

ORIGINAL ARTICLE

A meta-analytic review of tipping compensation practices: An agency theory perspective

George C. Banks¹  | Haley M. Woznyj² | Sven Kepes³ |
John H. Batchelor⁴ | Michael A. McDaniel³

¹University of North Carolina at Charlotte

²Longwood University

³Virginia Commonwealth University

⁴University of West Florida

Correspondence

George C. Banks, Belk College of Business, University of North Carolina at Charlotte, 9201 University City Blvd., Charlotte, NC 28223.
Email: gcbanks@gmail.com

Abstract

Tipping represents a form of compensation valued at over \$50 billion a year in the United States alone. Tipping can be used as an incentive mechanism to reduce a principal–agent problem. An agency problem occurs when the interests of a principal and agent are misaligned, and it is challenging for the principal to monitor or control the activities of the agent. However, past research has been limited in the investigation of the extent to which tipping is effective at addressing this problem. Following an examination of 74 independent studies with 12,271 individuals, meta-analytic results indicate that there is a small, positive relation between service quality and percentage of a bill tipped ($\hat{\rho} = .15$ without outliers). Yet, in support of the idea behind tipping, relative weights analyses illustrate that service quality was a stronger predictor of percentage of the bill tipped than food quality, frequency of patronage, and dining party size. Evidence also suggests that racial minority servers tend to be tipped less than White servers (Cohen's $d = .17$), and women tend to be tipped more than men (Cohen's $d = .15$). Still, given the magnitude of the effect, one might question if tipping is an effective compensation practice to reduce the principal–agent problem. We discuss theoretical and practical implications for future research.

Determining how to structure compensation is a critical strategic decision for organizations. Compensation is important to employee attitudes and behaviors, to the success of organizational functioning, and ultimately, to a firm's sustained competitive advantage (Gupta & Shaw, 2014). From the perspective of both organizations and employees, pay-for-performance tends to be the most desired general approach for compensating employees (Heneman & Werner, 2005; Kepes, Delery, & Gupta, 2009). Tipping, a particular approach to pay for performance, is especially popular; in fact, it is one of the most popular types of compensation practices in the United States and around the world (Lynn, 2015a, 2015b; Lynn, Zinkhan, & Harris, 1993). Under a tipping system, customers voluntarily give some amount of money, called tips or gratuities, above and beyond the contracted price of a service after that service has been rendered (Lynn & McCall, 2000). Recent estimates suggest that in the U.S. restaurant industry alone, close to \$47 billion

annually is tipped (Azar, 2011), up from about \$9 billion in the 1980s (Pearl, 1985); an increase of over 400% in just three decades.

Yet, despite the fact that the literature on tipping has grown substantially in the past decades, several prominent gaps remain, such as a strong theoretical framework. This is a major concern as compensation represents one of the hardest topics to study in management and applied psychology, yet it is also one of the most critical topics for organizations (e.g., pay constitutes a very large part of an organization's expenses; Gerhart, Rynes, & Fulmer, 2009). Structuring compensation becomes complex because a principal-agent problem may exist where the interests of principals (e.g., owners of service establishments) and agents (e.g., individual service providers) might not align (Bodvarsson & Gibson, 1997; Jacob & Page, 1980). More formally, this problem occurs "when (a) the desires or goals of the principal and agent conflict and (b) it is difficult or expensive for the principal to verify what the agent is actually doing" (Eisenhardt, 1989, p. 58). This goal incongruence can give way to agency costs (Berrone & Gomez-Mejia, 2009; Lynn, Kwortnik, & Sturman, 2011). Agency theory suggests that such costs can be reduced with outcome-based (e.g., incentives) and behavior-based or control mechanisms (e.g., monitoring; Eisenhardt, 1989).

In this work, we address this first gap, in part, by developing a comprehensive framework that applies agency theory (Dalton, Hitt, Certo, & Dalton, 2007; Eisenhardt, 1989) along with expectancy theory (Vroom, 1964). Drawing upon the formal version of agency theory (Grossman & Hart, 1983; Holmström, 1979; Oyer & Schaefer, 2011) helps us to understand that, in order to align the interests of principals and agents, agency costs stemming from moral hazards (i.e., "a lack of effort on the part of the agent"; Eisenhardt, 1989, p. 61) must be reduced. However, the theory falls short in explicitly providing a means of doing so. Expectancy theory can help to explain how to motivate specific behaviors to address the moral hazards. Thus, we explicitly apply expectancy and agency theories to understand how tipping can be used as an incentive mechanism to reduce moral hazards (Holmström, 1979) and, therefore, address the principal-agent problem.

As a second gap, inconsistent results from previous research call into question the extent to which tipping actually represents a pay-for-performance compensation practice. Theoretically, customers give tips (pay) to reward high-quality service (performance). However, empirical evidence suggests that the relation between service quality and the percentage tipped varies from negative (e.g., Barkan & Israeli, 2004; Lee, 2015) to nil (e.g., Lee, 2015), to positive (e.g., Lynn et al., 2008). Consequently, the degree to which service quality is related to the tips that a server receives (i.e., the rewards customers provide) is unknown, which may shed some light on the relatively high turnover rate in the restaurant industry (National Restaurant Association, 2015). Effective compensation practices are designed to attract highly qualified people, motivate them to perform at a high level, and to retain the high performers (Gerhart & Rynes, 2003). In the simplest sense, research has questioned whether tipping is successful in fulfilling these objectives.

Third, there is a need to identify contingency factors that could help to explain the inconsistent findings between service quality and received tips (Lynn & McCall, 2000). For instance, factors such as cultural social norms, patronage frequency, or the race/ethnicity of the server and/or customer could affect the percentage tipped, regardless of the quality of service provided to the customer (Lynn, 2006). It is important to account for such factors to properly explain the conditions under which tipping is an effective compensation practice for attracting, motivating, and retaining high-quality performers in the service industry.

We begin our meta-analytic review with the development of a comprehensive theoretical framework by applying agency theory and expectancy theory. This model serves to explain the individual-level motivating dynamics within the service industry. In doing so, we propose a series of hypotheses and research questions. We test our model using meta-analytic techniques with data from 74 independent samples containing 12,271 participants. We also identify and test important contingency factors that may moderate the extent to which tipping represents a pay-for-performance compensation practice. We use a recommended comprehensive battery of meta-analytic and publication bias techniques to assess the robustness of our obtained results. Such a comprehensive approach has not been applied in the vast majority of meta-analytic reviews to date (Kepes, Banks, McDaniel, & Whetzel, 2012; Kepes, McDaniel, Brannick, & Banks, 2013). We report the obtained results and conclude with recommendations for future research and practice to improve tipping practices in the service context. Further, we discuss how our theoretical model can guide future research.

1 | AGENCY THEORY

The main tenet of agency theory argues that a problem arises when one party (the principal) contracts with another party (the agent) to perform and make decisions on behalf of the principal (Eisenhardt, 1989; Lubatkin, Lane, Collin, & Very, 2005). The principal-agent problem can occur because both parties in the exchange relationship, the principal and the agent, may have diverging or conflicting interests. It could be costly for the principal to monitor (Holmström, 1979) or check the behaviors of the agents to ensure they are acting in ways that align with the principal's interests (Deckop, Mangel, & Cirka, 1999; Eisenhardt, 1989). Models based on agency theory typically operate on the assumption that agents are opportunistic and will act to maximize their own interests, potentially at the expense of the principal's interests (Cohen & Holder-Webb, 2006).

From this rational perspective, agents seek to work as little as possible. When principals lack the information needed to monitor their agents and the tasks require specialized knowledge, "moral hazards" can emerge whereby agents act in self-interested ways (e.g., shirk duties; fail to help coworkers; Gomez-Mejia & Balkin, 1992). Moral hazards may be particularly harmful in instances where information asymmetries are large and monitoring is quite difficult (Holmström, 1979). To address the principal-agent problem and reduce subsequent moral hazards, principals can implement mechanisms, such as control, to align the interests of principals and agents and limit self-serving behavior of the agents (Eisenhardt, 1989; Gomez-Mejia & Balkin, 1992). Similarly, principals can use incentive mechanisms to promote behaviors that are in the best interest of the principal.

Scholars have begun to express concerns about the narrow, almost exclusive application of agency theory to the corporate governance context and thus question its applicability in settings other than large, for-profit corporations, which limits the theory's utility (e.g., Berrone & Gomez-Mejia, 2009; Dalton et al., 2007). For instance, it has been proposed that agency theory should be a useful framework to understand compensation for jobs other than executives that are high in autonomy and where oversight is limited (Gomez-Mejia & Balkin, 1992), like individual service providers such as bartenders or restaurant servers (Lynn et al., 2011). The applicability of agency theory in such contexts remains understudied and provides opportunities to extend the theory beyond the corporate governance setting.

2 | THE APPLICATION OF EXPECTANCY AND AGENCY THEORIES

Expectancy theory (Lawler, 1971; Vroom, 1964) is used to explain how to motivate behavior. That is, it explains how employees assess the probability that their effort (Grossman & Hart, 1983) will lead to a given level of performance (i.e., expectancy), the probability that the achieved level of performance will be rewarded (i.e., instrumentality), and the extent to which they anticipate valuing the associated reward (i.e., valence). A compensation practice to motivate high levels of effort and performance could be designed by taking into consideration expectancy, instrumentality, and valence.

Agency theory contributes to expectancy theory by considering circumstances where moral hazards exist, thereby focusing on how to motivate behaviors that are in the best interest of both parties (principals and agents). Simultaneously, expectancy theory contributes to agency theory by helping to explain potential incentive mechanisms that may reduce moral hazards. Consequently, we use both agency and expectancy theory to build a theoretical framework to understand tipping as a means to address the principal-agent problem that arises in service contexts (see Figure 1 for an illustration). In the paragraphs that follow, we walk through the logic of our theoretical framework using an example in a restaurant context with servers, restaurant owners, and customers who act as proxies of the principal.

Servers (agents) work on behalf of the owner (principal) of a service establishment (Lynn et al., 2011). Although principals want their establishment to be successful (in terms of profits, return customers, etc.), agents want to secure high pay. Pursuant to the principal-agent problem, the agent's interests (Figure 1, box 1) may not always be aligned with the principal's interests (Oyer & Schaefer, 2011). When interests are not aligned, the potential for moral hazards is high. Job performance in the service industry is defined as providing high-quality service to customers. Principals

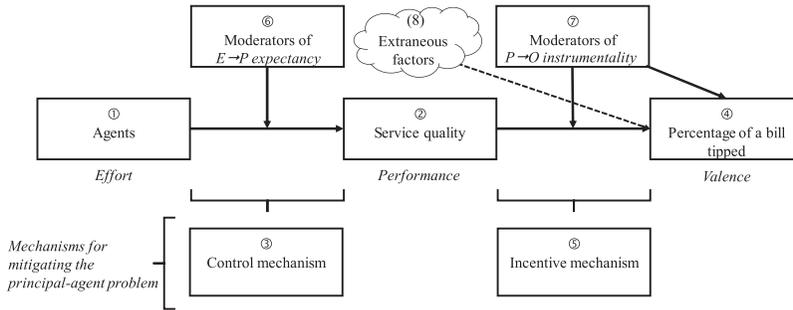


FIGURE 1 The principal-agent problem in a service context

(owners) want agents (servers) to provide high-quality service to customers to ensure that customers have positive experiences in their restaurant and return. For instance, servers may frequently refill customer drinks and ask whether “everything is to one’s expectations.” Positive customer experiences, in turn, enhance the restaurant’s reputation and increase the likelihood of new and repeat customers, which is the primary interest of principals. Thus, it is in the owner’s best interest for the servers to provide high-quality service to customers (Figure 1, box 2).

However, performance (i.e., service quality) is often very difficult and costly for principals to monitor and assess (Holmström, 1979). This is partly because performance is very idiosyncratic; different customers are likely to have quite distinct expectations for high-quality service. To address this, employees in the service sector typically have substantial autonomy and freedom; they can use their discretion when delivering service to customers, and in doing so, they have the ability to directly influence customer satisfaction and their perceived service quality (Brewster, 2015; Lynn & McCall, 2000). Agents in a service context understand better than the principals the diverse needs of customers and are better able to discern which actions or behaviors will lead to customers’ perceptions of high service quality.

Further complicating the principal-agent problem in such situations are information asymmetries (Holmström, 1979), which occur because principals are not likely to have all the information necessary to efficiently monitor and control the behaviors of their agents (Figure 1, box 3; Berrone & Gomez-Mejia, 2009). Control mechanisms are not practical in this setting. Unlike jobs in manufacturing where owners can conduct quality control checks, it is more difficult for owners to conduct such checks in a service context due to the personalized nature of the product and thus the performance (i.e., service quality). Such situations can lead to moral hazards because agents have the opportunity to engage in self-serving behaviors that are detrimental to the principal’s interests without the principal being aware. In order to diminish such moral hazards, agency theory suggests that one way to reduce the costs associated with monitoring (e.g., time, energy, money) is to structure incentive mechanisms in a way that aligns the principals’ and agents’ interests (e.g., Deckop et al., 1999; Eisenhardt, 1989).

As such, we draw upon expectancy theory and agency theory to provide insight regarding how to structure such incentive mechanisms (Figure 1, box 5). Owners want their servers to provide high-quality service and servers want the rewards associated with providing such high-quality service (i.e., tips). Customers act as the proxies for the principal when assessing the performance of the service and, accordingly, rewarding the server. This approach only works if the interests of the owner and the customer are aligned. The owner–customer relation is complex too, and owners often invest in shaping the relationship (e.g., through coupons, customer relationship management, etc.). Ideally, both parties want high-quality service whereby quality is defined by the customer. This makes the customer *and* the owner happy (good service increases the chances that the customer returns and spends more money). Customers, as recipients of the service, are in a good position to monitor, evaluate, and reward the quality of the service provided.

Customers know what they expect in terms of servers’ behaviors that make up good service quality (Figure 1, box 2). Thus, customers can reward servers’ efforts and performance by providing tips (Figure 1, box 4). To the extent that servers believe that their effort will result in customers’ perceptions of high-quality service (i.e., expectancy) and that providing high service quality will result in higher tips (i.e., instrumentality; Lawler, 1971), which they value (i.e., valence; Lawler, 1971), they will continue to provide high service quality. Thus, tipping represents an incentive

mechanism (Figure 1, box 5) that aligns principals' and agents' interests (Oyer & Schaefer, 2011). Although the expectancy relationship is an important component of the proposed model and process, the focus of this paper is on tipping as a pay-for-performance compensation practice. As such, we concentrate on the performance–outcome (i.e., the service quality–tip relation; Figure 1, box 5) relation in particular. Customers' tips not only help owners to verify that servers engage in high-quality service—behaviors that are in the best interest of the owner; tips also provide an incentive for servers to provide good service because of the monetary reward, which is in the best interest of the server (Lynn et al., 2011). Following this logic, one would assume that servers are motivated to provide better service quality when the instrumentality is strong, that is, when customers are willing to pay for better service. Hence, we test the instrumentality condition in Hypothesis 1 and propose:

Hypothesis 1: There will be a positive relation between service quality and percentage of the bill tipped.

The relation between service quality and tip (i.e., the incentive mechanism; Figure 1, box 5) rests on the assumption that servers value the tips that they receive (i.e., valence, Figure 1, box 4; Hackman & Porter, 1968; Lawler, 1971). Tip intensity refers to the absolute (i.e., tip amount in dollars) or relative (percentage of a bill tipped) amount tipped; for example, a \$25 (or 30%) tip would be of higher intensity than a \$2.50 (or 3%) tip. If tipping is an effective pay-for-performance compensation system, the expectation of a high-intensity tip should be incentivizing for the server. For absolute amounts, drawing upon previous compensation research (Mitra, Gupta, & Jenkins, 1997; Mitra, Tenhällä, & Shaw, 2016), the expected tip amount may have to cross a particular threshold to be considered valuable, and thus incentivizing, to the server. That is, the expectation of a small tip (e.g., \$2.50 or 3% of the bill) may not elicit good customer service because the server does not value it (i.e., there is low valence); thus, with such a small tip, servers are unlikely to be motivated to display high levels of effort and pursue high service quality. However, anticipation of a larger tip (e.g., \$25 or 30% of the bill) should elicit high service quality. Thus, as a test of the valence of percentage of a bill tipped (Figure 1, box 4), we hypothesize the following:

Hypothesis 2: The relation between service quality and tip will be stronger in instances where tip intensity is high in either absolute (2a) or relative terms (2b).

When dining in a restaurant, customers also have experiences that may or may not be within the control of the service provider, but influence tips (and the effectiveness of tipping as an incentive mechanism) nonetheless. These extraneous factors can moderate the instrumentality relationship (i.e., the service quality–tip relation; Figure 1, box 7) as well as influence customers' decisions to tip directly (Figure 1, box 8). For example, research has shown that ratings of patronage frequency, dining party size, and food quality influence the amount tipped (Lynn, 2006; Lynn & McCall, 2000; Rogelberg et al., 1999). Patronage frequency ought to be related to the amount tipped as customers are likely to tip a higher percentage if they frequent a restaurant often in order to ensure good service in the future. Furthermore, dining party size may also relate (negatively) to percentage tipped as there may be a diffusion of responsibility in which larger parties assume that others are providing an adequate tip and which may reduce the amount that each individual tips (Seiter & Weger, 2010). Finally, we also expect food quality to relate to the amount tipped. Customers who experience delicious (or distasteful) food are likely to consider this when determining how much tip to leave, as it is a part of the whole dining experience (Lynn, Jabbour, & Kim, 2012).

Yet, despite the importance of these factors, there is reason to expect that service quality should be the strongest predictor of tip percentage. The primary assumption in the use of tipping as an incentive mechanism is that achieved levels of performance, or service quality, are appropriately rewarded (i.e., instrumentality; Azar, 2009; Lynn, 2015a, 2015b). In addition, principals implement tipping as a form of compensation to encourage high service quality. In a tipping context, this implies that customers reward employees with an extra sum of money when they have provided a service that meets expected levels or goes beyond their expectations (Lynn & Graves, 1996; Lynn & McCall, 2000).

Consequently, service quality is likely to be more strongly related to tips than patronage frequency because the quality of the immediate service experience is likely to weigh more in a customer's decision to tip a certain amount

compared to their desire to encourage good service in the future (Azar, 2007a, 2010). Furthermore, service quality is also likely to be more important than dining party size; although a diffusion of responsibility in tipping among the party is possible, each tipper is still likely to consider the quality of service they received when determining how much to tip (Rogelberg et al., 1999). Finally, food quality is primarily a reflection of the restaurant's chefs and only to a noticeably lesser extent the performance of the server (Medler-Liraz, 2012). Although tips are a reward for a server's performance, we expect service quality to have greater relative importance than food quality when predicting percentage tipped. In sum, we propose the following hypothesis, which tests the relative importance of service quality (Figure 1, box 2) compared to extraneous factors (Figure 1, box 8):

Hypothesis 3: Service quality has greater relative importance when predicting percentage of a bill tipped than does patron frequency (H3a), dining party size (H3b), or food quality (H3c).

3 | BOUNDARY CONDITIONS IN THE INSTITUTION OF TIPPING

We now turn to potential boundary conditions in the institution of tipping. Given that one primary objective of our study is to understand the extent to which tipping represents an effective pay-for-performance compensation practice (and thus a viable incentive mechanism to align principal and agent interests), this section centers on the identification of contingency factors to help explain the inconsistent findings in past studies between service quality and tips. Due to data constraints, we do not propose formal research questions for all presented ideas and arguments.

3.1 | Expectancy

Various individual difference variables and job characteristics can influence the extent to which service providers expect that their effort will yield a certain level of performance (Figure 1, box 6; Hackman & Porter, 1968; Van Eerde & Thierry, 1996). For example, servers' individual characteristics, such as self-efficacy, affect their belief that they can achieve the desired level of performance (Gupta, Ganster, & Kepes, 2013). Job characteristics, such as job autonomy, role ambiguity, and role overload may also influence the relation between effort and performance (Hackman & Oldham, 1975; Kahn, Wolfe, Quinn, Snoek, & Rosenthal, 1964). For example, if a customer has a particular request, like a customized food order (e.g., a sandwich without onions), a server's efforts are only likely to result in the desired outcome if he or she has the discretion or autonomy to carry out the request. As another example, service in the dining area of a restaurant may be slow if the kitchen is understaffed. As a result, the effort servers extend may not lead to a comparable level of service quality (performance) had the kitchen been properly staffed (Grossman & Hart, 1983).

3.2 | Instrumentality

There is reason to question whether the relation between service quality and tipping behavior (i.e., instrumentality) is not as strong as one might hope. Several factors may affect, and potentially undermine, the relation between high service quality and percentage of the bill that is tipped (Figure 1, box 7). For instance, social norms within a geographic region (e.g., country where customers are from or a restaurant is located) are likely to influence the extent to which service quality results in tips (Azar, 2011; Bodvarsson & Gibson, 1997). For example, a social norm has developed in the United States whereby tipping has become an expectation for many jobs in the service sector. Although customers are not legally obligated to pay anything above the contracted price of the service, failure to do so may result in social disapproval from others, feelings of guilt, or even poor service in the future (Azar, 2004b; Lynn et al., 1993). Hence, the service quality–percentage tipped relation may be reduced because of a norm to tip even when service quality is poor. If servers are tipped a high amount regardless of the quality of the service they provide, tips may lose their motivating force (i.e., instrumentality is low) because their tip is almost certain. In sum, it may be possible that service quality is rewarded more consistently in some countries (e.g., the United States) than other countries.

Although consistency in tipping is one potential explanation, differences also exist across countries in terms of expectations regarding percentage of a bill tipped. In some countries, tipping is not as customary as it is in the United States (Lynn, 1997, 2006). As an example, although tipping approximately 15–20% of the bill in restaurants is common in the United States, the average tip percentage in Romania is 5–10% (Lynn & Lynn, 2004). Consequently, the relation between service quality and tips might depend on the country that customers are from and the extent to which they adhere to and follow certain social norms from their home country. As such, we ask the following research question:

Research Question 1: To what extent does country moderate the relation between service quality and percentage of the bill tipped?

Although the above research question accounts for differences across countries, it does not account for variability within a country. Characteristics at the individual level are likely to affect the service quality–tips relation as well (Figure 1, box 7), and thus the effectiveness of tipping as a pay-for-performance compensation practice. For instance, demographic characteristics of customers could be a factor that explains within country variability. As an example, after controlling for socioeconomic status and service quality, research has found that Black customers tip less than Whites do. One potential explanation is that Blacks are less likely, due to cultural differences, than Whites to perceive that 15–20% of the bill is a customary tip size (Lynn, 2007; Lynn & Sturman, 2011).

Differences in tipping norms could also exist among other demographic groups. For instance, older customers might tip less due to greater experience with the social norms surrounding tipping when the “typical” or socially acceptable tip percentage was less than it is today (Margalioth, 2006). Tipping norms can also vary by gender. Women, for example, could be more equity sensitive or men could be more likely to tip based on server attractiveness rather than service quality (Parrett, 2015). In sum, the service quality–tips relation (i.e., instrumentality) might be stronger for some demographic groups than for others. To the extent that such customer demographics predict the percentage of a bill tipped, one may lose confidence in tipping as an effective pay-for-performance practice to align interests of principals and agents. Consequently, we propose the following research question:

Research Question 2: To what extent do demographic variables (e.g., gender, race) of the customer predict percentage of the bill tipped?

In addition to characteristics of the customer, demographic features of the server, such as ethnicity or race, could be an important factor for the instrumentality relationship as well (Figure 1, box 7). If this is the case, the potential for unfairness is likely to exist (Ayres, Vars, & Zakariya, 2005; Brewster & Lynn, 2014). Controlling for as many factors as possible, research has found that ethnic and/or racial minority servers receive fewer tips than their White counterparts, regardless of the demographics (e.g., race) of the customers (Ayres et al., 2005; Brewster & Lynn, 2014; Lynn et al., 2008). Scholars suggest such group differences could be the result of implicit racial or gender attitudes that people hold about particular groups or perhaps even unconscious biases (Brewster & Lynn, 2014; Lynn & Sturman, 2011; Lynn et al., 2008). Thus, the quality of service that the server provides could not matter much in determining the size of the tip. If customers hold prejudicial beliefs about a server’s race or gender, these customers may be more likely to leave a smaller tip regardless of the service quality they received. This could lead to group differences whereby a seemingly neutral practice (tipping) affects groups differentially (i.e., Blacks receive fewer tips than Whites do, on average). Or, if minority servers believe they will be tipped less because of their race, it may result in a self-fulfilling prophecy where minority servers may not exert as much effort because they do not believe their efforts will result in the desired outcome (i.e., a higher tip). We therefore ask the following research question:

Research Question 3: To what extent do demographic variables (e.g., gender, race) of the server predict percentage of a bill tipped?

3.3 | Methodological moderators

We planned to conduct several subgroup analyses to determine the robustness of the service quality–tip percentage relation. We thus propose the following research question to determine the effects of several methodological moderators:

Research Question 4: To what extent does study design (e.g., experimental lab study; quasi-experimental field study; regular field study), rating source (customer vs. other), and statistics reported (e.g., correlation coefficient; statistic converted to correlation coefficient) moderate the relation between service quality and percentage of a bill tipped?

4 | METHODS

4.1 | Systematic search

Following best practice guidelines (see Kepes et al., 2013), we systematically searched the literature on tipping for published and unpublished samples. Databases (*Ebscohost*, the *Social Science Citation Index*, and *Google Scholar*) were searched in September 2015. We searched the databases using combinations of the following keywords: *tips*, *tipping*, *gratuity*, *gratitude*, *pay*, *compensation*, and *money*. To identify unpublished manuscripts in the form of dissertations and conference papers, we used *ProQuest Dissertations and Thesis* as well as *PapersFirst*. We also searched for working papers at the *National Bureau of Economic Research* (NBER) working papers (<http://www.nber.org/papers.html>) and the *Social Science Research Network* (SSRN) (<http://www.ssrn.com/en/>).

Using the results of the *Google Scholar* search, we identified key references (Azar, 2004a, 2004b, 2007b; Bodvarsson & Gibson, 1997; Conlin, Lynn, & O'Donoghue, 2003; Lynn, 2001, 2007; Lynn & Grassman, 1990; Lynn & Latane, 1984; Lynn & McCall, 2000; Lynn et al., 1993) for a backward and forward reference search where we examined the articles that the key reference cited as well as the articles that have cited the key reference. In addition, calls for unpublished papers were posted on the *Academy of Management* OB and HR division ListServes. Finally, the prominent tipping scholar Michael Lynn provided a comprehensive bibliography of papers on the topic of tipping, which included 399 published and 187 unpublished papers. This bibliography was checked against the results of our systematic search; no new studies were added, providing support for the thoroughness of our systematic search.

4.2 | Coding

To be included in our study, several inclusion and exclusion criteria were established. First, we elected to include only studies that used adults of traditional work populations and focused explicitly on tipping as a percentage of a bill. Hence, studies that focused on tipping where no bill was involved, as in the case of hotel bellmen (e.g., Lynn & Gregor, 2001), were excluded. Second, studies that did not report the percentage of a bill tipped or the information necessary to calculate this variable were excluded (e.g., Azar, 2010; Azar, Yosef, & Bar-Eli, 2015). Third, studies had to be quantitative in nature and report either correlation coefficients or statistics that could be converted to correlations (e.g., Cohen's *d*). In the event that a study did not report the necessary statistical information, the primary study authors were contacted. The supplemental materials include the studies where additional data were obtained via personal communication. We used detection heuristics to screen for potential duplicate studies, which can occur when the same data are used in two different articles or when a conference paper is later published as a journal article (see figure 1 in Wood, 2008). Lynn and Thomas-Haysbert (2003) aggregated several samples to make one large primary sample. We disaggregated the samples by restaurant to facilitate the formation of moderator-based subgroups.

In total, the current meta-analytic review includes 74 independent samples and 12,271 participants. The samples come from the time range of 1978–2015. Interrater reliability across 28 coding decisions (e.g., coding of effect size estimates, sample sizes) between the first and second author was $\kappa = 1.0$. We report the data for each sample in Appendix A (e.g., reliability, sample size, correlation coefficient) in the Open Science Framework online repository (https://osf.io/rnmky/?view_only=1995b7c813e54002a3b241d9485198f1).

4.3 | Meta-analytic procedure and robustness checks

4.3.1 | Psychometric meta-analysis

Psychometric meta-analysis was used to analyze the data because it allows for the correction of methodological artifacts such as measurement error (Schmidt & Hunter, 2015). As many of the studies did not report a reliability coefficient for *service quality*, this information was imputed, based on an average of those reliabilities reported (average $\alpha = .88$), for those studies that did not report reliability estimates. All analyses were at the individual-level of analysis. We considered the potential for moderating variables by calculating an 80% credibility interval around the corrected parameter estimates. Wide credibility intervals or intervals that include zero can be interpreted as evidence consistent with moderating effects. We report 95% confidence intervals (CIs) as well.

4.3.2 | Relative weights analyses

The use of relative weights analyses in meta-analytic reviews have been growing in popularity (e.g., Banks, Davis McCauley, Gardner, & Guler, 2016; Cole et al., 2012; Tonidandel & LeBreton, 2011). Such analyses are advantageous compared to more traditional regression or correlational analyses in examining the relative importance of predictors as they allow for an understanding of the predictive validity of correlated constructs. Standardized regression coefficients, for example, are not accurate indicators of the relative importance of predictors because they do not appropriately account for multicollinearity among predictors when partitioning variance; furthermore, bivariate correlations only consider the relationship between the predictor and outcome by itself (Tonidandel & LeBreton, 2011). Relative weights analysis address these issues because it uses orthogonal transformations of the predictors (Tonidandel & LeBreton, 2011). Thus, the analyses offer information regarding the amount of variance explained in a criterion (i.e., tip amount) that is attributable to each predictor variable. A relative weights analysis can be completed using an epsilon weight technique (Johnson, 2001; Tonidandel & LeBreton, 2011). Once computed, the weights are summed to R^2 , which can then be compared via ratios. For example, an epsilon weight of 0.20 is interpreted to be twice as large as an epsilon weight of 0.10, and the two weights together sum to $R^2 = .30$ (for a complete review, see Tonidandel & LeBreton, 2011).

4.3.3 | Outlier diagnostics and publication bias analyses

Sensitivity analyses refer to assessments of the extent to which findings of analyses remain stable when assumptions or aspects of data or analyses change (Kepes et al., 2013). We conducted a variety of sensitivity analyses on the observed correlations to evaluate the robustness of the results at the distribution level (Kepes & McDaniel, 2015). Meta-analytic data can be considered to be robust to the degree that findings remain stable across various analyses, including sensitivity analyses. Unless otherwise stated, all sensitivity analyses were conducted in *R* using the metafor package and with the recommended random-effects (RE) estimation model. First, we calculated the RE meta-analytic estimates. Next, we conducted one sample removed analyses to examine the influence of each individual sample on the meta-analytic results (Borenstein, Hedges, Higgins, & Rothstein, 2009). Then, we used several publication bias analyses (e.g., trim and fill, selection models; precision-effect test, precision effect estimate with standard error [PET-PEESE]; probability of the chi-square test of excess significance [P-TES]) to triangulate the location of the “true” mean effect size estimate (Kepes et al., 2012; Orliczky, 2011); to identify the possible range of results (i.e., mean correlations) rather than relying on a single estimate. As recommended, we performed the trim and fill analysis with the fixed-effects model and the *L* estimator (Kepes et al., 2012; Sutton, 2005).

We conducted selection model analyses using the a priori *p*-value cut-points to model moderate and severe instances of publication bias suggested by Vevea and Woods (2005). In addition, we conducted PET-PEESE analyses (Stanley & Doucouliagos, 2014) and tests of excess significance (P-TES; Francis, 2014; Ioannidis & Trikalinos, 2007). In contrast to the other analyses, P-TES does not provide an effect size estimate that is adjusted for publication bias. A distribution of effects with a probability of less than .1 is typically considered to lack credibility (Francis, 2014).

To identify outliers, we used Viechtbauer and Cheung's (2010) comprehensive battery of outlier and influence diagnostics. Our obtained results are presented with and without identified outliers, and we only assess the presence of bias

TABLE 1 The effectiveness of tipping compensation practices

| | <i>K</i> | <i>N</i> | \bar{r} | <i>SD_r</i> | $\hat{\rho}$ | <i>SD_ρ</i> | <i>CV_{LL}</i> | <i>CV_{UL}</i> | <i>CI_{LL}</i> | <i>CI_{UL}</i> | %Var |
|--|----------|----------|-----------|-----------------------|--------------|-----------------------|------------------------|------------------------|------------------------|------------------------|-------|
| Common predictors of percentage of a bill tipped | | | | | | | | | | | |
| Service quality | 74 | 12,271 | .20 | .22 | .22 | .24 | -.07 | .50 | .16 | .27 | 11.0 |
| Food quality | 39 | 3,900 | .09 | .38 | .09 | .38 | -.38 | .56 | -.03 | .21 | 7.0 |
| Frequency of patronage | 36 | 3,027 | .06 | .13 | .06 | .13 | -.03 | .16 | .02 | .11 | 67.0 |
| Party size | 48 | 4,805 | -.11 | .14 | -.11 | .14 | -.23 | .01 | -.15 | -.07 | 53.30 |
| Service quality–percentage of a bill tipped: Moderator subgroups | | | | | | | | | | | |
| Tip intensity | | | | | | | | | | | |
| High (absolute amount) | 4 | 868 | .19 | .04 | .20 | .05 | .20 | .20 | .13 | .26 | 99.90 |
| Low (absolute amount) | 42 | 5,899 | .10 | .17 | .11 | .18 | -.09 | .30 | .05 | .16 | 26.08 |
| High (relative amount) | 7 | 1,621 | .12 | .11 | .13 | .12 | .01 | .25 | .04 | .22 | 33.80 |
| Low (relative amount) | 43 | 7,367 | .24 | .27 | .25 | .28 | -.10 | .60 | .17 | .34 | 7.48 |
| Nationality | | | | | | | | | | | |
| USA | 68 | 11,432 | .21 | .23 | .22 | .24 | -.07 | .51 | .17 | .28 | 10.75 |
| International | 6 | 839 | .11 | .14 | .12 | .15 | -.03 | .27 | .01 | .24 | 37.21 |
| Design | | | | | | | | | | | |
| Lab study | 2 | 145 | .19 | .18 | .21 | .19 | .03 | .39 | -.04 | .46 | 41.97 |
| Quasi-experimental field study | 6 | 1,111 | .21 | .14 | .23 | .14 | .07 | .39 | .12 | .34 | 27.18 |
| Field study | 63 | 10,348 | .20 | .24 | .22 | .25 | -.09 | .52 | .15 | .28 | 10.13 |
| Rating source | | | | | | | | | | | |
| Customer | 69 | 10,221 | .24 | .22 | .26 | .24 | -.02 | .54 | .20 | .31 | 12.35 |
| Other | 5 | 2,050 | .00 | .06 | -.00 | .06 | -.04 | .04 | -.05 | .05 | 76.46 |
| Effect size type | | | | | | | | | | | |
| Correlation | 64 | 9,904 | .13 | .13 | .14 | .14 | -.00 | .27 | .10 | .17 | 38.62 |
| Converted effect size | 10 | 2,367 | .51 | .26 | .55 | .28 | .19 | .90 | .37 | .72 | 3.36 |

Notes: *k* = number of independent samples; *N* = total sample size; \bar{r} = sample-size-weighted mean observed correlation; *SD_r* = sample-size-weighted observed standard deviation of correlations; $\hat{\rho}$ = mean true-score correlation (corrected for unreliability in service quality measure); *SD_ρ* = standard deviation of corrected correlations; *CV_{LL}* and *CV_{UL}* = lower and upper bounds, respectively, of the 80% credibility interval; *CI_{LL}* and *CI_{UL}* = lower and upper bounds, respectively, of the 95% confidence interval around the mean true-score correlation; %Var = percentage of variance attributable to statistical artifacts. The correlations between service quality and other predictors of percentage of the bill tipped are available upon request. These correlations were used in the relative weights analyses.

in distributions consisting of at least 10 samples. The obtained results on distributions without outliers tend to be more robust and credible (Kepes & McDaniel, 2015; Viechtbauer & Cheung, 2010). For more details regarding the handling of outliers in this study, see Appendix B (https://osf.io/rp86m/?view_only=35c20ddae75b49509143cd45946b82eb).

5 | RESULTS

5.1 | Tests of hypotheses

The main results of the meta-analytic review are presented in Table 1. We began by testing Hypothesis 1, which stated that there would be a positive relation between service quality and percentage of the bill tipped. In general, there was support for a small to moderate magnitude relation ($\hat{\rho} = .22$, *k* = 74, *N* = 12,271, *SD_ρ* = .24). We also compared the predictive validity of service quality ratings relative to other common predictors of percentage of a bill tipped. We

identified that food quality ($\hat{\rho} = .09, k = 39, N = 3,900$) and frequency of patronage ($\hat{\rho} = .06, k = 36, N = 3,027$) were positively related to percentage of the bill tipped as well. Dining party size was negatively correlated with the percentage tipped ($\hat{\rho} = -.11, k = 48, N = 4,805$).

We next tested Hypothesis 2, which stated that the relation between service quality and percentage tipped would be stronger in cases where tip intensity was higher in either absolute (2a) or relative terms (2b). To determine tip intensity in an absolute sense, we classified all tips that exceeded a flat \$7.25, the current federal minimum wage per hour, as high intensity and all tips below as low intensity. We reasoned that any tip that provides a server with an entire hour's pay from waiting on just one table to be high in intensity. The results of the analyses show there was a moderate magnitude difference between the high-intensity tip amount ($\hat{\rho} = .20, k = 4, N = 868, 95\% \text{ CI } [.13, .26]$; see Table 1) and the low-intensity amount ($\hat{\rho} = .11, k = 42, N = 1,621, 95\% \text{ CI } [.05, .16]$) in the absolute case (Hypothesis 2a). Tip intensity in a relative sense was determined by any tips that exceeded the 20%-of-the-bill-tipped threshold, which tends to be the norm in the United States (Margalioth, 2006). Counter to Hypothesis 2b, we found that there was a larger magnitude relation in the low-intensity condition ($\hat{\rho} = .25, k = 43, N = 7,367, 95\% \text{ CI } [.17, .34]$) than in the high-intensity condition ($\hat{\rho} = .13, k = 7, N = 1,621, 95\% \text{ CI } [.04, .22]$). It is possible that the smaller magnitude relation in the high-intensity condition (as compared to low intensity) may be due to the influence of other factors besides simply service quality. For example, factors such as high food quality or certain personalities (e.g., agreeable people) may lead customers to tip a large percentage of the bill, which creates a high-intensity situation. Continuing with the personality example, agreeable people are kind and cooperative (Barrick & Mount, 1991); as a result, they may feel it necessary to provide servers with large tips. However, if the service quality was poor or mediocre, the relationship between service quality and tip percent would be weak. The results of Hypothesis 2 may be interpreted with caution because of potential range restriction issues as the low-intensity group may be bound by tips of zero and thus have lower variation than the high tip intensity group (and also interpreted carefully because of the smaller sample size in the high tip intensity distribution).

Hypothesis 3 predicted that service quality has greater relative importance than other predictors of percentage of a bill tipped. Moreover, we proposed that the variance in tip percentage explained is mostly attributable to service quality. Aligned with our prediction, service quality showed greater relative importance by large margins over frequency of patronage (H3a; 93.5% vs. 6.5%), dining party size (H3b; 80.2% vs. 19.8%), and food quality (H3c; 91.6% vs. 8.4%). Hence, there was support for Hypothesis 3.

In addition to testing the hypotheses, we sought to answer several research questions. First, we asked whether the country in which a study was conducted moderated the relation between service quality and percentage of a bill tipped. We found that, in fact, studies conducted in the United States reported a stronger relation ($\hat{\rho} = .22, k = 68, N = 11,432, 95\% \text{ CI } [.17, .28]$) than studies conducted in other countries ($\hat{\rho} = .12, k = 6, N = 839, 95\% \text{ CI } [.01, .24]$). There was an apparent practical difference ($\hat{\rho} = .22$ vs. $.12, \Delta = 45\%$). However, given the size of the second meta-analytic distribution ($k = 6$), we urge caution when interpreting the results and suggest that new analyses be conducted as more data become available.

Our second and third research questions related to demographic variables of the customer and the server. We asked whether demographic characteristics of customers could be used to predict the percentage of a bill tipped. We found that White customers tended to tip more than racial minorities (Cohen's $d = -.63, k = 5, N = 2,537$). In terms of age, there was only a very small, negative correlation between customers' age and percentage tipped ($\hat{\rho} = -.03, k = 3, N = 623$). With regard to male and female customers, the former tended to tip a larger percentage of the bill (Cohen's $d = -.11, k = 10, N = 1,349$). We also asked whether demographics of the server could predict the percentage tipped. Our results indicated that White servers tended to be tipped more than minority servers (Cohen's $d = .16, k = 3, N = 1,062$), and that female servers tended to be tipped more than male servers (Cohen's $d = .06, k = 9, N = 1,394$). Again, we urge caution when interpreting results from small distributions of samples.

Our fourth question related to the potential moderating role of methodological variables such as study design, rating source, and statistics reported in the original article. We found that lab studies ($\hat{\rho} = .21, k = 2, N = 145, 95\% \text{ CI } [-.04, .46]$), quasi-experimental field studies ($\hat{\rho} = .23, k = 6, N = 1,111, 95\% \text{ CI } [.12, .34]$), and field studies ($\hat{\rho} = .22, k = 63, N = 10,348, 95\% \text{ CI } [.15, .28]$) showed similar magnitude relations. Given these results, it is apparent that the groups were

not different from each other from a practical perspective ($\hat{\rho} = .21$ vs. $.23$ vs. $.22$). Yet again, one of these distributions is very small, and we urge caution when interpreting the results.

Our results also showed that studies that used customers' ratings of service quality reported a much stronger relation ($\hat{\rho} = .26$, $k = 69$, $N = 10,221$, 95% CI [.20, .31]) than studies using other sources (a server's perception of service quality) to measure service quality ($\hat{\rho} = .00$, $k = 5$, $N = 2,050$, 95% CI [-.05, .05]). Here, there is a noticeable practical difference between the groups ($\hat{\rho} = .26$ vs. $.00$). Finally, we found that studies that reported a correlation coefficient tend to have a much weaker relation ($\hat{\rho} = .14$, $k = 64$, $N = 11,432$, 95% CI [.10, .17]) than those for which we converted the effect size (e.g., Cohen's d) to a correlation ($\hat{\rho} = .55$, $k = 10$, $N = 2,367$, 95% CI [.37, .72]). In this condition, there is a very apparent difference in means between the two conditions ($\hat{\rho} = .14$ vs. $.55$, $\Delta = 293\%$) likely due to the presence of outliers after which the converted effect size estimate drops to $\hat{\rho} = .12$ (outliers removed: $r = .98$, $n = 84$; $r = .66$, $n = 1,600$).

5.2 | Outlier diagnostics and publication bias analyses

Table 2 presents the results of our outlier diagnostics (one sample removed analysis and Viechtbauer and Cheung's (2010) multifaceted influence diagnostics) and publication bias analyses. The first three columns display general information about the meta-analytic distribution (i.e., the name of the distribution analyzed, k , and N). Column 4 contains the weighted mean observed correlation calculated in the Hedges and Olkin meta-analytic tradition. For comparison purposes, in column 5, we present the meta-analytic estimates in the Schmidt and Hunter (2015) meta-analytic tradition once measurement error had been corrected in the service quality variable ($\hat{\rho}$). Column 6 presents the I^2 statistic, which indicates the ratio of true heterogeneity to total variation. Column 7 contains tau (τ), the between-sample standard deviation. In column 8, we present the results of the one-sample removed analysis (Borenstein et al., 2009), including the minimum and maximum as well as the median weighted mean observed correlation. In general, the obtained results illustrate that the results are relatively robust prior to the removal of outliers but that robustness increases with the removal of the outliers (the conclusions drawn from the analyses remains relatively stable).

Next, we present in columns 9–11 (Table 2) the results of the trim and fill analyses. The results illustrate that there exists asymmetry and effect sizes are imputed to the right of the respective distribution to create symmetry. This finding might be counter to what one would expect to see if publication bias was present, but Kepes, Banks, and Oh (2014) provided an explanation for this type of publication bias. With the removal of identified outliers, the number of imputed samples needed to achieve a symmetrical distribution is substantially reduced, suggesting that some of the data points have a greater influence on the shape of the distribution. In columns 11 and 12 (Table 2), we present the results of the moderate and severe selection model analyses. The results of the moderate selection model illustrate the potential for negligible publication bias. The severe selection models show the potential for publication bias under an assumption of severe publication bias. In our analyses, there were several instances where this method provided noncredible results due to inflated variance estimates (Kepes et al., 2012). Finally, the last two columns present the PET-PEESE and the P-TES.

The results of the sensitivity analyses illustrate that the degree of bias ranges from negligible (e.g., trim and fill for "total [wo]," $k = 68$) to moderate (PET-PEESE for "Customer [wo]," $k = 66$) and severe (e.g., moderate selection model for "USA" $k = 68$). Before the removal of the identified outliers, most publication bias methods indicated the presence of noticeable bias (i.e., $>20\%$, Kepes et al., 2012), suggesting that the meta-analytic means tended to be misestimated. However, after the removal of the outliers, the degree of publication bias was often markedly reduced. In addition, for some analyzed distributions, especially before the removal of outliers, a few publication bias methods (e.g., severe selection models and PET-PEESE) provided nonsensical results. Yet, after the removal of the identified outliers, fewer noncredible mean estimates were obtained and the obtained ones tended to converge. In fact, after the removal of the identified outliers, our findings tend to suggest that publication bias was mostly negligible. Therefore, heterogeneity due to outliers had been mistakenly attributed to publication bias by the publication bias methods. Taken together, the results of our sensitivity analyses illustrate that our results, after the removal of outliers, are largely robust. Hence, one can have greater confidence in the obtained results after the removal of outliers than before their removal (Kepes &

TABLE 2 Robustness checks

| | k | N | F̄ | F̂ | I ² | τ | osr | Trim & fill | | Selection models | | | PET-PEESE | | Excess significance |
|---|----|--------|------|-----|----------------|------|------------------|-------------|----|---------------------|--------------------|--------------------|--------------------|-------|---------------------|
| | | | | | | | | FPS | ik | t&f F̄ _o | sm F̄ _o | sm F̄ _o | pp F̄ _o | P-TES | |
| Service quality—percentage of a bill tipped | | | | | | | | | | | | | | | |
| Total | 74 | 12,271 | .18 | .22 | 93.5 | 0.30 | 0.15, 0.19, 0.18 | R | 30 | .29 | .11 | - | .85 | .10 | |
| Total (wo) | 68 | 8,775 | .14 | .15 | 24.6 | 0.05 | 0.13, 0.14, 0.14 | - | 0 | .14 | .12 | - | .16 | .73 | |
| Tip intensity | | | | | | | | | | | | | | | |
| Low (absolute amount) | 42 | 5,899 | .17 | .11 | 91.4 | 0.29 | 0.12, 0.18, 0.17 | R | 18 | .31 | .06 | - | .90 | .76 | |
| Low (absolute amount) (wo) | 33 | 2,493 | .15 | .15 | 0.0 | 0.0 | 0.14, 0.15, 0.15 | R | 4 | .16 | .13 | .09 | .20 | .66 | |
| Low (relative amount) | 43 | 7,367 | .21 | .25 | 95.8 | 0.38 | 0.17, 0.22, 0.21 | R | 17 | .33 | .12 | - | .88 | .03 | |
| Low (relative amount) (wo) | 34 | 2,645 | .16 | .17 | 0.0 | 0.0 | 0.15, 0.17, 0.16 | R | 1 | .16 | .14 | .12 | .20 | .85 | |
| Nationality | | | | | | | | | | | | | | | |
| USA | 68 | 11,432 | .19 | .22 | 93.9 | 0.31 | 0.16, 0.19, 0.19 | R | 28 | .29 | .11 | - | .85 | .11 | |
| USA (wo) | 55 | 5,043 | .15 | .16 | 1.92 | 0.01 | 0.14, 0.15, 0.15 | - | 0 | .15 | .13 | .10 | .18 | .91 | |
| Design | | | | | | | | | | | | | | | |
| Field study | 63 | 10,348 | .18 | .22 | 94.3 | 0.33 | 0.15, 0.19, 0.18 | R | 26 | .29 | .10 | - | .86 | .21 | |
| Field study (wo) | 56 | 5,900 | .14 | .15 | 14.3 | 0.04 | 0.14, 0.15, 0.14 | - | 0 | .14 | .13 | .08 | .17 | .66 | |
| Rating source | | | | | | | | | | | | | | | |
| Customer | 69 | 10,221 | .19 | .26 | 93.1 | 0.31 | 0.16, 0.20, 0.19 | R | 27 | .29 | .12 | - | .88 | .20 | |
| Customer (wo) | 66 | 8,271 | .15 | .16 | 41.9 | 0.08 | 0.15, 0.16, 0.15 | R | 6 | .17 | .13 | - | .19 | .63 | |
| Effect size type | | | | | | | | | | | | | | | |
| Correlations | 64 | 9,904 | .15 | .14 | 54.7 | 0.11 | 0.14, 0.16, 0.15 | R | 3 | .16 | .12 | - | .15 | .63 | |
| Correlations (wo) | 61 | 8,174 | .14 | .15 | 34.0 | 0.06 | 0.14, 0.15, 0.14 | - | 0 | .14 | .12 | - | .17 | .40 | |
| Converted effect size | 10 | 2,367 | 0.39 | .55 | 97.9 | 0.59 | 0.20, 0.42, 0.41 | R | 3 | 0.50 | - | - | .94 | 1.00 | |
| Converted effect size (wo) | 8 | 683 | 0.11 | .12 | 0.00 | 0.00 | - | - | - | - | - | - | - | - | |

Notes: wo = without outliers; k = number of correlation coefficients in the analyzed distribution. Publication bias analyses were not conducted for distributions with less than k = 20. k = number of independent samples; N = total sample size; F̄_{one} = random-effects weighted mean observed correlation in the Hedges and Olkin meta-analytic tradition; F̂ = mean true-score correlation in the Hunter and Schmidt meta-analytic tradition (corrected for unreliability in service quality measure); I² = ratio of true heterogeneity to total variation; τ = between-sample standard deviation; osr = one-sample removed, including the minimum and maximum effect size and the median weighted mean observed correlation. Trim and fill = trim and fill analysis; FPS = funnel plot side (i.e., side of the funnel plot where samples were imputed; L = left, R = right); ik = number of trim and fill imputed samples; t&f F̄_o = trim and fill adjusted observed mean (the weighted mean of the distribution of the combined observed and the imputed samples); sm F̄_o = one-tailed moderate selection models adjusted observed mean (and its variance); sm_s F̄_o = one-tailed severe selection model's adjusted observed mean; PET-PEESE = precision-effect test-precision effect estimate. pp F̄_o = PET-PEESE adjusted observed mean. P-TES = the probability of the chi-square test of excess significance.

McDaniel, 2015). For our data, outliers tended to have an adverse effect on many of the meta-analytic results because the results were often noticeably smaller in magnitude after outlier removal. We recommend that future research refer to these results rather than the ones from the distributions including outliers; they are more robust and should have greater credibility.

6 | DISCUSSION

Compensation, especially pay-for-performance, is a vital management practice for organizations in general, and human resource professionals in particular. To the former, employee compensation is a large expense of the overall organizational budget; efficient use of the funds is imperative for success (Gerhart, Milkovich, & Murray, 1992; Gerhart & Rynes, 2003). To the latter, compensation facilitates the job performance of employees; it is a crucial component of attracting, motivating, and retaining high performers. Despite the practical interest and applicability of compensation to guide employee behaviors, it is a topic that remains underresearched within the academic literature (Deadrick & Gibson, 2007; Gupta & Shaw, 2014). To address this gap in the literature, we applied expectancy and agency theory to understand the structure of compensation practices in the service industry. In doing so, we advanced the literature in several ways.

First, we advanced the management and applied psychology literature on compensation. We applied formal agency theory (Grossman, & Hart, 1983; Holmström, 1979) and expectancy theory (Adams, 1963). By applying these two theories, one defines performance in a way that is aligned with the interests of the owner (Oyer & Schaefer, 2011), which, in our case, are also essentially the interests of the customers. This can help to identify and potentially minimize the display of servers' self-interested behaviors that are not in the owner's interest. Such an approach accounts for the high costs associated with ensuring that employees behave in ways that align with the organization's goals. Agency theory assumes that principal and agent interests are not necessarily aligned and explains how compensation practices can be used to improve the alignment of interests between the two parties (i.e., variation in a monetary payment is linked to the principal's benefit; Oyer & Schaefer, 2011). By using agency theory and expectancy theory together (or in combination), we sought to achieve a more holistic understanding of how to structure effective compensation systems.

Second, we contribute to the literature by partially testing the theoretical framework we advance. Particularly, we investigated how the compensation practice of tipping can be used to address the principal-agent problem and incentivize servers to provide high-quality service (which is in the best interest of the owner). Previous research has questioned the extent to which tipping represents a pay-for-performance compensation practice. The results of this research suggest that the relation between service quality and percentage tipped varies in both direction and magnitude of the effect (e.g., Barkan & Israeli, 2004; Lynn et al., 2008; Lee, 2015).

Using meta-analytic techniques, we examined the extent to which high levels of service quality (performance) were appropriately rewarded in the form of tips (pay). Our results challenge both theory and conventional wisdom. We found a small to moderate correlation between tipping and service quality of .15 ($\hat{\rho} = .22$ with outliers; see Table 2). Based on the results of our sensitivity analysis, we view this as the best estimate of the service quality-tipping relation to date. Given this result, one might argue that tipping is an ineffective pay-for-performance compensation practice as it might not motivate employees to perform at a high level. That is, performing highly (i.e., providing high service quality) does not strongly equate to higher pay (in the form of tips). Therefore, the small to moderate magnitude relation suggests that tipping practices could be a relatively ineffective means of addressing the principal-agent problem. For instance, customers can tip service providers even when the service quality offered is subpar, which means that service providers are being rewarded for behaviors that are not in the owner's best interest.

Yet, when we compare tipping to other pay-for-performance practices, the magnitude of the service quality-percentage tipped relation is perhaps not necessarily so small. Jenkins, Mitra, Gupta, and Shaw (1998) found a meta-analytic estimate for performance quality ($\rho = .08$) and performance quantity ($\rho = .34$) that are comparable in size to our estimate with outliers ($\rho = .22$).¹ The service quality-percentage of a bill tipped relation before the removal of

outliers is also similar in magnitude to meta-analytic estimates from Garbers and Konradt (2014) for both quantitative and qualitative performance. Therefore, a mean effect size of .22 ($\rho = .15$ without outliers) is perhaps not so small in comparison to other relevant meta-analytic estimates of pay-for-performance practices and systems. Finally, if we consider the Bosco, Aguinis, Singh, Field, and Pierce (2015) study on correlation effect size benchmarks, which also did not account for outliers or publication bias, an effect size magnitude of .22 is not small by any means, especially not for a relation between an objective characteristic and performance. Still, we draw the general conclusion that the apparent effectiveness of tipping as a compensation practice is limited.

In addition, the results of the relative weights analyses illustrate the relative importance of service quality in predicting tip percentage. Although factors such as patronage frequency, dining party size, and food quality have been shown to be related to the percentage of a bill tipped (Lynn, 2006; Lynn & McCall, 2000; Rogelberg et al., 1999), our findings suggest that service quality shows dominance over those other factors in predicting the percentage of a bill tipped. Understanding the proportionate contribution of predictors aids in the development and refinement of theory (Tonidandel & LeBreton, 2011). The findings suggest that, out of the predictors explored, customers consider service quality the most when deciding the percentage of a bill tipped. Despite the fact that the magnitude of the effect size is not large, tipping serves to encourage and reward service quality above other factors. However, it is important to note the limitations of relative weights analysis; like regression and virtually all statistical techniques, it is subject to the effects of sampling and measurement error as well as model misspecification (Tonidandel & LeBreton, 2011).

Third, we explored possible explanations for the small to moderate relation between tipping and service quality. Therefore, individual differences in tipping may exist based on age, race, gender, and a myriad of other variables (Brewster & Lynn, 2014; Brewster & Mallinson, 2009; Lynn & Thomas-Haysbert, 2003; Parrett, 2015). Indeed, our findings provided evidence that some demographic differences of customers and servers affect the percentage of a bill that is tipped. These factors are likely to somewhat undermine the effectiveness of tipping as a pay-for-performance practice. As these factors are beyond the control of individual service providers, but influence the amount that is tipped, they diminish the motivating potential of tipping.

It is also interesting to note that the magnitude of the service quality–percentage tipped relation remains largely the same when considering moderators, such as study artifacts (e.g., design, source of service quality ratings), the country in which the study was conducted, and the intensity of the tip. It was particularly discouraging that the magnitude of the relation did not seem to increase when the tipping intensity is high (this was true in a relative, but not an absolute sense). This finding stands in contrast to motivational theories, such as expectancy theory, which argue that employees are more motivated to perform at a high level if they expect their performance will result in a highly valuable reward (Hackman & Porter, 1968). In the case of tipping, our obtained results indicate this is not necessarily the case, at least when using a relative measure for tipping intensity. However, some of our distributions contained relatively few samples (i.e., $k < 10$). Caution should be displayed when making inferences based on results from such distributions. Yet, we also note that conclusions based on a meta-analytic sample size of five or seven should be substantially more trustworthy than conclusions from one or two primary studies.

In addition, although the overall conclusion regarding the relationship between service quality and tipping remains stable, this study extends the prior work on tipping, specifically the meta-analysis conducted by Lynn and McCall (2000) in several important ways. First, we provide an updated and more comprehensive summary of the tipping literature; we included 74 independent samples with a total N of 12,271 (compared to $k = 13$ and $N = 2,547$ in Lynn and McCall's (2000) original meta-analysis). Second, while Lynn and McCall (2000) mention that the practice of tipping is theoretically aligned with economic rationalizations and equity theory, this study develops and advances a theoretical framework, based on expectancy and agency theory, to explain tipping as a compensation practice. This study also conducts a number of additional analyses compared to Lynn and McCall (2000), which helps to more comprehensively understand the literature. For example, this study runs a series of moderator analyses in an attempt to explain inconsistent findings in the relationship between service quality and tipping. We also explore the relative importance of service quality compared to other factors that have been shown to influence tipping. In addition, we conduct sensitivity analyses to better understand the influence of outliers and publication bias on the obtained results.

6.1 | Practical Implications

A recent article in the *Wall Street Journal* entitled “*Is it time to end tipping?*” highlights the practical significance of this research (Zagorsky, 2016). The article presents arguments for and against a recent trend in the restaurant industry to ban tipping. The evidence presented in our study supports the latter line of arguments; our results show that tipping may not be an overly effective pay-for-performance practice and owners should seriously consider disbanding the practice, or in the least, conduct local validation studies in their own restaurants. Thus, it could be necessary to develop other ways to structure compensation practices in service contexts so that organizations can attract, motivate, and retain high performers. However, developing such alternatives may prove challenging.

One alternative to the traditional tipping approach is the implementation of tip pooling practices. Unfortunately, there are insufficient data to evaluate the effectiveness of such practices. Theoretical arguments, however, would suggest that such practices might not be an effective alternative. Tip pooling is very similar to group- or team-level pay-for-performance practices and systems. Applying the tenants of expectancy theory, servers working under a tip pooling system would have a lower motivational force because performance is measured at the unit level. Because any individual has less control of group-level performance than individual-level performance, individual expectancy perceptions should be noticeably lower for unit-level pay-for-performance practices. In the event that an employee works harder than a coworker does, an employee would likely expect to be rewarded to a greater degree than the coworker. However, if all tips are pooled, high-performing servers could experience equity distress (Huseman, Hatfield, & Miles, 1987) and subsequently reduce their effort. Consequently, all servers may provide lower levels of service quality. Such a practice might not dramatically reduce the percentage of money tipped given the current magnitude relation between service quality and tip percentage. Hence, the practice of tipping would still not serve as a quality control mechanism (although pooling tips may help solve some of the potential discrimination problems with tipping).

A second alternative approach could be the implementation of automatic service charges, which are used in some European countries (Margalioth, 2006; Segrave, 1998). In such cases, restaurant servers are paid a standard wage similar to employees in other service contexts. This type of practice does not rely on customers to reward the service quality delivered by servers. Thus, it would be more of an egalitarian compensation system. Given that tipping is a questionable pay-for-performance practice, owners may just outsource risks and create precarious employments. Fixed payment, like encountered in some European countries, may be a direct reaction to such observations, and may reflect the trade-off between economic efficiency and social security. Yet, under such a system, equity and expectancy theory arguments suggest that servers should be unlikely to be motivated to provide high levels of service because tips (pay) are unrelated to service quality (performance) that servers provide. Further, agency theory would suggest that the principal-agent problem still exists because the interests between individual service providers and owners are misaligned due to difficulties in monitoring behaviors. We encourage future research to explore alternative compensation structures that might address the principal-agent problems and motivating dynamics in service contexts.

In addition to determining equity, comparisons and interactions among peers working in the service industry is a particularly interesting area for discussion and an area ripe for future research to explore. For example, restaurants employ a number of different positions, including servers, bussers, bartenders, and kitchen staff, among others. Each position has an important role (e.g., providing food and drink, clean tables and atmosphere, friendly service) in providing customers with the end product. It is possible that the roles and behaviors of others influence the expectancy of servers and subsequently their decision to provide high-quality service or not. As mentioned previously, if the kitchen or bar is understaffed, or the hostesses are not friendly when greeting customers, servers may feel less motivated to provide high-quality service. That is, customers may not be willing to tip servers for their service if the other aspects of the service experience are not of high quality (e.g., delayed or cold food, rude interactions). Thus, it would be interesting to explore how the behaviors of other restaurant employees might influence servers' decisions to provide high-quality service.

In addition to the technical challenges of designing alternative compensation systems, other barriers to revising tipping practices exist. First, despite being strongly opposed initially, tipping appears to be institutionalized in the United States (Margalioth, 2006). Most likely, it would take a noticeably more effective alternative that benefits multiple

stakeholders for change to occur. In addition, some have pointed to the potential for tax evasion as an incentive for both principals and agents to underreport the amount of tips given as compensation (Lynn, 2006). That is, tipping allows individual service providers to achieve higher income while reducing employers' payroll and tax burden. This is primarily true in the case of cash payments of tips. Companies and their employees have to pay less in taxes if such wages (i.e., tips) go unreported. As many restaurant servers have a base salary below the federal minimum wage, there could be a strong incentive for them to not report all cash tips and for owners to not monitor this activity closely. In sum, although there seems to be some evidence against the use of tipping as a pay-for-performance practice, it is unlikely that changes will occur any time soon to such an entrenched deep-rooted practice.

Finally, consistent with past research (Brewster, 2015; Brewster & Lynn, 2014; Brewster et al., 2009), we found demographic differences in the amount that servers are tipped. Hence, this study contributes to the notion that discrimination is possible in tipping practices. Although we could not control for as many factors as Brewster and colleagues, we do provide estimates that have been corrected for artifactual variance. As organizations strive to provide fair pay for their employees, we encourage practitioners to consider removing tipping as a practice from their restaurants or at least individually evaluate the extent to which tipping practices lead to fair compensation among their employees.

6.2 | Limitations and Future Directions

As with any meta-analytic review, this study was limited by the primary studies that were available for inclusion. However, we see these limitations as opportunities for future research. First, although compensation may be a driver of motivation and high performance (e.g., Heneman & Werner, 2005; Oyer, & Schaefer, 2011), it is important to note that it is not the only source of motivation. That is, servers may be intrinsically motivated to provide high-quality service or face social pressures from their coworkers or others. Although not the focus of this study, an investigation looking at other drivers of high-quality service would complement our results regarding tips.

Second, many of the studies did not report reliability estimates, especially for the measures of service quality, owing in part to the use of single-item measures. Reliability estimates were calculated based on those estimates that were provided, but because of the problems with single-item measures, we encourage future researchers to use multi-item scales. Furthermore, relevant information (e.g., reliability estimates) should always be reported in primary studies (American Psychological Association, 2008). Fourth, we were not able to test the contingency factors represented in box 6 of our model (Figure 1) because the primary study data were insufficient. For example, we were not able to consider if tipping helped to retain talent in service jobs. In order to understand the boundary conditions of our theoretical framework further, research should examine potential moderators of the expectancy relations. Contingency factors such as the ones outlined in our study can further advance and refine the scope and applicability of both expectancy and agency theory. Also, the number of studies for one set of these distributions (i.e., high tip intensity) is relatively small, making definite inferences questionable.

A third limitation is that the vast majority of research on tipping has been conducted in the restaurant context. This is not a surprise given how frequently people eat at restaurants and leave a tip. Still, we argue that tipping should also be investigated in other service contexts to determine whether the findings of our study generalize. Specifically, it would be interesting to explore if and to what extent organizations, industries, and countries differ in how they structure and implement tipping as a compensation practice. That is, do organizations vary in how they use tipping as a compensation practice? Or are the differences mostly occupational or industry driven? Finally, we wish to reiterate the results of hypothesis three relating to tip intensity should be interpreted with caution because of potential range restriction issues as the low-intensity group may be bound by tips of zero and thus have lower variation than the high tip intensity group.

7 | CONCLUSION

Some have argued that a principal-agent problem exists in service contexts where monitoring of employees is difficult and that the practice of tipping is one means to overcome such a challenge (Lynn et al., 2011). The current meta-analytic

review sought to evaluate the extent to which tipping is an effective pay-for-performance practice. We found only a small relation between service quality and percentage of a bill tipped in the restaurant context. We did find some evidence that demographic variables of servers and customers predict the percentage of a bill tipped, as did other characteristics, such as food quality, frequency of patronage, and party size. This suggests that the implementation of tipping practices is not necessarily effective in rewarding performance and aligning interests of principals and agents. In general, we conclude that tipping is a limited pay-for-performance practice. Our theoretical framework can be applied to principal-agent problems in other service contexts where monitoring is challenging. We encourage future research to continue investigating how to structure compensation systems to not only capitalize on what we have learned from expectancy and other motivation theories but to think more specifically about characteristics of compensation practices.

ACKNOWLEDGMENTS

Data contributed by Ofer Azar, Zachary Brewster, Bill Gibson, Michael Lynn, Hana Medler-Lirza, Matthew Parrett, and Jeremy Whaley were greatly appreciated. We wish to thank Steven Rogelberg for his additional comments and insights as well as Michael Lynn for sharing his bibliography on tipping papers. We also desire to thank James Diefendorff and two anonymous reviewers for their developmental comments.

NOTES

¹ Our study is the first that explicitly assessed the effects of outliers and publication bias when examining the meta-analytic effect of a pay-for-performance system (e.g., Jenkins et al. (1998) and Garbers and Konradt (2014) did not perform a comprehensive sensitivity analysis to assess the robustness of their results). Hence, although we view the meta-analytic means after the deletion of outliers as the more accurate and robust estimate (e.g., $\bar{r} = .14$, $\hat{\rho} = .15$; $k = 68$), we refer in this section to the meta-analytic results before outlier removal and publication bias analyses (e.g., $\bar{r} = .20$, $\hat{\rho} = .22$; $k = 74$) as those results are methodologically comparable to the ones reported in previously published meta-analytic reviews on pay for performance.

² Due to space constraints, references and input data for studies included in the meta-analytic review are available in an online repository (https://osf.io/rnmky/?view_only=1995b7c813e54002a3b241d9485198f1)

ORCID

George C. Banks  <http://orcid.org/0000-0002-2261-2935>

REFERENCES²

- Adams, J. S. (1963). Towards an understanding of inequity. *The Journal of Abnormal and Social Psychology*, 67(5), 422–436.
- American Psychological Association. (2008). Reporting standards for research in psychology: Why do we need them? What might they be? *American Psychologist*, 63, 839–851.
- Ayres, I., Vars, F. E., & Zakariya, N. (2005). To insure prejudice: Racial disparities in taxicab tipping. *Yale Law Journal*, 114, 1613–1674.
- Azar, O. H. (2004a). The history of tipping—From sixteenth-century England to United States in the 1910s. *The Journal of Socio-Economics*, 33(6), 745–764.
- Azar, O. H. (2004b). What sustains social norms and how they evolve? The case of tipping. *Journal of Economic Behavior & Organization*, 54(1), 49–64.
- Azar, O. H. (2007a). Do people tip strategically, to improve future service? Theory and evidence. *Canadian Journal of Economics/Revue canadienne d'économique*, 40(2), 515–527.
- Azar, O. H. (2007b). The social norm of tipping: A review. *Journal of Applied Social Psychology*, 37(2), 380–402.
- Azar, O. H. (2009). Incentives and service quality in the restaurant industry: The tipping–service puzzle. *Applied Economics*, 41(15), 1917–1927.
- Azar, O. H. (2010). Do people tip because of psychological or strategic motivations? An empirical analysis of restaurant tipping. *Applied Economics*, 42(23), 3039–3044.
- Azar, O. H. (2011). Business strategy and the social norm of tipping. *Journal of Economic Psychology*, 32(3), 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.
- Banks, G. C., Davis McCauley, K., Gardner, W. L., & Guler, C. E. (2016). A meta-analytic review of authentic and transformational leadership: A test for redundancy. *The Leadership Quarterly*, 27, 634–652.
- Barkan, R., & Israeli, A. (2004). Testing servers' roles as experts and managers of tipping behaviour. *The Service Industries Journal*, 24(6), 91–108.
- Barrick, M. A., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44, 1–26.
- Berrone, P., & Gomez-Mejia, L. R. (2009). Environmental performance and executive compensation: An integrated agency-institutional perspective. *Academy of Management Journal*, 52(1), 103–126.
- Bodvarsson, O. B., & Gibson, W. A. (1997). Economics and restaurant gratuities: Determining tip rates. *American Journal of Economics and Sociology*, 56(2), 187–203.
- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. West Sussex, UK: Wiley.
- Bosco, F. A., Aguinis, H., Singh, K., Field, J. G., & Pierce, C. A. (2015). Correlational effect size benchmarks. *Journal of Applied Psychology*, 100(2), 431–449.
- Brewster, Z. W. (2015). Perceptions of intergroup tipping differences, discriminatory service, and tip earnings among restaurant servers. *International Journal of Hospitality Management*, 46, 15–25.
- Brewster, Z. W., & Lynn, M. (2014). Black-White earnings gap among restaurant servers: A replication, extension, and exploration of consumer racial discrimination in tipping. *Sociological Inquiry*, 84(4), 545–569.
- Brewster, Z. W., & Mallinson, C. (2009). Racial differences in restaurant tipping: A labour process perspective. *Service Industries Journal*, 29(8), 1053–1075.
- Cohen, J. R., & Holder-Webb, L. L. (2006). Rethinking the influence of agency theory in the accounting academy. *Issues in Accounting Education*, 21(1), 17–30.
- Cole, M. S., Walter, F., Bedeian, A. G., & O'Boyle, E. H. (2012). Job burnout and employee engagement: A meta-analytic examination of construct proliferation. *Journal of Management*, 38(5), 1550–1581.
- Conlin, M., Lynn, M., & O'Donoghue, T. (2003). The norm of restaurant tipping. *Journal of Economic Behavior & Organization*, 52(3), 297–321.
- Dalton, D. R., Hitt, M. A., Certo, S. T., & Dalton, C. M. (2007). The fundamental agency problem and its mitigation: Independence, equity, and the market for corporate control. *Academy of Management Annals*, 1(1), 1–64.
- Deadrick, D. L., & Gibson, P. A. (2007). An examination of the research-practice gap in HR: Comparing topics of interest to HR academics and HR professionals. *Human Resource Management Review*, 17(2), 131–139.
- Deckop, J. R., Mangel, R., & Cirka, C. C. (1999). Research notes. Getting more than you pay for: Organizational citizenship behavior and pay-for-performance plans. *Academy of Management Journal*, 42(4), 420–428.
- Eisenhardt, K. M. (1989). Agency theory: An assessment and review. *Academy of Management Review*, 14(1), 57–74.
- Francis, G. (2014). The frequency of excess success for articles in Psychological Science. *Psychonomic Bulletin & Review*, 21(5), 1180–1187.
- Garbers, Y., & Konradt, U. (2014). The effect of financial incentives on performance: A quantitative review of individual and team-based financial incentives. *Journal of Occupational and Organizational Psychology*, 87(1), 102–137.
- Gerhart, B., Milkovich, G. T., & Murray, B. (1992). Pay, performance, and participation (CAHRS Working Paper #92–28). Ithaca, NY: Cornell University, School of Industrial and Labor Relations, Center for Advanced Human Resource Studies.
- Gerhart, B., & Rynes, S. L. (2003). *Compensation: Theory, evidence, and strategic implications*. Thousand Oaks, CA: Sage.
- Gerhart, B., Rynes, S. L., & Fulmer, I. S. (2009). Pay and performance: Individuals, groups, and executives. *Academy of Management Annals*, 3, 251–315.
- Gomez-Mejia, L. R., & Balkin, D. B. (1992). Determinants of faculty pay: An agency theory perspective. *Academy of Management Journal*, 35(5), 921–955.
- Grossman, S. J. & Hart, O. D. (1983). An analysis of the principal-agent problem. *Econometrica*, 51, 7–45.
- Gupta, N., Ganster, D. C., & Kepes, S. (2013). Assessing the validity of sales self-efficacy: A cautionary tale. *Journal of Applied Psychology*, 98(4), 690–700.
- Gupta, N., & Shaw, J. D. (2014). Employee compensation: The neglected area of HRM research. *Human Resource Management Review*, 24(1), 1–4.
- Hackman, J. R., & Oldham, G. R. (1975). Development of the job diagnostic survey. *Journal of Applied Psychology*, 60(2), 159–170.

- Hackman, J. R., & Porter, L. W. (1968). Expectancy theory predictions of work effectiveness. *Organizational Behavior and Human Performance*, 3(4), 417–426.
- Heneman, R. L., & Werner, J. M. (2005). *Merit pay: Linking pay to performance in a changing world*. Greenwich, CT: Information Age Publishing.
- Holmström, B. (1979). Moral hazard and observability. *Bell Journal of Economics*, 10, 74–91.
- Huseman, R. C., Hatfield, J. D., & Miles, E. W. (1987). A new perspective on equity theory: The equity sensitivity construct. *Academy of Management Review*, 12(2), 222–234.
- Ioannidis, J. P., & Trikalinos, T. A. (2007). An exploratory test for an excess of significant findings. *Clinical Trials*, 4(3), 245–253.
- Jacob, N. L., & Page, A. N. (1980). Production, information costs, and economic organization: The buyer monitoring case. *The American Economic Review*, 70(3), 476–478.
- Jenkins, G. D., Mitra, A., Gupta, N., & Shaw, J. D. (1998). Are financial incentives related to performance? A meta-analytic review of empirical research. *Journal of Applied Psychology*, 83, 777–787.
- Johnson, J. W. (2001). Computer program exchange: Rweight: Computing the relative weight of predictors in a regression. *Applied Psychological Measurement*, 25, 342.
- Kahn, R. L., Wolfe, D. M., Quinn, R. P., Snoek, J. D., & Rosenthal, R. A. (1964). *Organizational stress: Studies in role conflict and ambiguity*. New York, NY: John Wiley.
- Kepes, S., Banks, G. C., McDaniel, M. A., & Whetzel, D. L. (2012). Publication bias in the organizational sciences. *Organizational Research Methods*, 15, 624–662.
- Kepes, S., Banks, G. C., & Oh, I.-S. (2014). Avoiding bias in publication bias research: The value of “null” findings. *Journal of Business and Psychology*, 29, 183–203.
- Kepes, S., Delery, J., & Gupta, N. (2009). Contingencies in the effects of pay range on organizational effectiveness. *Personnel Psychology*, 62(3), 497–531.
- Kepes, S., & McDaniel, M. A. (2015). The validity of conscientiousness is overestimated in the prediction of job performance. *PloS One*, 10(10), e0141468.
- Kepes, S., McDaniel, M. A., Brannick, M. T., & Banks, G. C. (2013). Meta-analytic reviews in the organizational sciences: Two meta-analytic schools on the way to MARS. *Journal of Business and Psychology*, 28, 123–142.
- Lawler, E. (1971). *Pay and organization effectiveness: A psychological view*. New York, NY: McGraw Hill.
- Lee, H. E. (2015). Does a server's attentiveness matter? Understanding intercultural service encounters in restaurants. *International Journal of Hospitality Management*, 50, 134–144.
- Lubatkin, M. H., Lane, P. J., Collin, S.-O., & Very, P. (2005). Origins of corporate governance in the USA, Sweden and France. *Organization Studies*, 26(6), 867–888.
- Lynn, M. (1997). Tipping customs and status seeking: A cross-country study. *International Journal of Hospitality Management*, 16(2), 221–224.
- Lynn, M. (2001). Restaurant tipping and service quality: A tenuous relationship. *Cornell Hotel and Restaurant Administration Quarterly*, 42(1), 14–20.
- Lynn, M. (2006). Tipping in restaurants and around the globe: An interdisciplinary review. In M. Altman (Ed.), *Handbook of contemporary behavioral economics: foundations and developments* (pp. 626–643). Armonk, NY: M.E. Sharpe Publishers.
- Lynn, M. (2007). Race differences in restaurant tipping: A literature review and discussion of practical implications. *Journal of Foodservice Business Research*, 9(4), 99–113.
- Lynn, M. (2015a). Explanations of service gratuities and tipping: Evidence from individual differences in tipping motivations and tendencies. *Journal of Behavioral and Experimental Economics*, 55, 65–71.
- Lynn, M. (2015b). Service gratuities and tipping: A motivational framework. *Journal of Economic Psychology*, 46, 74–88.
- Lynn, M., & Grassman, A. (1990). Restaurant tipping: An examination of three 'rational' explanations. *Journal of Economic Psychology*, 11(2), 169–181.
- Lynn, M., & Graves, J. (1996). Tipping: An incentive/reward for service? *Journal of Hospitality & Tourism Research*, 20(1), 1–14.
- Lynn, M., & Gregor, R. (2001). Tipping and service: The case of hotel bellmen. *International Journal of Hospitality Management*, 20(3), 299–303.
- Lynn, M., Jabbour, P., & Kim, W. G. (2012). Who uses tips as a reward for service and when? An examination of potential moderators of the service–tipping relationship. *Journal of Economic Psychology*, 33(1), 90–103.

- Lynn, M., Kwortnik, R. J., & Sturman, M. C. (2011). Voluntary tipping and the selective attraction and retention of service workers in the USA: An application of the ASA model. *The International Journal of Human Resource Management*, 22(9), 1887–1901.
- Lynn, M., & Latane, B. (1984). The psychology of restaurant tipping. *Journal of Applied Social Psychology*, 14(6), 549–561.
- Lynn, M., & Lynn, A. (2004). National values and tipping customs: A replication and extension. *Journal of Hospitality & Tourism Research*, 28(3), 356–364.
- Lynn, M., & McCall, M. (2000). Gratitude and gratuity: A meta-analysis of research on the service-tipping relationship. *The Journal of Socio-Economics*, 29(2), 203–214.
- Lynn, M., & Sturman, M. (2011). Is the customer always right? The potential for racial bias in customer evaluations of employee performance. *Journal of Applied Social Psychology*, 41(9), 2312–2324.
- Lynn, M., Sturman, M., Ganley, C., Adams, E., Douglas, M., & McNeil, J. (2008). Consumer racial discrimination in tipping: A replication and extension. *Journal of Applied Social Psychology*, 38(4), 1045–1060.
- Lynn, M., & Thomas-Haysbert, C. (2003). Ethnic differences in tipping: Evidence, explanations, and implications. *Journal of Applied Social Psychology*, 33, 1747–1772.
- Lynn, M., Zinkhan, G. M., & Harris, J. (1993). Consumer tipping: A cross-country study. *Journal of Consumer Research*, 20(3), 478–488.
- Margalioth, Y. (2006). Case against tipping. *The U. PA. Journal of Labor and Employment Law*, 9, 117.
- Medler-Liraz, H. (2012). Service quality and tipping: The moderating role of the quality of food. *International Journal of Hospitality Management*, 31(4), 1327–1329.
- Mitra, A., Gupta, N., & Jenkins, G. (1997). A drop in the bucket: When is a pay raise a pay raise? *Journal of Organizational Behavior*, 18(2), 117–137.
- Mitra, A., Tenhiälä, A., & Shaw, J. D. (2016). Smallest meaningful pay increases: Field test, constructive replication, and extension. *Human Resource Management*, 55(1), 69–81.
- National Restaurant Association. (2015). Hospitality employee turnover rose in 2014. <http://www.restaurant.org/News-Research/News/Hospitality-employee-turnover-rose-in-2014>
- Orlitzky, M. (2011). How can significance tests be deinstitutionalized? *Organizational Research Methods*, 15, 199–228.
- Oyer, P., & Schaefer, S. (2011). Personnel economics: Hiring and incentives. In O. Ashenfelter & D. Card (Eds.), *Handbook of labor economics* (pp. 1769–1823). Amsterdam/New York: Elsevier.
- Parrett, M. (2015). Beauty and the feast: Examining the effect of beauty on earnings using restaurant tipping data. *Journal of Economic Psychology*, 49, 34–46.
- Pearl, R. B. (1985). *Tipping practices of American households: 1984 (final report to the Internal Revenue Service under Contract TIR 82-21)*. Champaign, IL: Survey Research Laboratory, University of Illinois.
- Rogelberg, S. G., Ployhart, R. E., Balzer, W. K., & Yonker, R. D. (1999). Using policy capturing to examine tipping decisions. *Journal of Applied Social Psychology*, 29(12), 2567–2590.
- Schmidt, F. L., & Hunter, J. E. (2015). *Methods of meta-analysis: Correcting error and bias in research findings* (3rd ed.). Newbury Park, CA: Sage.
- Segrave, K. (1998). *Tipping: An American social history of gratuities*. Jefferson, NC: McFarland.
- 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.
- Stanley, T., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, 5(1), 60–78.
- Sutton, A. J. (2005). Evidence concerning the consequences of publication and related biases. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment, and adjustments* (pp. 175–192). West Sussex, UK: Wiley.
- Tonidandel, S., & LeBreton, J. M. (2011). Relative importance analysis: A useful supplement to regression analysis. *Journal of Business and Psychology*, 26, 1–9.
- Van Eerde, W., & Thierry, H. (1996). Vroom's expectancy models and work-related criteria: A meta-analysis. *Journal of Applied Psychology*, 81(5), 575–586.
- Vevea, J. L., & Woods, C. M. (2005). Publication bias in research synthesis: Sensitivity analysis using a priori weight functions. *Psychological Methods*, 10(4), 428–443.

- Viechtbauer, W., & Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, 1(2), 112–125.
- Vroom, V. H. (1964). *Work and motivation*. New York, NY: Wiley.
- Wood, J. A. (2008). Methodology for dealing with duplicate study effects in a meta-analysis. *Organizational Research Methods*, 11, 79–95.
- Zagorsky, J. (2016). Is it time to end tipping? *The Wall Street Journal*. Retrieved from <http://www.wsj.com/articles/is-it-time-to-end-tipping-1456715614?mg=id-wsj>.

How to cite this article: Banks GC, Woznyj HM, Kepes S, Batchelor JH, McDaniel MA. A meta-analytic review of tipping compensation practices: An agency theory perspective. *Personnel Psychology*. 2018;71:457–478. <https://doi.org/10.1111/peps.12261>