$, where true null effects are defined as absolute differences between condition means that are less than the smallest meaningful amount (0.1 out of a five-point scale in this case), π' is the prior probability of a null effect being true, β is the chance of getting a nonsignificant result if the smallest meaningful effect is true, and α is the alpha level (0.05 in this case).

neither the processes underlying the tip-recommendation effects on tipping nor consumer tipping itself affect these other customer outcomes. In particular, the finding that asking for supernormative tips decreased tipping likelihood, but not customer satisfaction, repatronage, and spending (or increased the variability in these measures), suggests that the former effect is not due to strong anger or other negative affect that might be expected to also impact these latter outcome variables. Previous research has found that unsolicited recommendations can lead to negative affect and dissatisfaction, as well as reactive or contrary responses (Fitzsimons and Lehmann 2004), but that does not seem to have happened in this case. Why our supernormative tip recommendations appear to have created reactance (i.e., negative effects on compliance) without negative affect impacting satisfaction and other outcomes is not clear. Perhaps the fact that the tip recommendations benefited drivers rather than the firm softened consumers' emotional reactions to the firm, but not their behavioral reactions to the recommendations themselves. Developing and testing more ideas about the conditions under which recommendations produce reactance with and without negative affect toward the recommender is an interesting issue left to future research.

The fact that inducing people to tip more did not affect the tippers' satisfaction ratings, repatronage frequency, or subsequent spending also suggests that the self-perception/justification, warm-glow, and value-perception processes observed in other contexts did not operate here.[15] Characteristics of the current context that might be boundary conditions responsible for the failure to find evidence of these processes are discussed below.

5.3.1. Self-Perception/Justification and Warm-Glow Effects of Tipping. Voluntary compliance with small requests has been shown to evoke self-perception and consistency processes that increase subsequent compliance with larger requests (Burger 1999), and voluntary prosocial behavior has been shown to generate a positive feeling or warm glow that enhances customer satisfaction and repatronage intentions (Giebelhausen et al. 2017). Given this research and the fact that tipping is a voluntary behavior that benefits another person, it seemed likely that tipping would evoke self-perception/justification and warm-glow processes that increase customer satisfaction, repatronage, and spending, but such effects were not observed. Something about the current study conditions appears to interfere with these processes and their effects.

Self-perception/justification and warm-glow processes depend on the actors' attributions for their voluntary behavior. For example, both self-perception and warm-glow effects are diminished when external incentives for the voluntary behaviors diminish the behaviors' signaling about the self (Dillard et al. 1984, Burger and Caldwell 2003, Giebelhausen et al. 2016, Giebelhausen and Chun 2017). Thus, it is possible that these processes and their effects are limited to contexts where there is a dominant internal attribution for behavior. The current study context may not have met this condition, because tipping is driven by many motives, such as desires to reward good service, help servers, buy future service, buy social esteem, and comply with social norms (Lynn 2015a,b, 2016). The multifaceted nature of motivations for tipping may interfere with the self-attributions underlying self-perception/justification and warm-glow processes and, thereby, diminish the effects of tipping on subsequent attitudes and behaviors. Because many other real-world behaviors are similarly driven by multiple motivations, such a boundary condition for self-perception/justification and warm-glow processes and their effects is potentially important and deserves further investigation.

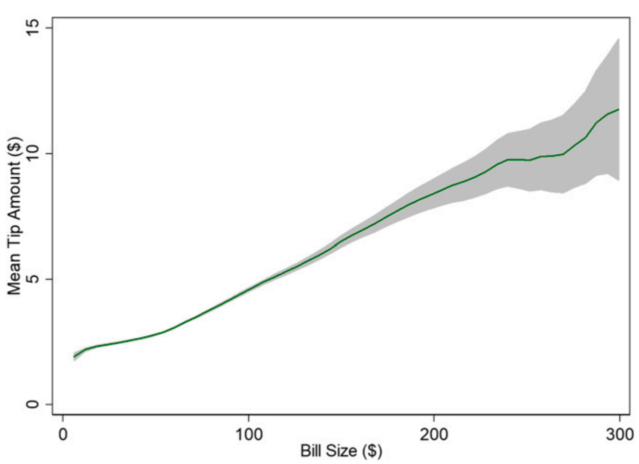
5.3.2. Value-Perception Effects of Tipping. Our finding that increasing customers' tip expenditures did not adversely affect their satisfaction, repatronage frequency, or spending suggests that those customers were not very sensitive to the costs associated with tipping. This conclusion is further supported by the fact that we found no marginal decrease in the positive relationship between tip amounts and bill size as bill size increased; that is, that the relationship between tip amounts and bill size was relatively linear, as shown in Figure 2. When we trim the five observations where bill size was above $750 (to reduce the influence of outliers), a regression of tip amount on bill size and bill size squared results in a quadratic term that is small and statistically insignificant ($B_{\text{quadratic}} = -0.00001$, $\text{S.E.}_{\text{clustered within customer}} = 0.00002$, $p > 0.50$). In other words, customers tipped roughly the same proportion of their bill sizes as those bill sizes increased, even though the costs of doing so increased with bill size. Thus, customers appear relatively insensitive to a broad range of costs associated with tipping.

The insensitivity to tipping costs in this study is interesting because it suggests that consumers may be less sensitive to voluntary price increases than to involuntary ones. There are three potential reasons for such a difference in price sensitivity. First, voluntary price increases are more easily avoided than are involuntary price increases, because voluntary pricing allows price-sensitive consumers to simply choose a lower payment, whereas involuntary pricing requires them to either pay the higher amount or forgo the intended purchase. Second, paying larger voluntary prices reflects or signals the givers' wealth and generosity more than does paying larger involuntary prices, and this reputational benefit may lessen the pain of paying. Finally, voluntary pricing empowers consumers both objectively and subjectively (Barone et al. 2017), and feelings of power may decrease price sensitivity by focusing consumers' attention on the acquisition of rewards and away from the prevention of pain and loss (Keltner et al. 2003, Yang et al. 2017). Of course, the current data only support insensitivity to voluntary price increases, not comparatively lower sensitivity to voluntary than to involuntary price increases. However, the former makes the latter more likely. Furthermore, other research finding that the perceived expensiveness of restaurants is affected more by menu prices than by expected tip amounts (Lynn and Wang 2013, Lynn 2017) provides additional support for the idea that consumers may be less sensitive to voluntary price increases than to involuntary ones.

5.4. Practical Implications

Our findings also have practical implications for those individuals and organizations seeking to increase tips or other voluntary contributions. In particular, our findings suggest that increasing consumer tipping (by itself) is neither a viable means of increasing customer satisfaction, repatronage, and spending nor a threat to these outcomes. As long as the means used to alter tipping do not directly affect customer satisfaction, then tips can be increased or decreased with few other consumer consequences. Furthermore, our results indicate that recommending larger tip amounts, at least within the $2–$10 or 5%–25% ranges studied here, is a safe means of increasing the amounts customers leave. Recommending larger tips increases tip revenues without long-term adverse effects on attitudes or behavior, even when

Figure 2. The Relationship Between Tip Amount and Bill Size (for the 99.25% of Transactions with Bill Size Below $300)



Note. The figure displays a local polynomial smooth of tip amount on bill size; the confidence intervals displayed are not adjusted for the presence of multiple transactions per customer.

those recommendations cause some recipients to not give in the short term.

The app-based laundry pick-up, cleaning, and delivery service we studied shares many characteristics with other app-based delivery services, such as GrubHub, Instacart, Postmates, and Amazon Flex, but is quite different from more traditional tipping contexts. For example, the current study context involves less server–customer contact and less service customization than does tipping in restaurants, bars, hair salons, and hotels. However, the greater server–customer contact and service customization in more traditional tipping contexts seems likely to decrease rather than increase customer irritation and anger with larger tip recommendations, so our conclusion that it is safe to increase the size of recommended tips should still apply to those traditional tipping contexts.

Our conclusion that increasing the tips customers leave will not lower customers' perceptions of value, satisfaction, repatronage frequency, and subsequent spending—in contrast to expectations based on the law of demand—should also apply across tipping contexts, because the costs of tipping more are the same across those contexts. Arguably, the customers in our study could be wealthier than most consumers, and this could limit the generalizability of our conclusions about sensitivity to the costs of tipping. However, other research finding that there is no quadratic trend in the positive relationship between restaurant tip amounts and bill size (Lynn and Sturman 2003) as well as the previously mentioned research finding that perceptions of restaurants' expensiveness are affected more by menu prices than by expected tip amounts (Lynn and Wang 2013, Lynn 2017) suggest that insensitivity to the costs of tipping is not limited to our sample and study context.

Things are a little different with respect to our conclusion that tipping does not enhance customer satisfaction and repatronage. Reasonable arguments can be made that the unusual features of the current study context may have diminished self-perception/justification and warm-glow processes. For example, less server–customer contact may reduce altruistic motives for tipping and, hence, warm-glow processes, whereas less service complexity/customization may reduce reward motives for tipping and, hence, self-perception/justification processes. Thus, it is possible that our failure to find tip-recommendation effects on customer attitudes and patronage speak only to the effects of tipping for app-based delivery services and that such effects would be stronger in more traditional tipping contexts. This possibility is worth testing in future research. However, if a multiplicity of motivations for tipping does undermine self-perception/justification and warm-glow processes, as discussed previously, then even research conducted in more traditional tipping contexts is unlikely to find strong tipping effects on customer attitudes and behavior because tipping is driven by many motives in most service contexts (Lynn 2015b, 2016). Positive, but weak, correlations of restaurant tip amounts with customer-service ratings and patronage frequency in the existing tipping literature (see Lynn and McCall 2000) lend credence to the generalizability of our null tipping effects on satisfaction and repatronage, because those weak correlations undoubtedly reflect satisfaction and patronage-frequency effects on tipping and provide little room for the reverse effects as well.

A final caution regarding the generalizability of our findings and their practical implications concerns labor market dynamics. As previously mentioned, our results suggest that if tips are paid out directly to workers, it is possible for firms to increase the earnings of tipped workers with no loss in revenue or profits. Firms may also want to capture some of this revenue for themselves, for example, by reducing their employee compensation in response to the worker's tip earnings. In fact, this practice has been the subject of some controversy in settings similar to the one that we study (see Houck 2019). Obviously, the current data do not speak to all the issues affecting the advisability of this tip-skimming practice. Importantly, they do not even speak to the more limited issue concerning the advisability of asking for larger tips when those larger tips are known or thought to be skimmed by firms in the form of lower wages. Our results were observed in a context in which service workers were presumed to benefit from the tips given. It remains an open question how applicable the results from this paper would be in a situation in which firms are perceived as the primary beneficiaries of larger tips, because they are believed to lower employee wages as employee tip-income increases.

5.5. Conclusions

The main finding from this paper is that larger suggested tip sizes increased the amount of tips received, while having no impact on overall customer satisfaction, repatronage, or spending. Because the tip suggestions were randomized across customers who were unaware of their participation in a study and because the sample size was large, we can have high confidence in the internal validity and the statistical-inference validity of the findings. Of course, the specific nature of the study context—an Internet app-based laundry delivery service—leaves open a question about the findings' generalizability to other contexts that should be examined in future research. Nevertheless, as discussed above, these findings provide insights to scholars interested in tipping as well as to those interested in more general value-perception, self-perception, self-justification, and warm-glow processes. They also provide insights to firms thinking about managing their employees' incomes and their customers' tipping behavior via tip recommendations and other tipping policies.

Acknowledgments
The authors thank Jordan Metzner, CEO and cofounder of Washio, for conducting the experiment reported here and for allowing us to analyze the data and publicly report the results in this paper.

Endnotes

1 Although their results were directionally consistent, Chandar et. al. (2019) found smaller effects than those reported by Haggag and Paci (2014) and speculate that defaults may play a smaller role in anonymous tipping decisions like the ones they studied than in tipping decisions visible to the service provider like the ones Haggag and Paci studied. Our study involved anonymous tipping via an app much like the study of Chandar et. al., but produced large effects more in line with size of effects reported by Haggag and Paci. This argues against anonymity weakening default effects on tipping. An alternative explanation is that default effects on tipping depend on strong tipping norms. Tipping was more common both in our study and in Haggag and Paci's study than in that of Chander et. al., suggesting that tipping norms were weaker in the Uber context.

2 Note that similar effects have been observed in charitable-donation contexts, where larger asks have been found to (i) decrease the likelihood that donations will be made, (ii) increase the size of those donations that are made, and (iii) have variable effects on total revenue generated (Weyant and Smith 1987, Doob and McLaughlin 1989, Schibrowsky and Peltier 1995, Desmet 1999, De Bruyn and Prokopec 2013).

3 Data on tipping prior to the app change are not available, so before–after comparisons of tipping are not possible. Furthermore, the company raised its prices substantially around the time it released the updated app supporting tipping and raised prices again a few months later, so the app update that supported tipping is confounded with price, which makes before–after comparisons of nontipping variables difficult to interpret. However, the randomly assigned tip recommendations are not affected by this confound, so we focus on their effects in this paper.

4 Among the missing values in the data are 214 service ratings of zero that were recoded as missing because zero was outside the scale's range, and it was unclear how those values were generated and saved or what they meant.

5 For the omnibus tests, we regress the outcome variable on a full set of indicators for each tip-suggestion condition and report the corresponding F (or χ^2) statistic for the model. This tested the null hypothesis that the mean for all of the conditions were equal to one another. In specifications containing multiple observations per customer, standard errors are adjusted for clustering at the customer level. In specifications containing only one observation per customer (such as those for patronage frequency), our omnibus test statistic is equivalent to the model statistic reported from an analysis of variance.

6 Because of differing sample sizes across outcomes and to reduce computation time, we perform the tests for multiple comparisons using several subgroupings of the dependent variables. The variables analyzed together are indicated in the notes to the corresponding tables. These analyses always account for the full set of nine paired comparisons that are the focus of our analysis.

7 Consistent with this possibility, the tendency to use a default tip and to save a tip preference were positively related ($B = 1.79$, S.E. = 0.03, $p < .001$) in a binomial logistic regression of default tip on saved preference using robust standard errors clustered within customer.

8 Because not all customers place the same number of orders, the number of customers placing order number x drops as x increases. Our sample restriction ensures that the pool of customers does not change as order number changes.

9 For customers who did not update the app, tip amounts are recorded as zero rather than missing. This limitation presents a problem for our analysis of any dynamic effects. However, it should not affect the validity of our other estimates, as there is no reason to believe that the propensity to update the app varied across the (randomly assigned) treatment groups.

10 For the first specification, the coefficient on the interaction term was 0.0564 with a standard error of 0.0199; and for the second, the coefficient was 0.0586 with a standard error of 0.0231, which is similar in magnitude to the corresponding coefficient in Table 5 using the larger sample.

11 Given the skewness of our satisfaction variable—the highest rating is given for about 74% of orders—we also confirm that the results are robust to alternative specifications. Instead of ordinary least squares, we used an ordered probit model to estimate the coefficients and standard errors for the omnibus test and naïve paired comparisons, and we find similar results. In addition, rather than using the five-point scale, we constructed an indicator for whether the rating equals five, the highest value. Again the results were similar using this measure: The omnibus test was not significant, and condition G again stood out as having slightly higher satisfaction, but those results were not significant after controlling for multiple comparisons.

12 These effects on tipping reflect only app-based tipping. Some customers may have given drivers cash tips rather than tipping through the app, and we have no way of seeing how the tip recommendations affected these cash tips. However, anecdotal evidence suggests that cash tips were relatively rare. In addition, any cash tips were probably given instead of (not in addition to) app-based tips. Thus, while the magnitude of our estimated negative effect on tip likelihood may be overstated in the presence of cash tipping, our estimated impacts on tip amount are likely to be a lower bound for the true effect. Most importantly, the possibility that tip recommendations could have affected cash tips as well as app-based tips does not challenge our finding that tip-enhancing recommendations do not affect customer satisfaction, repatronage frequency, and spending.

13 No software or formula known to us permits a better or more appropriate assessment of power because our data included clusters of unequal sizes.

14 The effects of *HvsG* and *KvsG* were not significant in pair-wise comparisons adjusting for multiple comparisons, but did have reliable (though trivial) effects on satisfaction ratings in unadjusted pair-wise comparisons. Thus, our classification of these effects as null is subject to challenge. This is a problem inherent to the application of binomial logic to continuous data, but does not detract from the value of the approach in cases with p-values farther from the alpha level.

15 Of course, it is possible that the true effects of tip recommendations on satisfaction, repatronage, and spending were hidden in our study by other, offsetting processes. For example, some tip recommendations decreased tipping likelihood while increasing the size of those tips left, and it is possible that the downstream consequences of these opposite effects on tipping canceled one another out. However, those pair-wise comparisons that produced opposite effects on tipping likelihood and tip size should have increased the variability in customer satisfaction, repatronage, and spending if tipping behavior affects these latter outcomes, and we found no such effects. Furthermore, a number of the pair-wise comparisons produced effects on tip size only (with no effects on tipping likelihood), and yet

these comparisons did not produce reliable effects on satisfaction, repatronage, or spending either. Alternatively, tipping may have both positive effects on satisfaction, repatronage, and spending through self-perception/justification and warm-glow processes and negative effects through value-perception processes, with these opposing processes and effects offsetting one another. However, this possibility requires a precise balance of tipping-induced warm-glow, self-perception/justification, and value-perception effects across many different tip-recommendation manipulations and outcome measures that is highly unlikely to occur by chance. Moreover, even in the unlikely event that such a robust balance among these processes and effects did exist, we would still conclude that tipping does not affect these outcomes, and our practical implications would remain the same.

References

Al-Ubaydli O, List JA (2013) On the generalizability of experimental results in economics: With a response to Camerer. NBER Working Paper 19666, National Bureau of Economic Research, Cambridge, MA.

Aronson E (1969) The theory of cognitive dissonance: A current perspective. *Adv. Experiment. Soc. Psych.* 4:1–34.

Azar OH (2007) The social norm of tipping: A review 1. *J. Appl. Soc. Psych.* 37(2):380–402.

Azar O (2011) Business strategy and the social norm of tipping. *J. Econom. Psych.* 32(3):515–525.

Barone MJ, Bae TJ, Qian S, d'Mello J (2017) Power and the appeal of the deal: How consumers value the control provided by Pay What You Want (PWYW) pricing. *Marketing Lett.* 28(3): 437–447.

Becker C, Bradley GT, Zantow K (2012) The underlying dimensions of tipping behavior: An exploration, confirmation, and predictive model. *Internat. J. Hospitality Management* 31(1):247–256.

Bray JP, Harris C (2006) The effect of 9-ending prices on retail sales: A quantitative UK based filed study. *J. Marketing Management* 22(5-6):601–617.

Burger JM (1999) The foot-in-the-door compliance procedure: A multiple-process analysis and review. *Personality Soc. Psych. Rev.* 3(4):303–325.

Burger JM, Caldwell DF (2003) The effects of monetary incentives and labeling on the foot-in-the-door effect: Evidence for a self-perception process. *Basic Appl. Soc. Psych.* 25(3):235–241.

Carr CL (2007) The FAIRSERV model: Consumer reactions to services based on a multidimensional evaluation of service fairness. *Decision Sci.* 38(1):107–130.

Chandar B, Gneezy U, List JA, Muir I (2019) The drivers of social preferences: Evidence from a nationwide tipping field experiment. NBER Working Paper 26380, National Bureau of Economic Research, Cambridge, MA.

Cialdini RB, Schroeder DA (1976) Increasing compliance by legitimizing paltry contributions: When even a penny helps. *J. Personality Soc. Psych.* 34(4):599–604.

Clee MA, Wicklund RA (1980) Consumer behavior and psychological reactance. *J. Consumer Res.* 6(4):389–405.

De Bruyn A, Prokopec S (2013) Opening a donor's wallet: The influence of appeal scales on likelihood and magnitude of donation. *J. Consumer Psych.* 23(4):496–502.

Desmet P (1999) Asking for less to obtain more. *J. Interactive Marketing* 13(3):55–65.

Dillard JP, Hunter JE, Burgoon M (1984) Sequential-request persuasive strategies: Meta-analysis of foot-in-the-door and door-in-the-face. *Human Comm. Res.* 10(4):461–488.

Doob AN, McLaughlin DS (1989) Ask and you shall be given: Request size and donations to a good cause 1. *J. Appl. Soc. Psych.* 19(12): 1049–1056.

Edwards JT, List JA (2014) Toward an understanding of why suggestions work in charitable fundraising: Theory and evidence from a natural field experiment. *J. Public Econom.* 114:1–13.

Estelami H (1999) The computational effect of price endings in multi-dimensional price advertising. *J. Product Brand Management* 8(3): 244–256.

Fazio RH (1987) Self-perception theory: A current perspective. Zanna MP, Olson JM, Herman CP, eds. *Social Influence: The Ontario Symposium*, Ontario Symposia on Personality and Social Psychology Series, vol. 5 (Lawrence Erlbaum Associates, Inc., Hillsdale, NJ), 129–150.

Fitzsimons GJ, Lehmann DR (2004) Reactance to recommendations: When unsolicited advice yields contrary responses. *Marketing Sci.* 23(1):82–94.

Gawronski B (2012) Back to the future of dissonance theory: Cognitive consistency as a core motive. *Soc. Cogn.* 30(6):652–668.

Giebelhausen M, Chun H, Cronin JJ, Hult T (2016) Adjusting the warm glow thermostat: How incentivizing participation in voluntary green programs moderates their impact on service satisfaction. *J. Marketing* 80(4):56–71.

Giebelhausen M, Chun H (2017) Replicating and extending our understanding of how hospitality managers can adjust the warm glow thermostat. *Cornell Hospitality Quart.* 58(2):122–133

Giebelhausen M, Lawrence B, Chun H, Hsu L (2017) The warm glow of restaurant checkout charity. *Cornell Hospitality Quart.* 58(4): 329–341.

Guido G, Peluso A (2004) Consumers' perception of odd-ending prices with the introduction of the Euro. *J. Product Brand Management* 13(3):200–210.

Haggag K, Paci G (2014) Default tips. *Amer. Econom. J. Appl. Econom.* 6(3):1–19.

Hoover H (2019) Default tip suggestions in NYC taxi cabs. Working paper, Florida State University, Tallahassee.

Houck B (2019) Instacart is revising its outrageous tipping policy, following public outcry. Accessed March 19, 2019, https://www.eater.com/2019/2/6/18214354/instacart-delivery-worker-pay-tipping-policy-doordash.

Ioannidis J (2005) Why most published research findings are false. *PLoS Med.* 2(8):1418–1422.

Keltner D, Gruenfeld DH, Anderson C (2003) Power, approach, and inhibition. *Psych. Rev.* 110(2):265–284.

List JA, Shaikh AM, Xu Y (2016) Multiple hypothesis testing in experimental economics. *Experiment. Econom.* 22(4):773–793.

Lynn WM (2006) Tipping in restaurants and around the globe: An interdisciplinary review. Altman M, ed. *Handbook of Contemporary Behavioral Economics: Foundations and Developments* (M. E. Sharpe Publishers, Armonk, NY), 626–643.

Lynn M (2015a) Service gratuities and tipping: A motivational framework. *J. Econom. Psych.* 46:74–88.

Lynn M (2015b) Explanations of service gratuities and tipping: Evidence from individual differences in tipping motivations and tendencies. *J. Behav. Experiment. Econom.* 55:65–71.

Lynn M (2016) Motivations for tipping: How they differ across more and less frequently tipped occupations. *J. Behav. Experiment. Econom.* 65:38–48.

Lynn M (2017) Should U.S. restaurants abandon tipping? A review of the issues and evidence. *Psychosociol. Issues Human Resource Management* 5(1):120–159.

Lynn M (2018) Are published techniques for increasing service-gratuities/tips effective? P-curving and R-indexing the evidence. *Internat. J. Hospitality Management* 69:65–74.

Lynn M, McCall M (2000) Gratitude and gratuity: A meta-analysis of research on the service-tipping relationship. *J. Socio-Econom.* 29(2):203–214.

Lynn M, Sturman MC (2003) It's simpler than it seems: An alternative explanation for the magnitude effect in tipping. *Internat. J. Hospitality Management* 22(1):103–110.

Lynn M, Wang S (2013) The indirect effects of tipping policies on patronage intentions through perceived expensiveness, fairness, and quality. *J. Econom. Psych.* 39:62–71.

Lynn M, Withiam G (2008) Tipping and its alternatives: Business considerations and directions for research. *J. Service Marketing* 22(4):328–336.

Lynn M, Flynn SM, Helion C (2013) Do consumers prefer round prices? Evidence from pay-what-you-want decisions and self-pumped gasoline purchases. *J. Econom. Psych.* 36:96–102.

Maniadis Z, Tufano F, List JA (2014) One swallow doesn't make a summer: New evidence on anchoring effects. *Amer. Econom. Rev.* 104(1):277–290.

Namasivayam K, Upneja A (2007) Employee preferences for tipping systems. *J. Foodservice Bus. Res.* 10(2):93–107.

Open Science Collaboration (2015) Estimating the reproducibility of psychological science. *Science* 349(6251):aac4716.

Robins RW, John OP (1997) The quest for self-insight: Theory and research on accuracy and bias in self-perception. Hogan R, Johnson J, Briggs S, eds. *Handbook of Personality Psychology* (Elsevier, Amsterdam), 649–679.

Schibrowsky JA, Peltier JW (1995) Decision frames and direct marketing offers: A field study in a fundraising context. *J. Direct Marketing* 9(1):8–16.

Schwartz Z (1997) The economics of tipping: Tips, profits and the market's demand—Supply equilibrium. *Tourism Econom.* 3(3): 265–279.

Seiter J, Brownlee G, Sanders M (2011) Persuasion by way of example: Does including gratuity guidelines on customers' checks affect restaurant tipping behavior? *J. Appl. Soc. Psych.* 41(1):150–159.

Shamir B (1983) A note on tipping and employee perceptions and attitudes. *J. Occupational Psych.* 56(3):255–259.

Shy O (2015) Do tips increase workers' income? *Management Sci.* 61(9):2041–2051.

Thomas M, Morwitz V (2005) Penny wise and pound foolish: The left-digit effect in price cognition. *J. Consumer Res.* 32(1):54–64.

Varki S, Colgate M (2001) The role of price perceptions in an integrated model of behavioral intentions. *J. Service Res.* 3(3): 232–240.

Weyant JM, Smith SL (1987) Getting more by asking for less: The effects of request size on donations of charity 1. *J. Appl. Soc. Psych.* 17(4):392–400.

Wieseke J, Kolberg A, Schons LM (2016) Life could be so easy: The convenience effect of round price endings. *J. Acad. Marketing Sci.* 44(4):474–494.

Yang W, Li Q, Guo M, Fan Q, He Y (2017) The effects of power on human behavior: The perspective of regulatory focus. *Acta Psych. Sinica* 49(3):404–415.