Are Published Techniques for Increasing Service-Gratuities/Tips Effective?

P-Curving and R-Indexing the Evidence

Michael Lynn*

School of Hotel Administration

Cornell University

***International Journal of Hospitality Management,* 69(2018),65-74.**

* Direct correspondence regarding this manuscript to Mike Lynn, 552 Statler Hall, Cornell University, Ithaca, NY 14853-6902, wml3@cornell.edu, (607)255-8271.

Abstract

Recently developed statistical tools are used to assess the evidential value and replicability of the published experimental literature on ways to increase tips. Significantly right-skewed full and half p-curves indicate that the literature is more than a collection of Type 1 errors – it provides evidence of real effects. Moreover, those real effects are scattered across both replicated and non-replicated effects as well as across the work of each of the major contributors to this literature. An overall r-index of .55 indicates that over half of the reported effects would likely be replicated if the studies were repeated. More research is need to ascertain the reliability of specific effects – especially those reported by Gueguen, because lower power makes his effects less replicable than others in the literature. Nevertheless, readers can be reasonably confident that most of the techniques for increasing tips in this literature will work.

Key words: tip income, field experiments, restaurants

Are Published Techniques for Increasing Service-Gratuities/Tips Effective? P-Curving and R-Indexing the Evidence

1. Introduction

Consumers around the world often give voluntary sums of money (called "tips") to the hospitality, tourism and other service workers who have served them. The sizes of these individual payments vary across consumers and contexts, but often amount to 10 percent or more of the costs of the services being tipped (Lynn and Lynn, 2004) and can reach thousands of dollars (Denison, 2016). There are no good records of aggregate tips within or across nations, but Azar (2011) estimates that approximately $45 billion per year is tipped to U.S. waiters and waitresses alone, so worldwide tipping totals across all services are undoubtedly enormous.

Tipping impacts consumers' wallets (Lynn, 2017) and dining experiences (Lynn, in press), but arguably its biggest impact is on service workers' incomes. For example, Payscale (2015) reports that the percentage of total income from tips is 12% for baristas, 16% for restaurant hosts/hostesses, 24% for busboys/girls, 31% for bartenders, 42% for banquet captains, and 62% for waiters/waitresses. Given this impact on service workers' incomes, many scholars across diverse academic disciplines have conducted field experiments and quasi-experiments testing ways that servers or their managers can increase the tips consumers leave. They have found that servers can earn larger tips if they:

- use makeup (for waitresses)(Gueguen and Jacob, 2011; Jacob, et. al, 2009),
- make their hair blond (for waitresses) Gueguen, 2012; Jiang and Galm, 2014) ,
- wear something unusual in their hair (for waitresses) (Jacob, Gueguen and Delfosse, 2012; Stillman and Hensley, 1980),

- wear red shirts or lipstick (for waitresses) (Gueguen & Jacob, 2012, 2014),
- introduce themselves by name (Garrity and Degelman, 1990),
- use customers' names (Adams and Pettijohn, 2016; Seiter, Givens and Weger, 2016; Seiter and Weger, 2013),
- squat next to or sit down at the table (Davis, et. al., 1998; Leodoro and Lynn, 2007; Lynn and Mynier, 1993),
- stand physically close to customers (Jacob and Guguen, 2012),
- touch customers (Crusco and Wetzel, 1984; Gueguen and Jacob, 2005; Hornik, 1992; Hubbard, et.al., 2003; Lynn, Le and Sherwyn, 1993),
- smile (Tidd and Lockard, 1978),
- compliment customers (Seiter, 2007; Seiter and Dutson, 2007; Seiter and Weger, 2010) ,
- mimic customers' verbal behavior (vanBaaren, et. al., 2003; Jacob and Gueguen, 2013),
- entertain guests with puzzles or jokes (Gueguen, 2002; Rind and Strohmetz, 2001b),
- forecast good weather to customers (Rind, 1996),
- write various messages or draw various pictures on the check (Gueguen and Logeherel, 2000; Jacob, et. al., 2013; Rind and Bordia, 1995, 1996; Seiter and Gass, 2005)),
- use tip trays with credit cardo logos on them (McCall and Belmont, 1996), and
- give customers free candies (Strohmetz, et. al, 2002).

These studies have been reviewed in a free e-book for servers, called *MegaTips 2: Scientifically Tested Ways to Increase Your Tips* (Lynn, 2011), and have received a lot of attention in the press and online (e.g., Gillman, 2017; Shin, 2014).

Unfortunately, the social sciences are undergoing a replication crisis that calls into question the reliability of this published tipping literature. Scholars have discovered that questionable research practices (collectively called "p-hacking") combine with a bias against publishing null results to make Type 1 errors far more common than typically believed (Simmons, Nelson and Simonsohn, 2011; Sterling, Rosenbaum and Weinkam, 1995). In fact, even frequently studied and apparently well-established phenomena such as the effects of ego depletion (Carter and McCullough, 2014), power posing (Simmons and Simonsohn, 2017), and money priming (Vadillo, Hardwicke and Shanks, 2016) appear to be smaller and less reliable than assumed. The effects of different tip enhancing techniques, which are less frequently studied and replicated, may also be unreliable. This possibility seems particularly plausible given the subtle and/or transitory nature of many of the purported tip enhancing behaviors and the fact that tipping has been found to be only weakly related to service quality (Lynn and McCall, 2000). Further support for skepticism regarding at least one of these tip enhancing techniques comes from Lynn, et.al's (2017) recently reported failure to conceptually replicate Gueguen and Jacob's (2014) finding that waitresses received more and larger tips when they wore red shirts than when their shirts were another color. Given these reasons for skepticism, the collective body of field experiments and quasi-experiments on ways service workers and managers can increase consumer tipping is evaluated in the paper below. Recently developed statistical tools -- called p-curves and the replicability-index – are used to assess the evidential value and replicability of the published literature on ways to increase tips.

2. Overview of P-Curve and R-Index Analyses

2.1. P-curve analysis

P-curve analysis uses the frequency distribution of p-values (aka, p-curve) associated with significant effects in a set of studies to assess the likelihood that selective reporting is the sole explanation for the set of effects (Simonsohn, Nelson and Simmons, 2014;Simonsohn, Simmons and Nelson, 2015). If selective reporting cannot explain a set of significant effects, that set is said to contain evidential value. Evidential value can be inferred from the shape of the p-curve as described below.

A uniform frequency distribution of significant p-values is indicative of no evidential value -- in other words, the significant effects are likely to be selectively reported Type 1 errors. This follows from the fact that p-values reflect the probability of getting a result that large or larger by chance alone. When the null hypothesis is true, there is a 5 percent chance of getting a p-value of .05 , a 4 percent chance of getting a p-value of .04, a 3 percent chance of getting a p-value of .03, etc… This means that p-values between .05 and .04 should occur 1 percent of the time, as should p-values between .04 and .03 as well as p-values between .03 and .02. In other words, when the null hypothesis is true, significant p-values should be uniformly distributed across values less than .05, .04, .03, .02 and.01. Thus, a uniform frequency distribution of significant p-values is expected when the null hypothesis is true and such a distribution provides no evidence of a real effect.

A right-skewed frequency distribution of significant p-values (in other words, with more p-values of .01 than .05) is indicative of some evidential value – of a true effect underlying at least some of the findings in the set. The shape of p-curves is a function only of effect size and

sample size. For any true non-zero effects, p-curves associated with unbiased tests of those effects are right-skewed with the amount of skewness increasing with true effect sizes and/or sample sizes. Thus, right-skewed p-curves are consistent with expectations when non-zero effects are tested and such a distribution indicates that at least some of the tested effects are real.

A left-skewed frequency distribution of significant p-values (in other words, with more p-values of .05 than.01) is indicative of intense p-hacking. P-hacking is the use of questionable research practices (such as making post-hoc decisions about how much data to collect and which data points and/or measures to retain in the analyses) to produce "significant effects" and increase the odds of publication. P-hacking tends to produce left-skewed frequency distributions of p-values because the p-hacking effort required to produce significant tests when true effects are zero increases exponentially as the alpha-level (or target p-value) decreases and alpha/p-levels of .05 are typically sufficient to get published. Thus, left-skewed p-curves are consistent with expectations when p-hacking is responsible for a set of significant effects and such a distribution indicates that the set of findings provide no evidence of a real effect.

P-curve analysts assess the statistical significance of a p-curve's right skewness using two tests.[1] First, for each p-value less than .05 (the "full p-curve"), they compute the probability (called the "pp-value") of getting a value that extreme conditioned on having a value of at least .05 and convert those pp-values to z-scores, which are then combined using Stouffer's method. This is the most powerful test of evidential value, but it is potentially biased by ambitious p-

[1] If a p-curve is not reliably right skewed, p-curve analysts also typically test whether the p-curve is significantly flatter than expected if the studies had a power of 33 percent. This test is a slightly more complicated analog to the right-skewness test and is used to assess whether or not a set of studies contains sufficient information to assess evidential value when the p-curve is not significantly right skewed. If this flatness test is significant, then the set of studies does contain sufficient information to assess evidential value and the absence of right-skewness is meaningful – i.e., indicative of no evidential value in the studies. If it is not significant, then the set of studies does not contain enough information to judge whether it does or does not contain evidential value.

hacking that targets alpha-levels below .05. To address this concern, a second test is typically done to see if the p-values less than .025 (the "half p-curve") are right skewed. For each of these p-values, the probability of getting a value that extreme conditioned on having a value of at least .025 is calculated and those pp-values are converted to z-scores, which are then combined using Stouffer's method. This latter test uses less information and has less power than the full p-curve test, but is also less biased by ambitious p-hacking. These two tests are used together to get both of the benefits that each provides – to get both statistical power and resistance to ambitious p-hacking. If the half p-curve's skewness is reliable at the .05 level or both the full and half p-curves' skewness is reliable at .10 level, then a set of studies is judged to contain evidential value.

2.2. R-index analysis

R-index analysis assesses the replicability of a set of studies by comparing the proportion of reported results that are statistically significant (called the "success rate") with expectations given the median observed (or post-hoc) power of the studies (Schimmack, 2016). Specifically, the replicability-index is calculated using the following formula:

R-Index = Median Observed Power – (Success Rate – Median Observed Power)

This index does not reflect the average probability of replication for a set of studies (a R-Index of .22 does not imply an average replicability of 22 percent), but it is monotonically related to the average probability of replication in the set of studies. Thus, comparisons of R-indices across authors, journals, schools, etc… will reflect their rankings in terms of replicability. Furthermore, the R-index over-estimates average true power that is less than 50 percent and under-estimates average true power that is more than 50 percent, so an R-index below .50 indicates that the

average replicability of the set of studies is below 50 percent and an R-index above .50 indicates that the average replicability of the set of studies exceeds 50 percent.

3. Identification of Studies and Effects to be Analyzed

This paper presents p-curve and r-index analyses of the body of published field experiments and quasi-experiments testing viable ways service workers and their employers can increase tip amounts. Relevant studies were identified through (i) personal knowledge of the literature, (ii) perusing Michael Lynn's online bibliography of tipping research and then checking the abstracts or complete papers of those listed articles whose titles suggested they may report field experiments, and (iii) searching Google Scholar using the terms "tipping experiment restaurant OR bar OR hair OR taxi OR service." In addition, an early draft of the paper was sent to several of the prominent contributors to this literature for comments and they identified two overlooked studies. All identified field studies that manipulated viable ways service workers and their employers can increase tip amounts and that observed the effects of those manipulations on actual tipping behavior were retained for analysis. Purely correlational studies and scenario based experiments (e.g., Grandey, et. al., 2005; Shih, Jai and Blum, 2016) were excluded as outside the domain of inquiry. Also excluded were five studies testing the effects on tipping of manipulations that the current author judged to be impractical as tip enhancing techniques. Those five studies involved giving customers excess change (Azar, Yosef and Bar-Eli, 2015), interrupting and delaying service (Fromkin, Goldstein and Brock, 1977), instructing customers to empathize with the service provider (Davis, Jiang, Williams, Drolet and Gibbs, 2017), having customers see others tip (Gueguen, 2007), and adopting an IT theft monitoring system (Pierce, Snow and McAfee, 2015). One final study was omitted because needed information about the statistical test of the relevant effect was misreported (Bernritter, Ooijen and Muller, 2017).

Statistical tests of main effects or simple main effects of the potential tip enhancing actions from each included study were recorded and analyzed.[2] In most cases, main effects were used. However, if main effects were not available or if they were qualified by significant cross-over interactions, then simple main effects were used.[3] Multiple simple main effects from a single study were all included in the analyses if they came from separate sub-samples (e.g., men and women). For studies with more than one treatment condition, overall F-tests with greater than 1 degree of freedom in the numerator were used when available. If appropriate overall F-tests were not available, but independent contrasts were, then the contrast with the largest sample size and involving a no-treatment control were used in the current analyses.

Often more than one tipping measure was used in the original study. If available, analyses of tips (including tips of zero) as a percentage of the bill were recorded and used in the current analyses. If authors reported analyses of the decision to tip or not and then analyses of tip amounts (in percentages or currency) among those leaving a tip, then the former analyses, which involved the larger samples, were recorded and used in the current analyses.

---

[2] The only exception was a study by Seligman, et. al. (1985) which manipulated explanations for naturally occurring slow and fast service. In that case, the test of the interaction of service speed and explanation type was recorded and used this analysis. This interaction reflected the effects of self-serving attributions -- making internal (vs external) attributions for good service and external (vs internal) attributions for bad service -- so amounted to a coherent technique for enhancing tips.

[3] Simonsohn, Nelson and Simmons (2014) argue that simple main effects from studies examining the attenuation of an effect should not be included in p-curve analyses. The reason is that p-values for main effects are typically smaller than those for interactions and publication bias favoring significant interactions will censor some significant simple main effects (namely, those associated with non-significant interactions) from the published records making the distribution of p-values for those simple main effects non-uniform. In other words, simple main effects from attenuated interactions bias the p-curve analysis in favor of finding evidential value. However, in the mostly applied literature studied here, significant interactions are not the primary focus of investigation even when they are tested. The real focus of these papers is demonstrating ways to increase tips and showing one significant simple main effect is usually sufficient to get published. Thus, in this literature, where non-significant interactions are not a barrier to publication, testing the attenuation of an effect does not imply that the p-curve for associated simple main effects is non-uniform. Therefore, simple main effects from attenuated interactions were included in the current p-curve analysis. However, the p-curve results reported in the main text are robust with respect to this decision. All simple main effects from attenuated interactions in the analysis came from Gueguen's work and omitting these simple main effects from the p-curve analysis did not substantially change the results for his body of work – without simple main effects, his full p-curve $z = -2.85$, $p < .003$ and half p-curve $z = -1.89$, $p < .03$.

The studies and statistical-tests used in the current analyses are summarized in Appendix A. P-curve analyses were conducted using the online app (version 4.0) at http://www.p-curve.com/. R-index analyses were conducted using the R-index Calculation Spreadsheet available online at http://www.r-index.org/.

4. Results and Discussion

P-curve and r-index analyses of the collective body of field experiments and quasi-experiments examining way to increase tips are summarized in Table 1 and discussed below.

4.1. Evidential value and replicability of the overall literature

The frequency distribution of significant p-values in the experimental literature on ways to increase tips is presented in Figure 1. The full p-curve is reliably right-skewed ($z = -12.27$, $p < .0001$) as is the half p-curve ($z = -12.24$, $p < .0001$), so this literature does contain evidential value – it is not merely a collection of Type 1 errors.

The proportion of all reported effects that were significant (aka, the "success rate") was .83 while the median observed power was .69, so the percentage of significant effects exceeds expectations based on estimated power by 14 percent in this analysis. This "inflation rate," which is reliably different from zero at the .02 level, suggests that here is some publication bias in this literature, though it could also be due to fraud and/or p-hacking. The r-index is .55, which suggests that roughly half the effects would be replicated if the studies were repeated. This is not to say that the other (likely non-replicated) half of the reported effects are Type 1 errors. With a median observed power of only .69, the probability of making a Type 2 error is around 30 percent and that probability is even higher if the true (non-zero) effect sizes are smaller than the

effect sizes observed in the studies to be replicated. Thus, the rate of Type 1 errors in the literature is likely to be considerably lower than the estimated rate of failures to replicate.

4.2. Evidential value and replicability of replicated and non-replicated effects

The literature analyzed above includes some effects that have been replicated. Specifically, there is more than one study finding significant effects of waitress' blond hair, waitress' hair decorations, server's mimicry of customer, server's touching of customer, server's squatting beside or sitting at the table, server's complimenting customer, server's giving customer candy, servers' communications about current or future weather conditions, and server's presentation of the bill on a tip tray with credit card logos. Since it seems more likely that replicated effects are real than that non-replicated ones are, the replicated and non-replicated effects in the literature were analyzed separately (see Table 1).

Both replicated and non-replicated effects had significantly right-skewed full and half p-curves (all $|z|$'s < 7.00, p's $< .0001$), so both sets of effects contain evidential value. As expected, the r-index for replicated effects is larger than that for non-replicated effects (r-indices = .61 and .51 respectively), but half or more of the effects in both sets of studies should be replicable if the studies were repeated.

4.3. Evidential value and replicability of major contributors' work

Over half of the literature on ways to increase tips can be attributed to just four authors (and their co-authors) – i.e., Nicholas Gueguen (often with Celine Jacob), Michael Lynn, Bruce Rind (often with David Strohmetz), and John Seiter (often with Harry Weger). The effects from each of these authors and all others combined were analyzed separately to see if their work differed in evidential value and/or replicability (see Table 1). Each of the major contributors and

the collective set of other contributors all had significantly right-skewed full and half p-curves (all |z|'s < 2.15, p's < .02), so their work contains evidential value. However, the replicability of Gueguen's work is notably lower than that for all other authors (r-index = .37 vs .77, .64, .61 and .61 respectively). This low r-index, which indicates that less than half of his effects would likely replicate if the studies were repeated, is due to both a low observed power of .60 and a high inflation rate of .23.[4] However, this does NOT imply questionable research practices on Gueguen's part. His median observed power is lower than that of others despite having a substantially larger median sample size than do the other authors (median n = 289 vs. 116, 70, 107 and 108 respectively), because he uses non-parametric tests with a binomial primary dependent variable (tip vs no tip) whereas others use parametric tests with a continuous primary dependent variable (dollar or percentage tip amounts).[5] Gueguen's choices of dependent measures and tests are the best available options given tipping norms in his home country of France, so achieving high power is more difficult and costly for him than for others studying in this area. Gueguen's unusually high inflation rate (excess success given median observed power) is mostly due to his low power as his success rate is not abnormally high for this literature. Low power means that he is more likely than others to make Type 2 errors and (with a similar success rate in published effects as others) to have more null results that are never published. Although Gueguen's low r-index does not imply questionable research practices, it does indicate that his

[4] Rind also has a low power (of .62). However, his success rate is also low, so his inflation rate is basically zero giving him a more or less average r-index for this literature. These results suggest that Rind is particularly successful at publishing results that fall short of conventional levels of statistical significance.

[5] The r-index analyses for Gueguen in Table 2 and the text included main effects across sex or simple main effects for both men and women from four studies testing tip enhancing tactics that were primarily expected to affect men, so whose effects on women and across sex were expected to be small. Including these small effects in the analysis may have reduced Gueguen's median observed power and, ultimately his r-index. To test this possibility, the r-index analyses were repeated with the effects from these four studies coming only from men – using simple main effects for men to represent these studies' results. This analysis produced results only slightly better than those in the main text – success rate =.94, median observed power = .67, inflation rate = .27, and r-index = .41.

effects more than those of other scholars in this area need to be replicated to insure their reliability.

5. Summary and Conclusions

This paper assesses the evidential value and replicability of an applied literature. The studies in this literature are almost always justified as tests of ways servers can earn larger tips - not as tests of theory. In fact, the processes underlying these effects are almost always speculative and rarely (if ever) measured or tested. Even if the studies were clear tests of underlying theoretical processes, this is a heterogeneous literature that consists of many different effects undoubtedly driven by many different processes. Thus, the current analyses do not change our confidence in any particular theory or theoretical process. Rather, their value lies in increasing our confidence in the practical usefulness of the entire literature social scientific literature on tip enhancing techniques.

Given known publication bias and the replication crisis in the social sciences (Simmons, Nelson and Simonsohn, 2011; Sterling, Rosenbaum and Weinkam, 1995), it was entirely plausible that most if not all of this literature consisted of Type 1 errors and provided no useful insight into ways to increase tips. Indeed, it was a fear of this possibility that drove the author to conduct the current analyses. Fortunately, the results of these analyses have allayed those fears. Significantly right-skewed full and half p-curves indicate that the literature is more than a collection of Type 1 errors – it provides evidence of real effects. Moreover, those real effects are scattered across both replicated and non-replicated effects as well as across the work of each of the major contributors to this literature. An overall r-index of .55 indicates that over half of the reported effects would likely be replicated if the studies were repeated and a median observed power of .69 indicates that at least two-thirds of the unreplicated effects would be false negatives

(or Type 2 errors). More research is need to ascertain the reliability of specific effects – especially those reported by Gueguen, because lower power makes his effects less replicable than others in the literature. Nevertheless, servers and their managers can rely upon the effectiveness of most (and possibly all) of the tip enhancing techniques in the literature to actually work as described. Thus, servers should employee these techniques to increase their tips and managers should encourage and facilitate the use of these techniques to increase their servers' tips (Lynn, 2005).

On the whole, the social science literature on ways to increase tips was revealed by the current analyses to describe real effects and, thus, to have practical value. Nevertheless, only a little more than half of the studies' results would be replicated if the studies were repeated. This low replication rate does not mean the original effects are false positives (or Type 1 errors), because the vast majority of those failures to replicate would be false negatives (Type 2 errors). However, it does indicate that the studies are under-powered, and, thereby highlights the need for researchers to use larger sample sizes. Of course, the current analyses apply only to the field experiment literature on ways to increase tips, but the low power and replicability in this literature is typical for neuroscience and psychology articles (see Button, et. al., 2013; Szucs and Ioannidis, 2017) and probably for hospitality articles too. Hopefully, this paper will encourage hospitality researchers to increase their sample sizes and statistical power.

The current analyses also indicate that the overall success rate in the field experiment literature on ways to increase tips is reliably larger than it should be given the median observed power of the studies in it. This suggests that the success rate has been artificially elevated thru p-hacking and/or publication bias and, thereby, highlights the need for researchers to be more forthcoming about unreported variables and analyses in their published studies as well as about

unpublished studies failing to find expected results. Reporting all the variables collected, analyses run, and studies conducted does come at costs of journal pages and readers' time and effort, but failure to report these things and the null results associated with them makes effects seem larger and more reliable than they actually are. Single-paper meta-analyses (see McShane and Bockenholt, 2017) represent one way to more completely and efficiently report research methods, analyses and results. Whether they use this particular tool or not, hospitality scholars should find ways to completely yet efficiently report all their findings – including null results – and hospitality reviewers and editors should reward them for doing so.

References

Adams, V. P., & Pettijohn, T. F., II. (2016). Using a customer's name to personalize checks and increase restaurant tipping behavior. *International Journal of Management Sciences, 6*(10), 491-497.

Azar, O.H. (2011). Business strategy and the social norm of tipping. *Journal of Economic Psychology,* 32, 515-525.

Azar, O. H., Yosef, S., & Bar-Eli, M. (2015). Restaurant tipping in a field experiment: How do customers tip when they receive too much change?. *Journal of Economic Psychology, 50,* 13-21.

Bernritter, S. F., van Ooijen, I., & Müller, B. C. N. (2017). Self-persuasion as marketing technique: The role of consumers' involvement. *European Journal of Marketing, 51*(5/6), 1075-1090.

Button, K.S., Ioannidis, J.P.A., Mokrysz, C., Nosek, B.A., Flint, J. Robinson, E.S.J., & Munafo, M.R. (2013). Power failure: Why small sample size undermines the relaiability of neuroscience. *Nature Reviews Neuroscience*, 14, 365-376.

Carter, E.C. & McCullough, M.E. (2014). Publication bias and the limited strength model of self-control: Has the evidence for ego depletion been overestimated? *Frontiers in Psychology*, 5, doi:10.3389/fpsyg.2014.00823.

Crusco, A. H., & Wetzel, C. G. (1984). The Midas touch: The effects of interpersonal touch on restaurant tipping. *Personality and Social Psychology Bulletin, 10*(4), 512-517.

Davis, C., Jiang, L., Williams, P., Drolet, A. & Gibbs, B.J. (2017). Predisposing customers to be more satisfied by inducing empathy in them. *Cornell Hospitality Quarterly*, 58 (3), 229-239.

Davis, S. F., Schrader, B., Richardson, T. R., Kring, J. P., & Kieffer, J. C. (1998). Restaurant servers influence tipping behavior. *Psychological Reports, 83*, 223-226.

Denison, R. (2016). Five celebrity tippers that will make you want to clock in. http://www.snagajob.com/resources/5-celebrity-tips-that-will-make-you-want-to-clock-in/, accessed July 19, 2017.

Fromkin, H. L., Goldstein, J. H., & Brock, T. C. (1977). The role of "irrelevant" derogation in vicarious aggression catharsis: A field experiment. *Journal of Experimental Social Psychology, 13*(3), 239-252.

Garrity, K., & Degelman, D. (1990). Effect of server introduction on restaurant tipping. *Journal of Applied Social Psychology, 20*(2), 168-172.

Gillman, S. (2017). Twenty ways to increase your tip income. https://www.frugalforless.com/increase-your-tip-income/, accessed July 19, 2017.

Grandey, A. A., Fisk, G. M., Mattila, A. S., Jansen, K. J., & Sideman, L. A. (2005). Is "service with a smile" enough? Authenticity of positive displays during service encounters. *Organizational Behavior and Human Decision Processes, 96*, 38-55.

Guéguen, N. (2002). The effects of a joke on tipping when it is delivered at the same time as the bill. *Journal of Applied Social Psychology, 32*(9), 1955-1963.

Guéguen, N. (2007). The effect of modeling on tipping behavior. *Studia Psychologica, 49*(3), 275-282.

Guéguen, N. (2012). Hair color and wages: Waitresses with blond hair have more fun. *The Journal of Socio-Economics, 41,* 370-372.

Guéguen, N. (2013). Helping with all your heart: The effect of cardioid dishes on tipping behavior. *Journal of Applied Social Psychology, 43,* 1745-1749.

Guéguen, N., & Jacob, C. (2005). The effect of touch on tipping: An evaluation in a French bar. *Hospitality Management, 24,* 295-299.

Guéguen, N., & Jacob, C. (2011). Enhanced female attractiveness with use of cosmetics and male tipping behaviors in restaurants. *Journal of Cosmetic Science, 62*(3), 283-290.

Guéguen, N., & Jacob, C. (2012). Lipstick and tipping behavior: When red lipstick enhance waitresses tips. *International Journal of Hospitality Management, 31,* 1333-1335.

Guéguen, N., & Jacob, C. (2014). Clothing color and tipping: Gentleman patrons give more tips to waitresses with red clothes. *Journal of Hospitality & Tourism Research, 38*(2), 275-280.

Guéguen, N., & Legoherel, P. (2000). Effect on tipping of barman drawing a sun on the bottom of customers' checks. *Psychological Reports, 87*, 223-226.

Hornik, J. (1992). Tactile stimulation and consumer response. *Journal of Consumer Research, Inc., 19*, 449-458.

Hubbard, A. S. E., Tsuji, A. A., Williams, C., & Seatriz, V., Jr. (2003). Effects of touch on gratuities received in same-gender and cross-gender dyads. *Journal of Applied Social Psychology, 33*(11), 2427-2438.

Hülsheger, U. R., Lang, J. W. B., Schewe, A. F., & Zijlstra, F. R. H. (2015). When regulating emotions at work pays off: A diary and an intervention study on emotion regulation and customer tips in service jobs. *Journal of Applied Psychology, 100*(2), 263-277.

Jacob, C., & Guéguen, N. (2012). The effect of physical distance between patrons and servers on tipping. *Journal of Hospitality & Tourism Research, 36*(1), 25-31.

Jacob, C., & Guéguen, N. (2013). The effect of employees' verbal mimicry on tipping. *International Journal of Hospitality Management, 35,* 109-111.

Jacob, C., & Guéguen, N. (2014). The effect of employees' clothing appearance on tipping. *Journal of Foodservice Business Research, 17*, 483-486.

Jacob, C., Guéguen, N., Ardiccioni, R., & Sénémeaud, C. (2013). Exposure to altruism quotes and tipping behavior in a restaurant. *International Journal of Hospitality Management, 32*, 299-301.

Jacob, C., Guéguen, N., & Boulbry, G. (2010). Effects of songs with prosocial lyrics on tipping behavior in a restaurant. *International Journal of Hospitality Management, 29,* 761-763.

Jacob, C., Guéguen, N., Boulbry, G., & Ardiccioni, R. (2009). Waitresses' facial cosmetics and tipping: A field experiment. *International Journal of Hospitality Management, 29,* 188-190.

Jacob, C., Guéguen, N., & Delfosse, C. (2012). She wore something in her hair: The effect of ornamentation on tipping. *Journal of Hospitality Marketing & Management, 21,* 414-420.

Jewell, C. N. (2008). Factors influencing tipping behavior in a restaurant. *Psi Chi Journal of Undergraduate Research, 13*(1), 38-48.

Jiang, C., & Galm, M. (2014). The economic benefit of being blonde: A study of waitress tip earnings based on their hair color in a prominent restaurant chain. *Journal of Behavioral Studies in Business, 7,* 1-6.

Karagiorgakis, A., & Malone, D. (2014). The effect of clothing and method of payment on tipping in a bar setting. *North American Journal of Psychology, 16*(3), 441-451.

Khadjavi, M. (2016). Indirect reciprocity and charitable giving—evidence from a field experiment. *Management Science, Articles in Advance,* 1-11. https://doi.org/10.1287/mnsc.2016.2519

Lee, N. Y., Noble, S. M., & Biswas, D. (2015). Hey big spender! A golden (color) atmospheric effect on tipping behavior. *Journal of the Academy of Marketing Science,* 1-21. https://doi.org/10.1007/s11747-016-0508-3

Leodoro, G., & Lynn, M. (2007). The effect of server posture on the tips of whites and blacks. *Journal of Applied Social Psychology, 37*(2), 201-209.

Lynn, M. (in press). The effects of tipping on consumers' satisfaction with restaurants. *Journal of Consumer Affairs,* forthcoming.

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

Lynn, M. (2011). MegaTips 2: Twenty Tested Techniques to Increase Your Tips. CHR Tool, Vol. 2, No. 1.

Lynn, M. (2005). Increasing servers' tips: What managers can do and why they should do it. *Journal of Foodservice Business Research*, 8 (4), 89-98.

Lynn, M., Giebelhausen, M. Garcia, S., Li, Y. & Patumanon, I. (2016). Clothing color and tipping: An attempted replication and extension. *Journal of Hospitality and Tourism Research*, 40 (4), 516-524.

Lynn, M., & Gregor, R. (2001). Tipping and service: The case of hotel bellmen. *International Journal of Hospitality Management, 20*, 299-303.

Lynn, M., Le, J. M., & Sherwyn, D. S. (1998). Reach out and touch your customers. *Cornell Hotel and Restaurant Administration Quarterly, 39*(3), *60*-65.

Lynn, M. & Lynn, A. (2004). National values and tipping customs: A replication and extension. *Journal of hospitality and Tourism Research*, 28,356-364.

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

Lynn, M., & Mynier, K. (1993). Effect of server posture on restaurant tipping. *Journal of Applied Social Psychology, 23*(8), 678-685.

McCall, M., & Belmont, H. J. (1996). Credit card insignia and restaurant tipping: Evidence for an associative link. *Journal of Applied Psychology*, *81*(5), 609-613.

McShane, B.B. & Bockenholt, U. (2017). Single-paper meta-analysis: Benefits for study summary, theory testing, and replicability. *Journal of Consumer Research*, 43, 1048-1063.

Payscale (2015). Restaurant Report Key Stats, http://www.payscale.com/data-packages/restaurant-report/full-data, accessed July 19, 2017.

Pierce, L., Snow, D. C., & McAfee, A. (2015). Cleaning house: The impact of information technology monitoring on employee theft and productivity. *Management Science, 61*(10), 2299-2319.

Rind, B. (1996). Effect of beliefs about weather conditions on tipping. *Journal of Applied Social Psychology*, *26*(2), 137-147.

Rind, B., & Bordia, P. (1995). Effect of server's "thank you" and personalization on restaurant tipping. *Journal of Applied Social Psychology*, *25*(9), 745-751.

Rind, B., & Bordia, P. (1996). Effect on restaurant tipping of male and female servers drawing a happy, smiling face on the backs of customers' checks. *Journal of Applied Social Psychology, 26*(3), 218-225.

Rind, B., & Strohmetz, D. (1999). Effect on restaurant tipping of a helpful message written on the back of customers' checks. *Journal of Applied Social Psychology*, *29*(1), 139-144.

Rind, B., & Strohmetz, D. (2001a). Effect of beliefs about weather conditions on restaurant tipping. *Journal of Applied Social Psychology*, *31*(10), 2160-2164.

Rind, B., & Strohmetz, D. (2001b). Effect on restaurant tipping of presenting customers with an interesting task and of reciprocity. *Journal of Applied Social Psychology, 31*(7), 1379-1384.

Schimmack, U. (2016). The replicability-index: Quantifying statistical research integrity. https://wordpress.com/post/replication-index.wordpress.com/920

Seiter, J. S. (2007). Ingratiation and gratuity: The effect of complimenting customers on tipping behavior in restaurants. *Journal of Applied Social Psychology, 37*(3), 478-485.

Seiter, J. S., Brownlee, G. M., & Sanders, M. (2011). Persuasion by way of example: Does including gratuity guidelines on customers' checks affect restaurant tipping behavior. *Journal of Applied Social Psychology, 41*(1), 150-159.

Seiter, J. S., & Dutson, E. (2007). The effect of compliments on tipping behavior in hairstyling salons. *Journal of Applied Social Psychology, 37*(9), 1999-2007.

Seiter, J. S., & Gass, R. H. (2005). The effect of patriotic messages on restaurant tipping. *Journal of Applied Social Psychology, 35*(6), 1197-1205.

Seiter, J. S., Givens, K. D. & Weger, H. Jr. (2016). The effect of mutual introductions and addressing customers by name on tipping behavior in restaurants. *Journal of Hospitality Marketing & Management, 25,* 640-651.

Seiter, J. S., & Weger, H., Jr. (2010). The effect of generalized compliments, sex of server, and size of dining party on tipping behavior in restaurants. *Journal of Applied Social Psychology, 40*(1), 1-12.

Seiter, J. S., & Weger, H., Jr. (2013). Does a customer by any other name tip the same?: The effect of forms of address and customers' age on gratuities given to food servers in the United States. *Journal of Applied Social Psychology, 43*, 1592-1598.

Seiter, J. S., & Weger, H., Jr. (2014). The principle of reciprocity in hospitality contexts: The relationship between tipping behavior and food servers' approaches to handling leftovers. *Journal of Hospitality & Tourism Research,* 1-8. https://doi.org/10.1177/1096348014562425

Seligman, C., Finegan, J. E., Hazelwood, J. D., & Wilkinson, M. (1985). Manipulating attributions for profit: A field test of the effects of attributions on behavior. *Social Cognition, 3*(3), 313-321.

Shih, I. H., Jai, T. M., & Blum, S. C. (2016). Tipping: The case of hotel turndown service. *International Journal of Hospitality Management, 55,* 129-131.

Shin, L. (2014). The ultimate guide to living on tips, Part 1: How to earn more in tips, https://www.forbes.com/sites/laurashin/2014/04/30/the-ultimate-guide-to-living-on-tips-part-1-how-to-discreetly-increase-your-tips/#562855837d02, accessed July 19, 2017.

Simmons, J.P. & Simonsohn, U. (2017). Power posing: P-curving the evidence. *Psychological Science*, 28(5), 687-693.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological science*, *22*(11), 1359-1366.

Simonsohn, U., Nelson, L.D. & Simmons, J.P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*, 143, 535-547.

Simonsohn, U., Simmons, J.P. & Nelson, L.D. (2015). Better p-curves: Making p-curve analysis more robust to errors, fraud, and ambitious p-hacking, A reply to Ulrich and Miller (2015). *Journal of Experimental Psychology: General*, 144, 1146-1152.

Stephen, R., & Zweigenhaft, R. L. (1986). The effect on tipping of a waitress touching male and female customers. *The Journal of Social Psychology*, *126*(1), 141-142.

Sterling, T. D., Rosenbaum, W. L., & Weinkam, J. J. (1995). Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa. *The American Statistician*, *49*(1), 108-112.

Stillman, J. W., & Hensely, W. E. (1980). She wore a flower in her hair: The effect of ornamentation on nonverbal communication. *Journal of Applied Communications Research*, *8*(1), 31-39.

Strohmetz, D. B., & Rind, B. (2001). The impact of tipping recommendations on tip levels. *Cornell Hotel and Restaurant Administration Quarterly*, *42*(3), 71-73.

Strohmetz, D. B., Rind, B., Fisher, R., & Lynn, M. (2002). Sweetening the till: the use of candy to increase restaurant tipping. *Journal of Applied Social Psychology*, *32*(2), 300-309.

Szucs, D. & Ioannidid, J.P.A. (2017). Empirical assessment of published effects sizes and power in recent cognitive neuroscience and psychology literature. *PLoS Biol* 15 (3): e2000797. https://doi.org/10.1371/journal.pbio.2000797

Tidd, K. L., & Lockard, J. S. (1978). Monetary significance of the affiliative smile: A case for reciprocal altruism. *Bulletin of the Psychonomic Society*, *11*(6), 344-346.

Vadillo, M.A., Hardwicke, T.E. & Shanks, D.R. (2016). Selection bias, vote counting and money priming effects: A comment on Rohrer, Pashler and Harris (2015) and Vohs (2015). *Journal of Experimental Psychology: General*, 145 (5), 655-663.

van Baaren, R. B., Holland, R. W., Steenaert, B., & van Knippenberg, A. (2003). Mimicry for money: Behavioral consequences of imitation. *Journal of Experimental Social Psychology*, *39*, 393-398.

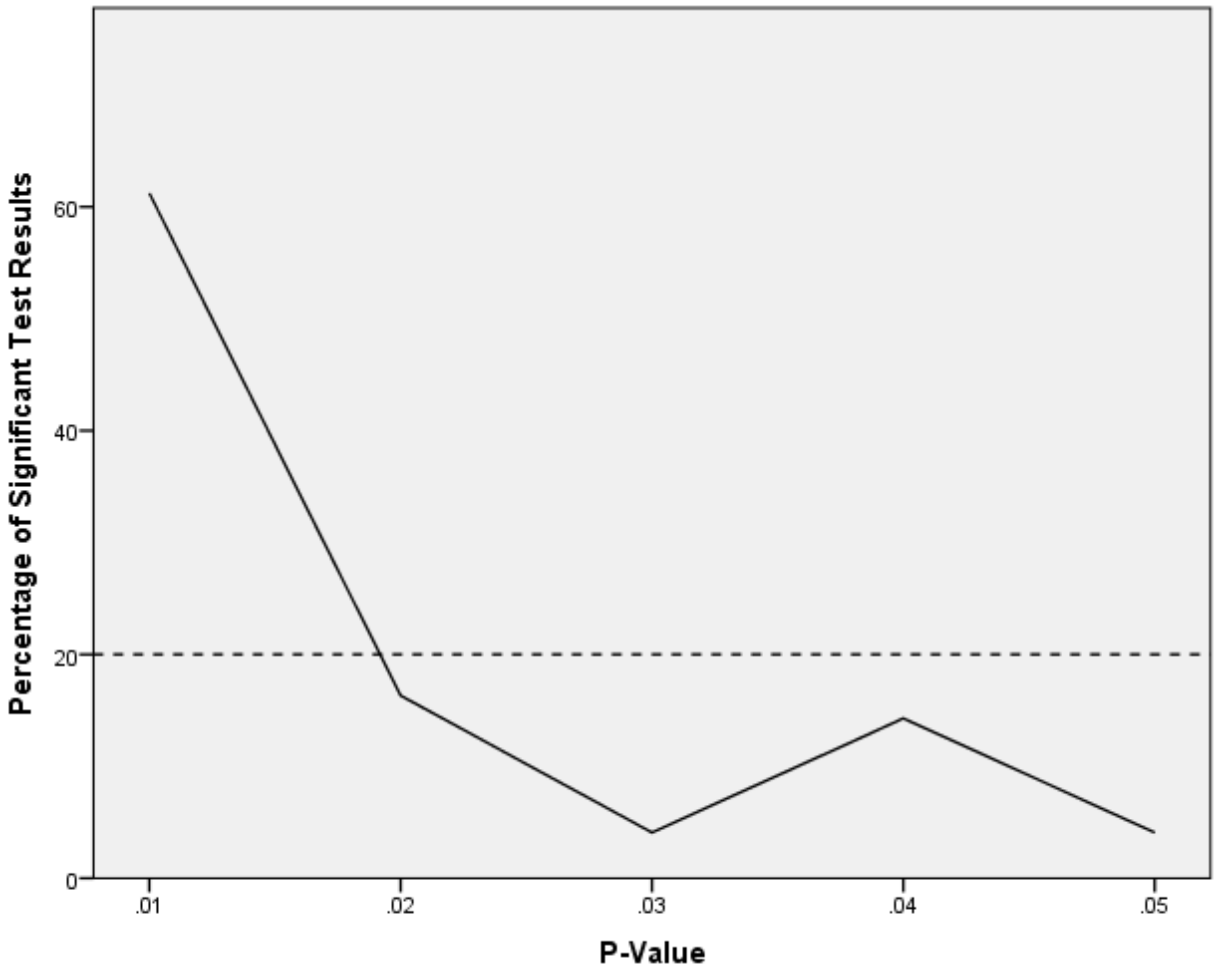

Figure 1. P-Curve for Overall Experimental Literature on Ways to Increase Tips (dashed line depicts expectations under the null-hypothesis)

Table 1. Summary of p-curve and r-index analyses of the field experiment literature on ways to increase tips.

	P-Curve Analyses		R-Index Analyses			
Set of Studies	Full p-Curve Right-Skewness Test	Half p-Curve Right Skewness Test	Success Rate	Median Observed Power	Inflation Rate[a]	R-Index
All Studies	z = -12.27, *p < .0001*	z = -12.24, *p < .0001*	.83	.69	.14 *p = .012*	.55
- Replicated Studies	z = -9.64, *p < .0001*	z = -10.07, *p < .0001*	.84	.72	.12 *p = .074*	.61
- Non-Replicated Studies	z = -7.64, *p < .0001*	z = -7.14, *p < .0001*	.81	.66	.15 *p = .063*	.51
- Studies by Gueguen	z = -2.97, *p < .0015*	z = -2.17, *p < .0148*	.83	.60	.23 *p = .033*	.37
- Studies by Lynn	z = -7.45, *p < .0001*	z = -8.09, *p < .0001*	.88	.82	.05 *p = .563*	.77
- Studies by Rind	z = -6.60, *p < .0001*	z = -6.13, *p < .0001*	.60	.62	-.02 *p =.566*	.64
- Studies by Seiter	z = -3.92, *p < .0001*	z = -4.00, *p < .0001*	1.00	.81	.19 *p = .185*	.61
- Studies by All Others	z = -8.85, *p < .0001*	z = -8.96, *p < .0001*	.86	.74	.13 *p = .138*	.61

[a] The p-values reported in this column are exact p-values from one-sample binomial tests of the hypothesis that the observed success rate equals the median observed power.

`

Appendix A: Description of studies and effects included in the analyses.

Author (Year)	Manipulation	Original Hypothesis/Research Question	Tip Measure	Test Statistics
Adams & Pettijohn (2016)	Restaurant customers' last name vs first name vs nothing written on check	"addressing a customer by their title and last name will increase the likelihood of a bigger tip" and "when a server personalizes a check by writing 'Thank you, (customers name)' this personalization will increase a server's tip."	%	F(2,100)=4.67 n=103
Crusco & Wetzel (1984)	Restaurant customer not touched vs touched on hand vs touched on shoulder	"A brief touch on the hand was expected to produce positive affect towards the waitress for both male and female customers and hence increase the amount of tip" and "…the shoulder touch condition may reduce tipping compared to the hand touch condition, and more so for male than for female customers."	%	F(2,108)=3.45 n=114
Davis, et. al. (1998)	Restaurant waitresses stood vs squatted (two counterbalanced time periods each)	"The present study … to enhance the generalizability of this (squatting increases tips) effect."	%	F(3,72)=151.57 n=112
Garrity & Degelman (1990)	Restaurant server introduced self by name vs not	"The purpose of the present study was to systematically examine the effect of server introduction (name introduction or no name introduction) on tipping behavior."	$	F(1,36)=24.02 n=42
Gueguen (2002)	Bar customers given humorous card vs advertisement card vs no card	"we predict that a humorous card, if read, will result in larger tips, compared to a situation in which the card is not shown or in which there is no humorous content on the card."	y/n	$\chi 2(2)=9.95$, n=211
Gueguen (2012)	Restaurant waitress wore red vs blonde vs dark vs brown wig	'we explored the role of blond hair color on waitresses' earnings in restaurants."	y/n	Male Customer: $\chi 2(3)=14.04$, n=503 Female Customer: $\chi 2(3)=3.12$, n=344 [P-curve robustness test omitted this study.]
Gueguen (2013)	Restaurant bill holder was square vs round vs heart-shaped	"It was hypothesized that such an object (bill dishes with a cardioid shape) would be associated with increased tipping behavior toward the employees."	y/n	$\chi 2(2)=11.72$, n=365
Gueguen & Jacob (2005)	Bar waitress touched customer vs not	"So because touch seems to be important in French culture, we expected that it could influence tipping positively."	y/n	$\chi 2(1)=4.73$, n=143

Gueguen & Jacob (2011)	Restaurant waitress wore makeup vs not	"we hypothesized that a waitress's makeup would increase tipping behavior, especially with men patrons."	y/n	$\chi2(1)=4.16$, n=174 [R-index robustness test used effect on males only: $\chi2(1)=4.51$, n=112]
Gueguen & Jacob (2012)	Restaurant waitresses wore red vs pink vs brown vs no lipstick	"Thus, it was then hypothesized that waitress's lipstick and particularly red lipstick would enhance tipping behavior, especially with men patrons."	y/n	Male Customer: $\chi2(3)=7.77$, n=319 Female Customer: $\chi2(3)=.50$, n=128 [P-curve robustness test omitted this study. R-index robustness test did not use effect on females]
Gueguen & Jacob (2014)	Restaurant waitresses wore black vs white vs red vs blue vs green t-shirts	"we hypothesized that waitresses wearing red clothes would receive larger tips from men than those wearing other colors"	y/n	Male Customer: $\chi2(5)= 12.29$, n=418 Female Customer: $\chi2(5)=1.19$, n=304 [P-curve robustness test omitted this study. R-index robustness test did not use effect on females]
Gueguen & Legoherel (2000)	Bar check had drawing of sun vs not	"we expected such a drawing at the bottom of a bar bill would more dispose the client to leave money for the waiter."	y/n	$\chi2(1)=6.22$, n=177
Hornik (1992)	Restaurant servers touched customers vs not	"This experiment is not only a replication but also a significant extension of their (Crusco & Wetzel's 1984) study. First, … investigated possible server-customer gender interaction effect."	%	Touch: $F(1,214)=6.40$ n=248
Hubbard, et. al. (2003)	Restaurant and bar servers touched customers vs not	"we expect the following: Patrons who are touched briefly on the shoulder will give their servers a larger gratuity…"	%	Restaurant: $F(1,192)=37.91$, n=200 Bar: $F(1,192)=71.51$, n=200
Hulsheger, et. al. (2015)	Hairdressers trained to regulate their emotions vs not	"H3: Participants receiving a self-training intervention in emotion regulation strategies will received more tips …"	$	$t(228)=2.75$, n=231
Jacob & Gueguen (2012)	Restaurant waitresses stood short vs medium vs far distance from customers	"we hypothesized that short distance between servers and patrons would enhance customer tipping."	y/n	$\chi2(2)=14.83$, n=478

Jacob & Gueguen (2013)	Restaurant waitress verbally mimicked customer vs not	"we hypothesized that mimicry would be associated with more tips than when no mimicry was displayed."	y/n	$\chi 2(1)=6.04$, n=237
Jacob & Gueguen (2014)	Pizza delivery driver wore suit vs casual clothes	"it was hypothesized that patrons ordering pizza at home would give a higher tip to a delivery man wearing a two-piece suit rather than neat but casual clothes."	%	t(53)=2.40, n=55
Jacob, Gueguen, Ardiccioni & Senemeaud (2013)	Restaurant check contained altruism vs neutral vs no quote	"It was hypothesized that exposure to altruism content would be associated with more generous tipping behavior."	y/n	$\chi 2(2)=10.7$, n=349
Jacob, Gueguen & Boulbry (2010)	Restaurant's background music had prosocial lyrics vs neutral lyrics vs regular lyrics	The purpose of our study was to test the effect of songs with prosocial lyrics on patrons' tipping behavior in the restaurant."	y/n	F(2,66)=7.36, n=72
Jacob, Gueguen, Boulbry & Ardiccioni (2009)	Restaurant waitresses wore makeup vs not	"we hypothesized that waitress's makeup would enhance tipping behavior, especially with men patrons."	y/n	$\chi 2(1)= 6.88$, n=274 [R-index robustness test used effect on males only: $\chi 2(1)=6.03$, n=186]
Jacob, Gueguen & Delfosse (2012)	Restaurant waitresses wore a plastic hair barrete decorated with a bird vs sprig vs rose vs nothing	"our objective was to replicate the study of Stillman and Hensley (1980)…"	y/n	$\chi 2$ (3)=7.81, n=665
Jewell (2008)	Restaurant waitresses touched customers vs not	"The last hypothesis was that interpersonal touch would increase tip percentages."	%	t(95)=1.50, n=97
Jiang & Galm (2014)	Restaurant waitresses dyed hair blond vs not	"Thus, blondeness could positively or negatively affect incomes of women in the labor market, depending on whether the beauty perception dominates, or the low intelligence perception prevails."	%	t(267)=2.69, n=282
Karagiorgakis & Malone (2014)	Bartendress' clothing covered little vs medium vs large vs regular amounts of skin	"It was hypothesized that less clothing (worn by bartendress) would result in significantly larger tips."	%	F(3,11)=2.74, n=15
Khadjavi (2017)	Hairdresser collected charitable donations vs collected and donated not	"The central hypothesis of our study is that tips increase with the kindness of the hairdresser."	$	$\chi 2(2)=7.81$, n=834

Lee, Noble & Biswas (2017)	Restaurant check holder was gold vs black	"Consumers presented with a gold-colored service prop will leave a larger tip(%) than consumers presented with a non-gold-colored (e.g., black or white) service prop."	%	F(1,226)=4.76 n = 235
Leodoro & Lynn (2007)	Restaurant waitress stood vs sat down at table.	"Which of these two processes dominates the other and, therefore, which racial group" (Whites or Blacks) "responds more positively" (thru tipping) "to having servers sit down at or lean over the table is an empirical question."	%	Whites: t(170)=2.00, n = 172 Blacks: t(126)=1.66, n = 128 [simple main effects should be used when interactions are cross-over]
Lynn & Gregor (2001)	Hotel bellman gave full vs limited service	"The present study begins to address these questions by examining the effects of service effort on the tips given to a hotel bellman."	$	t(48)=4.10, n =50
Lynn, Le & Sherwyn (1998)	Restaurant waiter touched customer briefly vs long time vs not at all	"We sought to determine whether the age of the customer influenced customers' reactions to being touched and whether a long versus a short touch created some difference in the customer's reaction to being touched."	%	Touch: F(2,102)=4.82, n = 105
Lynn & Mynier (1993)	Restaurant servers stood vs squatted	"Thus, it seems likely that squatting down will increase as server's tips because it makes the server seem friendlier."	%	S1: t(268)=7.41, n = 270 S2: t (146)=4.81, n = 148
McCall & Belmont (1996)	Restaurant check holder had credit card logo vs not	"it was hypothesized that restaurant tips would increase in the presence of such credit cues."	%	S1: F(1,74)=6.60, n = 77 S2: F(1,24)=4.89, n=27
Rind (1996)	Casino hotel customer told weather was rainy vs sunny and warm vs cold (in a 2x2 design)	"Hence, the present investigation was exploratory; no predictions were made."	%	F(3,201)=2.467, n = 205 [calculated from reported summary statistics]
Rind & Bordia (1995)	Restaurant server wrote on check thank-you vs thank-you plus name vs nothing	"It was predicted…that the addition of a written gratitude and personalization would increase tips over writing nothing. Furthermore… it was expected that personalizing the gratitude message would increase tips over the gratitude message alone."	%	Thanks vs nothing: t(48)=1.8, n = 51 [excluded - Thanks vs thanks plus name: t (48)=.08, n =51]
Rind & Bordia (1996)	Restaurant server drew a smiley face on check vs not	"it was hypothesized … that drawing a happy smiling face would increase tips when done by the female server but not when done by the male server."	%	Waitress: t(85)=1.68, n = 45 Waiter: t(85)=-1.15, n =44
Rind & Strohmetz (1999)	Restaurant waitress wrote helpful message on check vs not	"it was expected that customers would perceive the massage as a helpful 'tip' for them that cost the server some time in writing up, which would have the effect of increasing their willingness to return the favor by leaving a larger tip."	%	t(79)=1.93, n = 81

Rind & Strohmetz (2001a)	Restaurant waitress predicted good vs bad weather vs noprediction	"we hypothesized that induced beliefs about future weather conditions written on customers' checks would affect tipping."	%	F(2,57)=20.58, n = 60
Rind & Strohmetz (2001b)	Restaurant customers were shown vs given vs nothing a card with an interesting puzzle	"it was expected that the interesting task would heighten tip percentages, especially when customers were allowed to keep the task."	%	Card vs nothing: t (57)=2.95,n =60 [excluded - Shown vs given: t (57)=.30, n = 40]
Seiter (2007)	Restaurant waitresses complimented customers' food choices vs not	"Food servers will receive significantly higher tips when they compliment their customers than when they do not."	%	F(1,92)=4.60, n= 94
Seiter, et. al. (2011)	Restaurant check had gratuity guideline vs not	"Food servers will receive significantly higher tips when providing their customers gratuity guidelines in calculation –assistance format than when they do not provide such guidelines."	%	F(1,111)=4.69, n= 113
Seiter & Dutson (2007)	Hairstylist complimented customer vs customer's hairstyle vs nothing	"Hair stylists will receive significantly higher tips when they compliment their customers than when they do not."	%	F (2, 112)=4.38, n=115
Seiter & Gass (2005)	Restaurant waitresses wrote Have a Nice Day vs God Bless America vs United We Stand vs nothing on checks	"Research Question: Do female food servers who use patriotic messages earn significantly higher tips than those who do not?"	%	F(3,96)=7.10, n=100
Seiter, Givens & Wegner (2016)	Restaurant waitresses learned and used customers names vs not	"Food servers will earn larger tips when they invite mutual introductions and address customers by name than when they do not."	%	t(87)=3.4, n= 90
Seiter & Weger (2010)	Restaurant servers complimented customers' food choices vs not	RQ1: "Is there a significant difference between the tips received by food servers who give customers generalized compliments and the tips received by food servers who do not give compliments?"	%	t(344)=3.80, n=360
Seiter & Weger (2013)	Restaurant servers thanked customers with customers' first name vs last name vs title vs nothing	H1: "Immediacy (ie, addressing customers by name) will increase gratuities received by food servers." RQ1: "Does receiving more formal or less formal forms of address from servers have an effect on customers' tipping behavior?"	%	F(3,131)=5.96, n= 135 [calculated from reported summary statistics]
Seiter & Weger (2014)	Restaurant waitresses handled leftovers by handing customer box vs boxing leftovers vs	H: "Food servers who box customers leftovers will receive higher tips than food servers who do not."	%	F(4,197)=4.63, n= 199

	boxing and dating them vs boxing and adding customer name to them vs boxing, dating and naming them	RQ: "Is writing guests' names and/or the day's date on boxes of left-overs associated with higher tips for food servers compared with not writing such information?"		
Seligman, et. al. (1985)	Pizza delivery customers were told speed of service was due to driver effort or external factors	"Specifically, customers making a person attribution were expected to tip more when the pizza arrived early than when it arrived late. Customers making a situation attribution, however, were not expected to tip any more for an early than for a late delivery."	$	F(1,56)=4.49, n=60 [speed x attribution interaction]
Stephen & Zweigenhaft (1986)	Restaurant waitress touched female vs male vs no one in mixed sex dyads	"we hypothesized that it would be more profitable for a waitress to touch the female than the male."	%	F(2, 109)=13.79, n = 112
Stillman & Hensley (1980)	Restaurant waitresses wore flower in hair vs not	"Diners will leave a larger tip for a waitress who wears a flower in her hair than for the same waitress without a flower."	$	t(374)=3.13, n = 376
Strohmetz & Rind (2001)	Restaurant servers gave customers tip guidelines vs not	"We investigated the following questions: What impact does providing customers with a suggested gratuity have on the actual tip left for the server?"	%	t(108)=.62, n = 110
Strohmetz, Rind, Fisher & Lynn (2002)	Restaurant servers gave customers candy (1, 2 or 1+1 pieces) vs not	"there are at least two other reasons for expecting gifts of candy to affect" (increase) "tips."	%	S1: t(90)=5.25, n = 92 S2, candy vs none: t(76)=4.49, n = 80 [excluded - S2, one vs two: t(76)=4.70n= 60; S2, 1+1 vs 2: t(76)=2.06, n =40]
Tidd & Lockard (1978)	Cocktail waitress gave customers big vs small smile	"It was predicted that, to a broad smile from the waitress, single men would leave larger tips and return more smiles than single women."	$	Males: t (46)=5.58, n= 48 Females: t(46)=2.96, n = 48
Van Baaren, et. al. (2003)	Restaurant waitress verbally mimicked customer vs not	"The primary goal of the present studies is to investigate the behavioral consequences of mimicry. Does mimicry produce larger tips for waitresses?"	y/n	S1: $\chi 2(1)= 2.85$,n=59 S2:$\chi 2(1)=9.38$, n=141