Management Science

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

A Modeling Framework for Tipping in the Presence of a Social Norm

Laurens G. Debo, Ran I. Snitkovsky

To cite this article:
Laurens G. Debo, Ran I. Snitkovsky (2026) A Modeling Framework for Tipping in the Presence of a Social Norm. Management Science 72(4):3236-3255. https://doi.org/10.1287/mnsc.2021.03422

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

MANAGEMENT SCIENCE
Vol. 72, No. 4, April 2026, pp. 3236–3255
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

A Modeling Framework for Tipping in the Presence of a Social Norm

Laurens G. Debo,[a,*] Ran I. Snitkovsky[b]

[a] Tuck School of Business, Dartmouth College, Hanover, New Hampshire 03755; [b] Coller School of Management, Tel Aviv University, Tel Aviv-Yafo 6997801, Israel

*Corresponding author

Contact: laurens.g.debo@tuck.dartmouth.edu, https://orcid.org/0000-0002-2330-7078 (LGD); ransnit@gmail.com, https://orcid.org/0000-0002-0590-0361 (RIS)

Received: October 11, 2021
Revised: November 14, 2023; August 7, 2024
Accepted: August 24, 2024
Published Online in Articles in Advance: August 18, 2025

https://doi.org/10.1287/mnsc.2021.03422

Abstract. Tipping is a complex phenomenon cutting across various stakeholders—firms, customers, and workers. Analyzing the long-run impact of policies related to tipping is therefore challenging. To facilitate such analysis, we develop a modeling framework in which a tipping norm forms endogenously in a market consisting of a firm that offers service to potential customers. Customers choose whether to consume the service or not and, if yes, how much to tip the server afterward. With tipping, customers show appreciation to the server by sharing a fraction of their surplus but also undergo social pressure to comply with the prevailing norm. This tipping norm is shown to evolve endogenously through a dynamic process of sequential market adjustments over time: the average tip in each period determines the tipping norm for the following period, causing the firm to adapt the price and customers to adapt their tips accordingly. Characterization of this equilibrium outcome allows us to derive qualitative results on the long-run impact of different exogenous factors on tipping: we find that the equilibrium tip-to-price ratio increases when customers are more sensitive to social pressure, their range of service valuations spreads out, or they consider the service more valuable. Building on this framework, we further investigate several economic implications of tipping pertaining to social welfare, labor cost, and service quality, thus uncovering incentives and trade-offs to which the tipping mechanism gives rise from the firm, the worker, and the customer's perspectives.

History: Accepted by Karan Girotra, operations management.

Funding: Financial support by The Henry Crown Institute of Business Research in Israel, The Coller Foundation, and the Israeli Science Foundation [Grant 1361/23] is gratefully acknowledged.
Supplemental Material: The online appendix and data files are available at https://doi.org/10.1287/mnsc.2021.03422.

Keywords: decision analysis: theory • economics: game theory and bargaining theory • utility preference: theory

1. Introduction

1.1. Background, Motivation, and Goals

On an average day, roughly 33 million Americans leave tips in restaurants they visit (Lynn 2006). With its annual magnitude in U.S. restaurants estimated (as of 2010) at about $47 billion (Azar 2011), tipping has become, over time, a highly widespread practice, further extending beyond the hospitality sector to include occupations such as barbers, taxi drivers, and delivery personnel (Star 1988). Tips therefore account for a substantial portion of customers' budgets to acquire a wide range of services, and, at the same time, furnish a crucial source of income for many low-income workers: more than two million tipped employees work in the U.S. food and beverage sector, paid at the minimum (or subminimum) wage (Jones 2016).

Since it was introduced in the United States in the early 1900s, the practice of tipping has gradually grown in significance (Azar 2004b). The customary tip-to-price ratio in U.S. restaurants has generally increased over time from 10% in the '60s, going through 15% in the '80s to 20% (Azar 2004a) and above nowadays, putting more and more pressure on customers to tip. This escalation has triggered fiery debates among customers, firms, employees, and legislators across the United States. Some firms have even experimented with abolishing tipping, only to revert to the practice shortly after (Lynn 2017).

Despite its far-reaching implications on businesses' revenue, labor cost, and customer satisfaction, tipping has not received adequate attention in the management science literature. We use the lens of that literature to address a series of questions relevant to the debates about tipping. First, what economic forces drive the evolution of tipping norms, and why do these norms change over time? Second, what are the implications of tipping on the demand, price, and quality of the service, and how do they affect the economic rents among stakeholders? And third, how do tipping-related policies affect the distribution of said rents, and are such policies even sustainable in the long run?

Answering these questions is challenging for several reasons. First, tipping is self-perpetuating, as customers feel pressure to tip when others do (Lynn 2015, Azar 2020), making the cause and effect of tipping virtually impossible to isolate. Second, evaluating the consequences of tipping is difficult because tipping involves various social and psychological factors that are hard to quantify. Third, studying tipping experimentally in a controlled environment is challenging because the evolution of tipping norms is a long social process that is hard to replicate or test in a laboratory setting.

In light of these challenges, we adopt a model-based approach to disentangle the different economic drivers of tipping. To answer our research questions, we develop a modeling framework of a firm offering a single service at a posted price to customers with heterogeneous willingness to pay. Each customer chooses whether to purchase the service and how much to tip afterward. Our approach focuses on the two primary and most-cited motives for customers to tip (Azar 2020): *appreciation*, which is the customer's desire to reward the worker by sharing the generated surplus, and *social pressure*, which is the intangible cost of deviating from the norm.

We first characterize customer tipping behavior when the tipping norm benchmark, referred to as the *customary tip*, is considered exogenous. We find that tips weakly increase with the willingness to pay. Among customers with low willingness to pay, tipping is driven by social pressure, whereas among other customers, tipping is driven by appreciation. To explore the endogenous formation of the norm, we study a dynamic process in which the average tip in each period determines the customary tip, according to which the firm will set the price and the customers will tip in the next period. This process is shown to converge to a stable market equilibrium outcome. Analyzing the comparative statics of this equilibrium, we show that the tip-to-price ratio increases when customers are more inclined to conform with the norm, when their level of heterogeneity increases, or when service quality improves.

Finally, we build on the base model to discuss three applications:

1. The implications of *tip credits* (Section 3.1): A tip credit is a regulation that allows firms to count tips toward the minimum wage as part of workers' compensation. We find that with a tip credit, the firm can appropriate the tips, thus enjoying cheaper labor at the expense of the individual worker, whose total income per shift decreases. At the same time, because of the reduction in labor cost, the firm decreases the price, serves more demand, and hence increases the total number of employment hours.

2. The role of tips as a quality control mechanism (Section 3.2): According to our analysis, tips are not highly effective in motivating workers to exert effort that improves service quality. When social pressure is high, tips are high as well but are mainly driven by conformity and therefore are insensitive to service quality. When social pressure is low, tips are indeed more sensitive to service quality, yet too small in magnitude to reward the worker for putting effort.

3. The viability of tip abolition (Section 3.3): Tips facilitate *price discrimination*: Customers with higher surplus tip more, ending up paying more for the same service, and the rents accrued finance the workers' compensation. Eliminating tips, on one hand, prevents the firm from extracting rents but, on the other hand, allows the firm to set the workers' compensation directly rather than leaving it to customers to decide. We find that tip abolition is profitable only when the labor pool is immobile, in which case the firm is willing to give up the benefits of rent extraction at the expense of the workers' income.

1.2. Literature Review

Tipping has been studied in the literature mainly from an empirical or experimental perspective, yet to a much lesser extent from a theoretically analytic viewpoint. For conciseness, in this review, we concentrate on the key motivations for tipping and how they have been addressed in relevant economic models. For comprehensive overviews and surveys of the literature on tipping written by pioneers in this area, see Lynn (2006) and Azar (2007, 2020).

1.2.1. Economic Justification of Tipping. Tipping is a voluntary act performed after receiving service, making it challenging to rationalize. Economists have long struggled to explain why a rational, "homo-economicus" customer would tip (Lynn 2015). In theory, tipping can be justified for repeat customers, who may see a return on their tips in the form of better service in the future (Ben-Zion and Karni 1977). However, in a survey among U.S. and Israeli patrons, Azar (2010) does not find compelling empirical evidence to support this claim. Whether future-service considerations motivate repeat customers to tip, this rationale fails to explain why one-time customers, with no intention to return to the facility, in practice still tip (Kahneman et al. 1986).

Another common argument is that tipping reduces employee monitoring costs for managers (Jacob and Page 1980, Bodvarsson and Gibson 1997, Lynn and McCall 2000, Pencavel 2015) because estimating the quality of service is easier for a customer (who experiences the service) than it is for a manager. Thus, it is more efficient if the customer rewarded the worker directly rather than having the employer monitor the service performance. Yet this explanation addresses the benefits of tipping to businesses only, completely ignoring the customer's incentives. Furthermore, technology advances made tipping a less effective monitoring mechanism, and data suggest that tipping is not more prevalent in occupations conducive to customer monitoring (Azar 2005). Hence, in order to understand tipping, one needs to introduce human motivations that go beyond self-interest.

1.2.2. Behavioral Motivations of Tipping. The phenomenon of tipping has attracted the attention of scholars in social psychology, sociology, and hospitality management. Lynn (2015) draws on behavioral and experimental economics, as well as literature in various sociopsychological science disciplines, to identify several plausible motivations for tipping: helping servers, rewarding service, forward buying service, acquiring social esteem, and fulfilling a sense of obligation. Based on surveys conducted in the United States and Israel, Azar (2010, 2020) finds that the two most significant motivations for leaving a tip are showing appreciation to the service provider and conforming to a social norm. Whereas both appreciation and conformity can be addressed as psychological sources of intangible cost (or gain), they differ significantly from one another. Showing appreciation links a tipper to the worker through an idiosyncratic sense of fairness. By contrast, the social norm relates a tipper to other tippers through a shared desire to comply with a common, consensual conception that need not agree with that tipper's personal preferences. This interpretation views appreciation as what Azar (2004b) labels an "internal" psychological motivation for tipping, whereas compliance with a social norm can be thought of as an "external" one. We next elaborate on these two motivations.

1.2.3. Internal Motives: Appreciation and Fairness. There is a large body of experimental work demonstrating that people deviate from purely self-interested behavior when sharing money with others, who have no say in allocations, in order to induce fairness (Fehr and Gächter 2000). This behavior is best illustrated in laboratory experiments of dictator games, which constitute a large body of literature initiated by the work of Kahneman et al. (1986). In its simplest form, attributed to Forsythe et al. (1994), the game includes one subject, an "allocator," who receives a sum of money from the experimenter and is tasked to split that money with another subject, a "recipient." Ruffle (1998) draws an analogy between dictator games and tipping, where the allocator is the customer, the recipient is the service provider, and the sum of money split is the *surplus* generated to the customer in the service exchange. If allocators in the dictator game were purely self-interested, they would keep all the money for themselves. Yet in practice, allocators consistently offer strictly positive amounts (Fehr and Gächter 2000, List 2007). This departure from material self-interest has inspired researchers to develop theoretical utility models that capture different aspects of social preferences such as altruism, reciprocal fairness, and inequality aversion. Notable among these works is the model of Fehr and Schmidt (1999) and its variation by Charness and Rabin (2002), which we use as a building block in our customer utility model. For a comprehensive overview of dictator games, its variations, and related descriptive models, see Camerer (2011, chapter 2).

1.2.4. External Motives: Social Norms. Through tipping, customers relate not only to the service provider but also to peer customers in choosing a tip that aligns with the common tipping practice, thereby acquiring social esteem and fulfilling a sense of obligation. This connection is possible when customers share a common belief about tipping. Indeed, there are various ways through which these common practices spread: aside from being handed down from mouth to ear, tipping practices are disseminated via books and guides (Star 1988, Schein et al. 1989, Post and Senning 2022), and businesses often present tip suggestions on bills and on Square displays upon purchasing the service (Alexander et al. 2021). Incorporating social preferences in utility functions to relate the individual and its peers has recently gained acceptance in the behavioral and socioeconomic literature, with Levitt and List (2007) being one prominent example. In the tipping literature, Azar (2004a) develops a model in which he endogenizes the social norm, and customers incur disutility when they tip a different amount than what the norm prescribes. This approach aligns with the psychological concept of "cognitive dissonance" used to describe the discomfort that individuals experience in the presence of information inconsistent with their views or actions (Festinger 1957, chapter 8).

1.2.5. Social Comparisons and Our Model of Tipping. Our paper relates to a recent stream of literature in operations management acknowledging the influence of "social comparisons" on worker or consumer behavior. Such social comparisons arise in many different contexts and settings, such as labor (Roels and Su 2014, Tan and Netessine 2019), gaming (Mai and Hu 2022), selling (Tereyağoğlu and Veeraraghavan 2012, Momot et al. 2020), and services (Kostami et al. 2017). To the best of our knowledge, ours is the first paper in this line of literature to consider the context of tipping. Our model further departs from previous literature in considering not one but two distinguished types of social interactions, namely, the customer-server interaction (appreciation) and the customer-customer interaction (social pressure). As a result, customers in our model who, in the absence of social pressure, would tip less than the prevailing norm, in the presence of social pressure, may "top up" the tip to reduce their discomfort. Such behavior is strongly supported by numerous papers in psychology, the seminal one being Heider (1946), claiming that individuals show a tendency to increase the consistency between their personal attitudes and their social environment (Helbing 2010).

2. Model

2.1. Baseline Assumptions and Primitives

We consider a single firm selling a service in a market comprised of a continuum of potential, vertically differentiated customers who are heterogeneous in terms of how much they value the service. Without loss of generality, we normalize the potential customer-demand volume to one. After the firm posts a price, customers choose whether to consume the service or not and, if they do, how much to tip after service is provided. Following the tipping literature, we consider two key motivations for customers to tip:

1. The customer-worker interaction (appreciation): Customers receive surplus (service valuation net of price) from being served by the worker and are willing to share with the worker a fraction of this surplus, in the form of tips, at an amount they perceive as fair. Thus, the customer utility depends not only on the net value that service generates to them but also on how this net value is split between them and the worker.

2. The customer-customer interaction (social pressure): Customers face pressure to conform to the social norm of tipping. They compare their tip to a benchmark, which we refer to as the *customary tip*. When customers tip less than the customary tip, they incur disutility, referred to as "social-pressure cost."

Below, we formalize these two motivations. Similarly to Mussa and Rosen (1978), a customer's valuation is represented by θ, her *type*. These types are continuously distributed over $[\theta_L, \theta_H]$ (where $\theta_L < \theta_H$), with cumulative distribution function F and tail-distribution function $\bar{F}$. For convenience, we assume that F is strictly increasing over $[\theta_L, \theta_H]$.

A customer of type θ, after paying a price p, receives a surplus of $\theta - p$. By tipping t, the customer then splits this surplus into two shares, $\pi_C = \theta - p - t$ for herself and $\pi_W = t$ for the worker. Following Charness and Rabin (2002), we model the customer's distributional preference, v, as a weighted average of these shares, taking the form $\sigma\pi_C + (1-\sigma)\pi_W$ when $\pi_W \le \pi_C$ and $(1-\sigma)\pi_C + \sigma\pi_W$ otherwise[1] for some $\sigma > 0$. Thus, σ is the weight the customer assigns to the larger share of the surplus. As a function of the tip, t, we can write v as

$$v(t;\theta,p) = \begin{cases} \sigma(\theta - p - t) + (1-\sigma)t & \text{if} \quad t \le (\theta - p) - t, \\ (1-\sigma)(\theta - p - t) + \sigma t & \text{otherwise}. \end{cases} \tag{1}$$

To induce fairness, we assume $\sigma < 1/2$; that is, the customer puts the least weight on the larger share of the surplus (see also Fehr and Schmidt 1999).

To capture the pressure that customers undergo when tipping, we introduce a component in the customer utility function that we refer to as the *social-pressure cost*. As in Azar (2004a), this cost reflects the loss for customers caused by feelings of social rejection when they tip an amount that is less than the *customary tip* level, m. In subsequent analysis (Section 2.2), this customary tip m will be determined by the average tip of all customers. We model the social-pressure cost as a smooth, convex increasing function K, with $K(d) = 0$ for all $d \le 0$, where d represents how much the tip is below the customary tip; $d = m - t$. Thus, a customer who chooses to tip t incurs a social-pressure cost of $K(m - t)$. Note that the assumption $K(d) = 0$ for all $d \le 0$ implies that customers who tip at or above the customary tip m do not incur any social-pressure cost.[2]

Given the customary tip m and the price p, a customer of type θ who obtains service and tips t receives a total utility[3] of

$$u(t, \theta; m, p) = v(t; \theta, p) - K(m - t), \tag{2}$$

whereas the utility of not purchasing the service is normalized to zero. The price p is set by the firm and posted before the customers make decisions. Knowing their own type and the customary tip, each potential customer observes the price and decides whether to consume the service and, if so, how much to tip afterward. Customers make tipping and consumption decisions so as to maximize their overall utility.

Consider the customary tip m and the price p as given. Assuming a customer consumes the service and tips, this customer's type θ is mapped to a corresponding rational tip, $t^*(\theta; m, p)$. A customer hence will consume if the return from service, after tipping, is nonnegative. We refer to the set of types of consuming customers, $\mathcal{C}^*(m, p)$, as the *consumption set*. Below, we formalize the criteria for rational consumption and tipping in the market.

Definition 1 (Rational Customer Behavior). Given m and p, the consumption set, $\mathcal{C}^*(m, p) \subseteq [\theta_L, \theta_H]$, together with the equilibrium tip function, $t^*(.;m,p):[\theta_L,\theta_H]\to[0,\infty)$, jointly satisfy

$$\begin{cases}\text{Rational tipping:} & t^*(\theta;m,p)\in\arg\max_{t\geq 0}u(t,\theta;m,p)\\ & \text{for all }\theta\in\mathcal{C}^*(m,p)\\ \text{Rational consumption:} & \theta\in\mathcal{C}^*(m,p)\text{ if and only if}\\ & u(t^*(\theta;m,p),\theta;m,p)\geq 0.\end{cases}$$

With the consumption set $\mathcal{C}^*(m,p)$, the firm's revenue is expressed as $p\int_{\theta\in\mathcal{C}^*(m,p)}dF(\theta)$. Consistent with practice, this definition assumes the firm cannot pocket the tips (U.S. Department of Labor, Wage and Hour Division 2011).

2.2. Roadmap of Analysis

We analyze the market equilibrium in two stages. In the first stage, for a given price, p, and a given customary tip, m, we characterize rational customer consumption and tipping behavior (see Definition 1), in turn allowing us to solve for the price p that maximizes the firm's revenue (see Definition 2 below). In the second stage, we study the long-run consistent firm's pricing and customers' purchasing and tipping behavior, in which the customary tip aligns with the average tip (see Definition 3 below). Taken together, these two stages identify a unique pair, a price p^* and a customary tip m^*, which are the main dependent variables in our paper. Throughout Section 2, the employees (or workers) do not play a role in the analysis. We introduce workers in Section 3, in which we apply our analytic framework to study tip credit, quality control, and tip abolishing.

2.3. Exogenous Customary Tip

In this section, we analyze rational tipping and consumption under the assumption that the customary tip, m, is exogenous, first treating the price p as given and then characterizing the firm's revenue-maximizing price. We omit the m and p specifications wherever these are considered fixed, and we will reintroduce them in the notation in Section 2.2 when we discuss the endogenous formation of m and p. The key result of this section is Proposition 1, which, given m and p, uniquely characterizes the consumption set $\mathcal{C}^*$ and tip function t^* in a way that facilitates the firm's pricing problem (given m) formulated later in Definition 2.

2.3.1. Rational Customer Behavior. We start off the analysis by deriving preliminary results.[4] For a fixed price p, consider a θ-type customer, and let t be their tip. Recall from Equation (1) that the customer's distributional utility is v. By tipping an additional \$1, customers increase the worker's share, π_W, by \$1 and reduce their own share, π_C, by \$1, increasing the difference, $\pi_W-\pi_C$, by \$2. When the worker's share is smaller ($\pi_W\leq\pi_C$), the distributional utility, v, can be rewritten as $\pi_C+(1-\sigma)(\pi_W-\pi_C)$, and tipping \$1 changes v by $-1+(1-\sigma)\cdot 2$, that is, an increase of $1-2\sigma$, which is positive because of $\sigma<1/2$. Symmetrically, when the customer's share is smaller ($\pi_W>\pi_C$), then v is given by $\pi_C+\sigma(\pi_W-\pi_C)$, and tipping \$1 results in a change in v of $-1+\sigma\cdot 2$, which is negative. Thus, in the absence of social-pressure cost, the rational tip splits the surplus equally, $\pi_C=\pi_W$; hence, the tip is $(\theta-p)/2$.

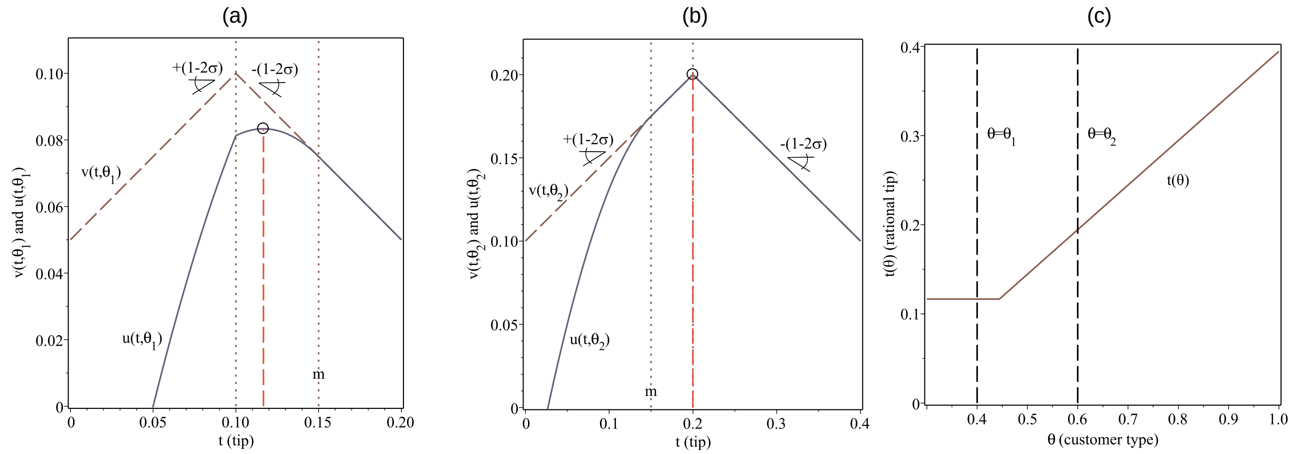
Next, we incorporate the social-pressure cost in the customer utility, u; see Equation (2). As $K(m-t)$ decreases as a function of the tip t, the customer's utility increases for $t<(\theta-p)/2$, implying that under rational tipping the customer would tip at least[5] $(\theta-p)/2$. In tipping more than $(\theta-p)/2$, the customer trades off the possible reduction in social-pressure cost with the disutility from deviating from evenly splitting the surplus. The customer utility is then maximized when the marginal decline in v, $1-2\sigma$, equals the marginal social-pressure cost $K'(m-t)$ or, equivalently, when the customer tips $m-\underline{d}$ for $\underline{d}=\arg\min_d\{-(1-2\sigma)d+K(d)\}$.

Recall that the social-pressure cost is minimized at any $t\geq m$. It follows that for a customer tipping $(\theta-p)/2$, if that tip amount is sufficiently large, $(\theta-p)/2\geq m-\underline{d}$, the resulting social-pressure cost will be insignificant (if at all). Such a customer will not feel obliged to further increase their tip beyond $(\theta-p)/2$. This reflects a situation where the customer values the service highly so that her feelings of appreciation obviate their seeking of social approval via tipping. Hence, we label the term $(\theta-p)/2$ as the customer's *appreciation tip*. However, if $(\theta-p)/2$ is too small, the customer will be driven by social pressure to increase her tip, specifically, to $m-\underline{d}$, which is the amount that balances the social-pressure cost and the cost of tipping more than the appreciation tip. Hence, we call $m-\underline{d}$ the *conformity tip*.

Figure 1, (a) and (b) illustrates rational tipping for two customer types. The dashed line represents $v(t,\theta)$. The solid curve represents the customer utility $u(t,\theta)=v(t,\theta)-K(m-t)$. The low customer type (θ_1) is pressured by the social norm to tip more than her appreciation tip (Figure 1(a)). In contrast, the appreciation tip of the high customer type, $(\theta_2-p)/2$, exceeds $m-\underline{d}$, and therefore, such a customer will tip based on appreciation (Figure 1(b)). Figure 1(c) illustrates the rational tip $t^*(\theta)$ for all customer types in $[3/10,1]$. It is also evident from Figure 1, (a) and (b), that the customer utility at the optimal tip amount is higher for the customer with the higher type. Lemma 1 formalizes these results.

Lemma 1. *Provided the customary tip, m, and price, p, for a rational customer of type $\theta\in[\theta_L,\theta_H]$ that consumes service, the optimal tip is given by $t^*(\theta)=\max\{(\theta-p)/2, m-\underline{d}\}$. Furthermore, the customer utility from consuming*

Figure 1. (Color online) Illustration of Rational Tipping for $K(d) = (\kappa/2)([d]^+)^2$ (Where $[d]^+ = \max\{d, 0\}$)



Notes. Parameters are: $\kappa = 15$, $\sigma = 1/4$ (which gives $\underline{d} = 1/30$), $m = 15/100$, and $p = 2/10$. (a) and (b) Illustration of $v(t, \theta)$ (dashed line) and $u(t, \theta) = v(t, \theta) - K(m - t)$ (solid line) for $\theta_1 = 4/10$ and $\theta_2 = 6/10$. The rational tip in (a) is the conformity tip $m - \underline{d} = 7/60$ and the appreciation tip $(\theta_2 - p)/2 = 2/10$ in (b). (c) Illustration of the rational tip $t(\theta)$ as a function of the customer types for $\theta \in [3/10, 1]$.

(and tipping) is strictly increasing in the type, that is, $(d/d\theta)u(t^*(\theta), \theta) > 0$ *for all* $\theta \in \mathcal{C}^*$.

A direct implication of Lemma 1 is that there exists a cutoff customer type, $\underline{\theta}$, for which the utility is equal to zero,[6] $u(t^*(\underline{\theta}), \underline{\theta}) = 0$. Given how much she intends to tip, this $\underline{\theta}$-type customer is indifferent between purchasing the service or not. Thus, any arbitrary θ-type customer consumes if and only if $\theta \geq \underline{\theta}$; hence, the consumption set contains the high-end customer types: $\mathcal{C}^* = [\underline{\theta}, \theta_H]$. Among the customers who consume the service (with types in $[\underline{\theta}, \theta_H]$), some are of sufficiently low types and tip the conformity tip, and the rest are of higher types who choose their appreciation tips. Depending on m and p, the consumption set may accommodate both "conformity tippers" and "appreciation tippers," in which case there will be some customer, of type $\hat{\theta}$, whose appreciation tip equals the conformity tip; that is, $(\hat{\theta} - p)/2 = m - \underline{d}$. Expressing $\underline{\theta}$ and $\hat{\theta}$ in terms of m and p, we can fully characterize the consumption set and tip function, as established in Proposition 1 below:

Proposition 1. *For every* $m \geq 0$ *there exists a unique value* $c(m) \geq 0$ *satisfying*

$$(1 - \sigma)c(m) = (1 - 2\sigma)m + L(\min\{m - c(m)/2, \underline{d}\}), \quad (3)$$

where $L(d) = -(1 - 2\sigma)d + K(d)$. *Additionally,* $c(m)$ *is strictly increasing and satisfies* $c(m) < m$. *Provided also the price* p, *under rational customer behavior, the consumption set is given by* $\mathcal{C}^* = [\underline{\theta}, \theta_H]$, *where* $\underline{\theta} = p + c(m)$, *and a customer of type* $\theta \in [\underline{\theta}, \theta_H]$ *tips*

$$t^*(\theta) = \begin{cases} m - \underline{d}, & \text{if } \theta \leq \hat{\theta} \\ (\theta - \underline{\theta} + c(m))/2, & \text{otherwise,} \end{cases} \quad (4)$$

where $\hat{\theta} = p + 2(m - \underline{d})$.

The term $c(m)$ captures the total loss of value (including social-pressure cost) incurred by the cutoff customer because of tipping. It is noteworthy to mention that this loss is smaller than the customary tip itself, $c(m) < m$. This is because the tip constitutes the worker's payoff, π_W, which contributes to the customer's utility via the term v (see Equation (2)); hence, it partially compensates the customer's material loss from tipping. Specifically, note that when a customer in our model tips $t = m$, the social-pressure cost is eliminated ($K(0) = 0$). The customer's share of surplus, π_C, then reduces by m, yet the worker's share, π_W, increases by m. As the customer also cares about the worker's share of the surplus, her utility loss associated with tipping would therefore be less than m. It follows that the cutoff customer, by choosing their tip optimally, ensures not to lose more than m; hence, $c(m) < m$. This resonates with Lynn and Brewster (2018), who argue that tipping decreases the customer's perception of the expensiveness of the service compared with service-inclusive prices. This observation will play an important role later in the Section 3 when we discuss abolishing tipping.

As explained earlier, $\hat{\theta}$ is the type of customer whose appreciation tip equals $m - \underline{d}$. It is therefore intuitive that when the customary tip m is small (in particular, when $m = 0$), then $\hat{\theta} < \underline{\theta}$, and all customers in the consumption set tip for appreciation. When m is moderate, that is, when $\underline{\theta} \leq \hat{\theta} \leq \theta_H$, some customers tip for conformity and others for appreciation. Otherwise, when m is sufficiently high, then $\hat{\theta} > \theta_H$ such that all purchasing customers tip the conformity tip, $m - \underline{d}$. This implies that all customers tip less than the customary tip, m, which, so far, we have treated as an exogenous quantity. In Section 2.2, we endogenize m by imposing equality between the customary tip and the average tip over the consumption set. This imposition will rule out the possibility that all customers tip below m.

2.3.2. Optimal Pricing. The characterization in Proposition 1 of the cutoff value (and hence also the demand for service) for any given price naturally raises the matter of price optimization: given m, each price p maps to a unique cutoff, $\underline{\theta} = p + c(m)$, which, in turn, determines the equilibrium demand for service, $\bar{F}(\underline{\theta})$, and hence the firm's profit, $p \cdot \bar{F}(\underline{\theta})$. We next formulate this pricing problem, assuming throughout that customers consume and tip according to Proposition 1. Thus, when referring to the tip function and/or cutoff type, we mean those characterized in Equation (4); therefore, we drop the "$*$" indication from the notation.

We analyze the firm's pricing decision as if it were setting the cutoff type $\underline{\theta}$ directly instead of the price. Henceforth, our independent variable in the pricing problem is thus $\underline{\theta}$ instead of p. We introduce $\underline{\theta}$ as an argument in the expressions below. Given the exogenous customary tip m, the price then takes the form $\underline{\theta} - c(m)$.

Definition 2 (Optimal Pricing). Given the customary tip, m, the optimal price for the firm is determined by $\underline{\theta}^{\text{opt.}}(m) - c(m)$, where $\underline{\theta}^{\text{opt.}}(m)$ satisfies[7]

$$\underline{\theta}^{\text{opt.}}(m) \in \arg \max_{\underline{\theta} \in [\theta_L, \theta_H]} (\underline{\theta} - c(m))\bar{F}(\underline{\theta}). \quad (5)$$

The firm's optimal cutoff $\underline{\theta}^{\text{opt.}}(m)$ defines a function mapping m to a type. Thus, for a customary tip m, we can characterize the firm's optimal cutoff type. An important observation is the following:

Lemma 2. *The optimal cutoff type, $\underline{\theta}^{\text{opt.}}(m)$, is continuously increasing in m and is strictly increasing over the set $\{m \text{ s.t. } \theta_L < \underline{\theta}^{\text{opt.}}(m) < \theta_H\}$.*

From the firm's perspective, the utility loss from tipping incurred by the cutoff customer, $c(m)$, is analogous to production cost; that is, it can be thought of as a cost per each "unit" of service sold. Recall from Proposition 1 that $c(m)$ increases in the customary tip, m. Not surprisingly, Lemma 2 states that as the customary tip increases, the firm is forced to focus on the higher end of the market. We next discuss how the price and the tipping behavior stabilize in equilibrium over time, when the customary tip emerges endogenously as the average tip in the population.

2.4. Endogenous Tipping Norm

In the previous section, we considered the customary tip m as a given exogenous quantity. In this section, we focus on the formation of a tipping norm, namely, on how the value m evolves over time through sequential adaptations of the firm's price and the customers' behavior.

For that purpose, consider a dynamic, infinitely long sequence of markets in discrete periods. Initially, the customary tip is fixed, and the firm sets a price. Similarly to Azar (2004a), the customary tip in the next period is determined by the average rational tip in the current period. At the beginning of the next period, the firm then updates its price based on the new customary tip, and customers adapt their decisions, resulting in a new market outcome. This process continues repeatedly, and when it stabilizes, we say that the market is in equilibrium. We aim to study the asymptotic behavior of this process and to characterize the tipping norm, that is, the equilibrium customary tip, together with its appropriate equilibrium price.

The two key results of this section are Propositions 2 and 3. Proposition 2 states that the equilibrium (i.e., asymptotic) pair of average tip and corresponding price exists uniquely. It further implies that starting at zero, the average tip increases monotonically to the equilibrium one, whereas the sequence of corresponding prices decreases monotonically. Hence, the tip-to-price ratio increases over time. Proposition 3 then discusses the comparative statics of the equilibrium average tip and price when social pressure intensifies. It is shown that when social pressure is weak, customers' tips in equilibrium are driven exclusively by appreciation. Yet, as social pressure grows stronger, the proportion of customers among those who consume that tip for conformity increases to one. This causes the equilibrium average tip to increase and the price to decrease; hence, the equilibrium tip-to-price ratio increases in social pressure.

We shall now describe in more rigor the process that governs the formation of the tipping norm. First, we reintroduce the customary tip m in the notation as an argument of the functions in which it appears. For each m and $\underline{\theta}$ the corresponding price is $\underline{\theta} - c(m)$ and customers consume and tip rationally. Thus, the consumption set is given by $[\underline{\theta}, \theta_H]$, and the tip, for each type $\theta \in [\underline{\theta}, \theta_H]$, is given by $t(\theta; m, \underline{\theta})$, as described in Proposition 1, where we rewrite $\hat{\theta}$ in Equation (4) as $\hat{\theta}(m, \underline{\theta}) = \underline{\theta} - c(m) + 2(m - \underline{d})$. The average rational tip is then given by

$$ET(m, \underline{\theta}) = \begin{cases} \dfrac{\int_{\underline{\theta}}^{\theta_H} t(\theta; m, \underline{\theta}) dF(\theta)}{\int_{\underline{\theta}}^{\theta_H} dF(\theta)} & \text{if } \underline{\theta} \in [\theta_L, \theta_H), \\ t(\theta_H; m, \theta_H) & \text{if } \underline{\theta} = \theta_H. \end{cases} \quad (6)$$

Periods are indexed by $k = 1, 2, \ldots$, and the average tip at period k is denoted by m_k, which serves as the customary tip for the next period. The customary tip at the beginning of the first period is denoted by m_0. At each period k, given the average tip at the previous period, m_{k-1}, the firm's price is set to $p_k = \underline{\theta}_k - c(m_{k-1})$, where the cutoff customer $\underline{\theta}_k$ is chosen optimally,

$\underline{\theta}_k = \underline{\theta}^{\text{opt.}}(m_{k-1})$. The resulting average tip at period k, which determines the customary tip for period $k+1$, is therefore updated via $m_k = BR(m_{k-1})$, where

$$BR(m) = ET(m, \underline{\theta}^{\text{opt.}}(m)). \tag{7}$$

Given the current period's customary tip m, $BR(m)$ is understood as the average *Best Response* tip in the next period, which takes into account both rational customer behavior (Definition 1) and the firm's optimal pricing (Definition 2). We define the equilibrium tip, equilibrium cutoff, and equilibrium price as $m^* = \lim_{k\to+\infty} m_k$, $\underline{\theta}^* = \lim_{k\to+\infty} \underline{\theta}_k$, and $p^* = \lim_{k\to+\infty} p_k$, respectively.

Our formulation suggests that if the limits m^* and $\underline{\theta}^*$ indeed exist, then they must satisfy $\underline{\theta}^* = \underline{\theta}^{\text{opt.}}(m^*)$ and $m^* = BR(m^*) = ET(m^*, \underline{\theta}^*)$, thereby motivating the following definition:

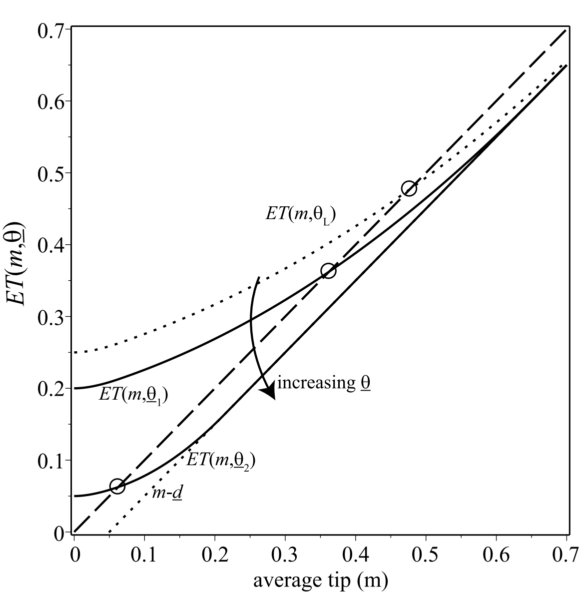
Definition 3 (Consistency of m and $\underline{\theta}$). For a customary tip m and a cutoff $\underline{\theta}$, the pair $(m, \underline{\theta})$ is called consistent if m and $\underline{\theta}$ satisfy

$$m = ET(m, \underline{\theta}). \tag{8}$$

In words, the cutoff and customary tip are consistent when the associated rational customer behavior yields an average tip equal to the customary tip.

Figure 2 illustrates $ET(m, \underline{\theta})$ as a function of m for two different values of $\underline{\theta}$, $\underline{\theta}_1 = 0.2$, and $\underline{\theta}_2 = 0.8$, where we set $\sigma = 1/4$, $K(d) = (\kappa/2)([d]^+)^2$ with $\kappa = 20$, and uniform customer types over [0, 1]. The figure

Figure 2. The Average Tip $ET(m, \underline{\theta})$ as a Function of m, Quadratic Social-Pressure Cost $K(d) = (\kappa/2)([d]^+)^2$, $\kappa = 10$, $\sigma = 1/4$, and Customer Types Uniformly Distributed over [0, 1]



Notes. In the top curve, $\underline{\theta}_1 = 0.2$, and in the bottom curve, $\underline{\theta}_2 = 0.8$. The dashed line is the identity (45-degree) line. The top dotted line corresponds with $\underline{\theta} = \theta_L$, which intersects the identity line at the highest possible consistent customary tip. The bottom dotted line is $m - \underline{d}$.

reveals several important characteristics of the average tip, $ET(m, \underline{\theta})$. First, it increases in the customary tip, m, which is intuitive and driven by social pressure. Second, it intersects the identity line exactly once; thus, given $\underline{\theta}$, the consistency criterion of Equation (8) is satisfied at exactly one value of m. Third, the intersection with the identity line is higher for lower values of $\underline{\theta}$. This is intuitive because when the cutoff customer type is lower, which means a lower price, each customer's surplus is higher and, hence, also their appreciation tip. It follows that the maximum possible consistent customary tip corresponds with $\underline{\theta} = \theta_L$. Finally, for any $\underline{\theta}$, when m is sufficiently high (specifically, such that $\hat{\theta}(m, \underline{\theta}) \geq \theta_H$), $ET(m, \underline{\theta})$ coincides with $m - \underline{d}$, below the identity line, indicating that all customers choose the conformity tip. Clearly, such a high value for m does not satisfy consistency (see Definition 3); therefore, in equilibrium, there must be some customers who tip their appreciation tip.

Rather than considering the evolution of the tipping norm over a sequence of periods, one could also consider a market of customers with a rational-expectations condition imposed (as in Muth 1961). The firm plays first by setting the price, and customers react by making rational consumption and tipping choices. The consistency condition in Equation (8), requiring that the believed average tip coincides with the actual one, should then be interpreted as a precondition on the equilibrium outcome, namely, the rational-expectations assumption. This alternative equilibrium criterion would result in an outcome identical to the asymptotic equilibrium discussed hereafter. However, in this paper, we choose the dynamic formulation over the rational-expectations one because it puts stronger emphasis on the time component in the evolution process of the tipping norm, which we wish to understand.

To keep the analysis focused and obtain sharp structural properties, throughout the remainder of the paper, we make the following two assumptions:

Assumption 1. *The lowest customer type θ_L is less than $\underline{\theta}^{\text{opt.}}(0)$.*

Assumption 2. *The type distribution, F, is of increasing failure rate (IFR).*

Because $\underline{\theta}^{\text{opt.}}(m)$ is increasing (see Lemma 2), Assumption 1 implies that $\underline{\theta}^{\text{opt.}}(m) > \theta_L$ for every $m \geq 0$. As a result, the solution to the firm's pricing problem in Equation (5) lies in $(\theta_L, \theta_H]$, and the market is not fully covered in equilibrium. Assumption 2 is a regularity on the type distribution function that imposes $(dF(\theta)/d\theta)/\bar{F}(\theta)$ is increasing. This implies that the firm's profit, as a function of $\underline{\theta}$, admits (at most) one extreme point for every m.

Equipped with these assumptions, we can now state Proposition 2, the main analytic result of this section. It argues that under Assumptions 1 and 2, the equilibrium m^* solving $m^* = BR(m^*)$ is unique. Together with the unique equilibrium cutoff $\underline{\theta}^* = \underline{\theta}^{\text{opt.}}(m^*)$, the price $p^* = \underline{\theta}^* - c(m^*)$, and the corresponding tip function $t(\theta; m^*, \underline{\theta}^*)$ for $\theta \in [\underline{\theta}^*, \theta_H]$, we obtain a full characterization of the equilibrium market outcome that aligns with Definitions 1, 2, and 3. Additionally, we prove that the average-tip sequence is increasing, whereas the price sequence is decreasing so that the tip-to-price ratio increases over time; $m_0/p_0 \le m_1/p_1 \le \cdots \le m^*/p^*$.

Proposition 2. *Equilibrium existence and uniqueness.*

i. *There exists a unique customary tip m^* satisfying the equilibrium condition $m^* = BR(m^*)$, associated with a unique equilibrium cutoff type $\underline{\theta}^*$ satisfying $\underline{\theta}^* = \underline{\theta}^{\text{opt.}}(m^*)$, and the corresponding equilibrium price is given by $p^* = \underline{\theta}^* - c(m^*)$.*

ii. *For any $m_0 \in [0, m^*]$, the average-tip sequence is increasing, $m_0 \le m_1 \le m_2 \le \cdots \le m^*$, whereas the price sequence is decreasing,*[8] *$p_0 \ge p_1 \ge p_2 \ge \cdots \ge p^*$.*

Next, we study the interplay between different exogenous forces and the equilibrium tipping behavior described in Proposition 2.

2.5. Comparative-Statics Analysis

Consider a family of social-pressure cost functions parameterized by κ, defined as follows. Let $\ell(d)$ be a smooth, convex increasing function, with $\ell(d) = 0$ for all $d \le 0$. For a given positive parameter $\kappa > 0$, we consider the social-pressure cost function $K(d) = \kappa \cdot \ell(d)$. Thus, K satisfies our assumptions in Section 2.2. Intuitively, κ captures the intensity of social pressure; as κ grows, customers become more compliant to the norm, with the cost of deviating from m becoming more significant.

It is natural to expect that in equilibrium, when social pressure is low, most customers in the consumption set will tip their appreciation tip, and when it is high, more customers will tip the conformity tip. The next proposition formalizes this intuition.

Proposition 3. *Let $K(d) = \kappa \cdot \ell(d)$ for some smooth, convex increasing function $\ell(d)$ satisfying $\ell(d) = 0$ for all $d \le 0$. Then,*

i. *For all $\kappa > 0$, both the equilibrium average tip, m^*, and the equilibrium cutoff, $\underline{\theta}^*$, are increasing in κ, whereas the equilibrium price, p^*, is decreasing in κ.*

ii. *There exist $\kappa_0 > 0$ and $\kappa_1 \ge \kappa_0$ such that if $0 < \kappa < \kappa_0$, then $\hat{\theta}(m^*, \underline{\theta}^*) < \underline{\theta}^*$ (thus, all customers tip for appreciation), and if $\kappa > \kappa_1$, then $\hat{\theta}(m^*, \underline{\theta}^*) > \underline{\theta}^*$ (thus, some customers tip for conformity). Moreover, if $\ell'(d)$ is concave, then $\kappa_0 = \kappa_1$.*

iii. *In the limit, as the level of social-pressure κ tends to infinity, the equilibrium consumption set remains proper (nondegenerate), and all customers purchasing service tip for conformity; that is, $\lim_{\kappa \to +\infty} \underline{\theta}^* < \lim_{\kappa \to +\infty} \hat{\theta}(m^*, \underline{\theta}^*) = \theta_H$.*

Proposition 3(i) argues that the equilibrium average tip increases as social pressure intensifies. This, in turn, exacerbates the burden of tipping on customers, which forces the firm to reduce the price; hence, the price and the average tip move in opposite directions. It follows that the tip-to-price ratio increases in social pressure. Proposition 3(ii) then states that for all κ below a certain threshold, all customers tip their appreciation tips, whereas for all κ above some second threshold, there will be customers tipping for conformity. A technical condition is introduced under which these two thresholds agree.

The result in Proposition 3(iii) is more delicate. Intuitively, it states first that no matter how strong the social pressure is, there will always be a nonnegligible set of customers who consume the service ($\underline{\theta}^* < \theta_H$); thus, the firm's profit is strictly positive. Yet, with the intensification of social pressure, more and more customers tip the conformity tip, $m^* - \underline{d}$, so that in the limit, appreciation tips disappear: $\hat{\theta}(m^*, \underline{\theta}^*) \to \theta_H$. As $\lim_{\kappa \to +\infty} \underline{d} = 0$, with increasing social pressure, the tip distribution becomes more concentrated around its average, $m^* > 0$, such that in the limit, all tips are identical[9] and equal m^*.

In what follows, we focus on the two extreme regimes, $\kappa \to 0$ and $\kappa \to +\infty$, which we refer to, respectively, as *weak* and *strong social norm*. We specialize customer types to the uniform distribution;[10] thus, the market for the service corresponds to a linear demand curve. Then, we can explicitly express the dynamics of the tipping norm formation over time, under each of the two regimes, as summarized in the next proposition.

Proposition 4. *Let $K(d) = \kappa \cdot \ell(d)$ for some smooth, convex increasing function $\ell(d)$ satisfying $\ell(d) = 0$ for all $d \le 0$, and assume customer types are uniformly distributed over $[\theta_L, \theta_H]$. Then, in the limit, for every $m \ge 0$,*

i. *under weak social norm (as $\kappa \to 0$), $p^{\text{opt.}}(m) \to \theta_H/2$ and $BR(m) \to \theta_H/8$;*

ii. *under strong social norm (as $\kappa \to +\infty$),*

$$p^{\text{opt.}}(m) \to \frac{\theta_H}{2} - \frac{1-2\sigma}{2(1-\sigma)} m \quad \text{and}$$

$$BR(m) \to m + \frac{\left(\left[\theta_H - \frac{3-2\sigma}{1-\sigma} m\right]^+\right)^2}{8\left(\theta_H - \frac{1-2\sigma}{1-\sigma} m\right)}.$$

In the weak-social-norm regime, with uniformly distributed types, the firm extracts all the surplus from the cutoff customer, this customer's tip becomes zero, and so does $c(m)$. The price is therefore set equal to the cutoff type $\underline{\theta}^* = \theta_H/2$, which is determined by the solution $\max_{\underline{\theta} \in [\theta_L, \theta_H]} \underline{\theta} \cdot (\theta_H - \underline{\theta})/(\theta_H - \theta_L)$ and, hence, does not depend on m. In terms of the evolution dynamics of the social norm, this implies that the firm updates its optimal price once (setting it at $\theta_H/2$), and the customary tip and cutoff type settle immediately at the first iteration of the process. Because customers are exclusively motivated by appreciation, the average tip is given by $\theta_H/8$, which is half of the average surplus, $(\int_{\theta_H/2}^{\theta_H}(\theta-\theta_H/2)d\theta)/(\theta_H-\theta_H/2)=\theta_H/4$. The tip-to-price ratio is therefore 1/4.

Under strong social norm, for $m_0=0$, the average-tip sequence converges monotonically to $m^*=\frac{1-\sigma}{3-2\sigma}\theta_H$, and the tip-to-price ratio approaches one. Unlike the weak-social-norm case, under strong social norm, the average-tip sequence converges much more slowly, and as $m_{k+1}=BR(m_k)$, it can be seen from Proposition 4(ii) that

$$\lim_{k\to+\infty}\frac{|m_{k+1}-m^*|}{|m_k-m^*|}=1;$$

thus, the convergence is linear. For example, with $\theta_H=1$, $\theta_L=0$, and $m_0=0$, under weak social norm, the tip-to-price ratio settles immediately at 25% and stays at 25% forever after, whereas under strong social norm, the tip-to-price ratio takes 176 iterations to reach 74%, whereas the equilibrium ratio is 75% (for $\sigma=1/4$). Consistent with Proposition 3, in the weak-social-norm regime, the equilibrium average tip is lower, the price is higher, and the tip-to-price ratio is smaller compared with that under strong social norm, as summarized in Table 1.

Note that with weak social norm, even though appreciation is the sole motivation for tipping, the parameter σ does not play a role. This is because in the absence of social pressure, the customer distributional preference forces an even split of the surplus between the worker and customer for any $\sigma<1/2$.

However, with $\kappa\in(0,\infty)$, the impact of σ on the equilibrium tip and tip-to-price ratio involves contradicting effects. Recall that σ captures how much the customer cares about the worker (who gets the larger share of surplus). When σ increases, the customer marginal disutility from tipping decreases; that is, tipping feels less expensive to the customer. Thus, customers are willing to tip more in order to comply with the social norm; namely, the conformity tip $(m-\underline{d})$ increases. However, increasing σ also decreases the cutoff customers' utility loss from tipping, $c(m)$, thereby allowing the firm to increase the price, which,

Table 1. Comparison of the Equilibrium Average Tip, Price, and Tip-to-Price Ratio Under Weak and Strong Social Norms for Uniformly Distributed Types over $[\theta_L,\theta_H]$

Equilibrium	Weak social norm		Strong social norm
Average tip (m^*)	$\theta_H\cdot 1/8$	<	$\theta_H\cdot(1-\sigma)/(3-2\sigma)$
Price (p^*)	$\theta_H/2$	>	$\theta_H/(3-2\sigma)$
Cutoff type ($\underline{\theta}^*$)	$\theta_H/2$	<	$\theta_H\cdot 2(1-\sigma)/(3-2\sigma)$
Tip-to-price ratio (m^*/p^*)	1/4	<	$1-\sigma$
Convergence time	One iteration		Linear

in turn, drags the tips down (as the customer surplus is finite).

Under strong social norm ($\kappa\to\infty$), customers cannot afford to deviate from the customary tip regardless of the value of σ, so they all tip exactly m. In this limiting case, $\underline{d}=0$, and when σ grows, the conformity tip (which equals m) remains constant, although customers incur less disutility from tipping it. Hence, the equilibrium price increases with σ, consequently causing both the tip and the tip-to-price ratio to decrease. In Section 5.7 in the Online Appendix, we illustrate the comparative statics of the tip-to-price ratio with respect to σ, as discussed above, under weak and strong social norms for various parameters.

In the two aforementioned extreme regimes, with uniformly distributed types, the equilibrium tip-to-price ratio does not depend on the parameters of the type distribution, θ_L, and θ_H (see Table 1). This, however, is not the case for any finite and strictly positive κ. In order to obtain crisp insights on how θ_L and θ_H influence the equilibrium outcome and, in particular, the tip-to-price ratio, we consider an arbitrarily fixed $\kappa>0$ and focus on a quadratic social-pressure cost function, $K(d)=\kappa\cdot\ell(d)$ with $\ell(d)=(1/2)([d]^+)^2$. We conclude our comparative-static analysis with the following result:

Proposition 5. *When the social-pressure cost function is $K(d)=\kappa\cdot(1/2)([d]^+)^2$ and customer types are uniformly distributed over $[\theta_L,\theta_H]$, then the equilibrium average tip (m^*) and tip-to-price ratio (m^*/p^*) do not vary with θ_L, yet they strictly increase in θ_H.*

Proposition 5 first states that changes in θ_L do not affect the equilibrium outcome. This is because when the market is not fully covered (Assumption 1) and types are uniformly distributed, the cutoff type chosen by the firm is bounded from below by $\theta_H/2$, which is the cutoff type the firm would set had customers incurred no social-pressure cost at all. As a consequence, any customer with service valuation less than $\theta_H/2$, and particularly the one of the θ_L type, never consumes the service and therefore does not influence the tipping behavior. Hence, in the case of uniformly distributed types, the impact of the type distribution kicks in via the change in θ_H. Ceteris paribus, when θ_H increases, the surplus of any consuming customer increases, causing, on one hand, the average tip to increase because of bigger appreciation tips, but on the other hand, it allows the firm to gain more by raising the price. Hence, it is a priori not clear whether the tip-to-price ratio is increasing, yet Proposition 5 confirms it is. Below, we discuss two interesting implications of this result.

First, note that by expanding the support of the type distribution about its mean, that is, replacing $[\theta_L,\theta_H]$ by $[\theta_L-\epsilon,\theta_H+\epsilon]$ for some $\epsilon\geq 0$, customer valuations spread out, and the variance of the type distribution increases while keeping its mean constant. This modification can be interpreted as a growth in the level of heterogeneity in the population, holding the mean of the distribution constant. Then, based on Proposition 5, we conclude that as potential customers become more heterogeneous, the tip-to-price ratio increases.

Second, when the interval $[\theta_L, \theta_H]$ is scaled up, that is, replaced by $[q\theta_L, q\theta_H]$ for some $q \geq 1$, all customers value the service more.[11] This situation can be interpreted as the firm improving its service quality. Proposition 5 then argues that as long as Assumption 1 is satisfied, this will increase the tip-to-price ratio.

2.5.1. Summary of Insights. Historically, the tip-to-price ratio has been generally increasing over the years[12] (Azar 2004a, Margalioth 2010, Clifton et al. 2018). In our model, successive adaptations of customers' tips and endogenous pricing cause the tip-to-price ratio to increase over short time scales, ultimately converging to a stable equilibrium. But more importantly, our framework also allows us to identify the external forces that increase the tip-to-price ratio in the long run:

1. An intensification of social pressure: Together with appreciation, social pressure compels some low-tip customers to increase their tips closer to the average, causing the average tip to increase. The higher the social pressure, the stronger this effect will be. New technologies, such as Square, may increase social pressure by making each patron's tip observable to others.

2. An expansion of the range of customer valuations (heterogeneity): Margalioth (2006) argues that "rich" (high-end) customers tip generously to signal their social status, hence dragging "poor" (low-end) customers to tip more via a conformity (social norm) mechanism. Thus, rising social inequalities over the years have exacerbated the pressure that affluent customers exert on less affluent customers, causing the tip-to-price ratio to increase. In Proposition 5, we establish a result that supports this claim.

3. An exogenous improvement of service quality: Researchers have not yet reached a consensus on the extent to which service quality can explain tipping. Though data show that tip percentages are positively correlated with service evaluation, this relationship is rather small in magnitude (Lynn and McCall 2000). We find, based on Proposition 5, that for uniformly distributed types, the tip-to-price ratio increases when customers value the service more. Interestingly, this effect dies out either when the social norm of tipping is weak or when it is strong. Table 1 shows that in both extreme cases, the tip-to-price ratio is independent of customers' range of valuations. In Section 3.2, we analyze in more depth whether tips motivate workers to provide higher-quality service.

3. Applications

In this section, we address several questions of interest related to tipped employees and labor. So far, the supply of labor (workers) remained passive in our model. In this section, the firm's main decision variable remains the price, but now we assume that the firm employs a continuum of workers proportionally to demand. This assumption is reasonable if we think of labor being measured in units of worker shifts or hours such that the number of service units rendered by a single worker during a shift is limited, similar to the model analyzed by Shy (2015). For ease of exposition, we assume that each unit demand, namely, the service of a single customer, requires one unit of supply, and we refer to this normalized unit supply as a "worker."

Below, we build on the theoretical foundations laid in Section 2 to investigate three important applications of tipping: (a) the economic implications of the "tip credit" policy—a regulation that allows the firm to pay workers less than a minimum wage in anticipation of additional tip income (in Section 3.1), (b) how tipping creates incentives for workers to improve service quality (in Section 3.2), and (c) under what conditions, if at all, tip abolishing can be feasible (in Section 3.3).

3.1. Tip Credit

U.S. federal law allows employers to reduce the minimum wage for tipped occupations by a fixed amount, referred to as the tip credit, as long as the employer tops up the employee's total income whenever it is less than the state minimum wage (Azar 2012, U.S. Department of Labor, Wage and Hour Division 2023). Thus, the tip credit allows the employer to rely on a limited amount of tips received by the worker to finance the worker's wage. For example, in New Hampshire, the state minimum wage is \$7.25 per hour, and the tip credit is \$3.99 per hour, so an employer must pay their tipped employee a wage of (at least) \$3.26 per hour. If, indeed, the employee's hourly wage is \$3.26 and, in addition, the employee receives $t < \$3.99$ in tips per hour, then the employer needs to complement the employee's hourly wage with $\$3.99 - t$ to ensure that the worker obtains the state minimum wage, $\$3.26 + t + \$3.99 - t = \$7.25$. Otherwise, if $t \geq \$3.99$, then the employer does not need to pay the employee any extra. In some other states such as California, the tip credit is zero. Therefore, an employee's hourly wage must be at least the state minimum, which, in California, is \$15.50 per hour, and employees collect all of their tips on top of that hourly wage.

Let $\underline{w}$ be the minimum wage (per unit of service) and δ be the tip credit. Thus, the firm pays the worker a fixed wage of $\underline{w} - \delta$ per unit of service, referred to as the *cash wage*. In Section 2, we have assumed that

$\underline{w} = \delta = 0$. Now, we assume that $\underline{w}$ is exogenous and study the comparative statics of the social welfare with respect to tip credit, $\delta \in [0, \underline{w}]$. Whenever the tip plus the cash wage falls below $\underline{w}$, the firm pays the shortfall, $\{\delta - t(\theta; \underline{\theta}, m)\}^+$. Therefore, provided $\underline{\theta}$ and m, we redefine the firm's profit function (see Definition 2) as

$$\Pi(\underline{\theta}, m) = (\underline{\theta} - c(m) - (\underline{w} - \delta)))\bar{F}(\underline{\theta}) - \int_{\underline{\theta}}^{\theta_H} \{\delta - t(\theta; \underline{\theta}, m)\}^+ dF(\theta).$$

As in the base model (Section 2), the equilibrium conditions are $\underline{\theta}^* = \arg\max_{\underline{\theta} \in [\theta_L, \theta_H]} \Pi(\underline{\theta}, m^*)$ and $m^* = ET(m^*, \underline{\theta}^*)$.

Because of the random assignment of customers to servers, the tip received by the worker varies and hence introduces uncertainty in the worker's income. As tipped workers are often low-paid workers, such income uncertainty may be expensive for them (Lynn 2017). Below, we utilize our model to study the effects of the tip credit on social welfare, taking into account the adverse effect of income variability on workers' welfare.

Consider a single service performed by an individual worker. This worker is assigned to a customer chosen uniformly at random from the consumption set, $[\underline{\theta}, \theta_H]$. Given m and $\underline{\theta}$, if the latter customer's type is θ, the worker's income is $w(\theta; \underline{\theta}, m) = \underline{w} + \{t(\theta; \underline{\theta}, m) - \delta\}^+$; that is, the worker earns the minimum wage plus all tips in excess of the tip credit. The latter can be equivalently described as the maximum between minimum wage and the cash wage plus the tip; that is, $w(\theta; \underline{\theta}, m) = \max\{\underline{w} - \delta + t(\theta; \underline{\theta}, m), \underline{w}\}$. Workers are averse to fluctuations in their income caused by the randomness in the tips they receive. Given the level of risk aversion s, the worker utility as a function of income w is concave increasing and expressed as $u_W(w) = (1 - e^{-s \cdot w})/s$ for $s > 0$ and $u_W(w) = w$ for $s = 0$, which is the constant absolute risk aversion (CARA) function. Here, a higher value of s captures higher worker risk aversion. The expected utility is

$$EU_W(\underline{\theta}, m) = \frac{\int_{\underline{\theta}}^{\theta_H} u_W(w(\theta; \underline{\theta}, m)) dF(\theta)}{\int_{\underline{\theta}}^{\theta_H} dF(\theta)},$$

and the *aggregate worker surplus* is $EU_W(\underline{\theta}, m) \cdot \int_{\underline{\theta}}^{\theta_H} dF(\theta)$. The worker's certainty equivalent is defined as the income level, $w_c(\underline{\theta}, m)$, satisfying $u_W(w_c(\underline{\theta}, m)) = EU_W(\underline{\theta}, m)$, which, in general, is less than the expected income because of risk aversion (for $s > 0$). The aggregate customer surplus is measured as the overall net utility for customers;[13] $\int_{\underline{\theta}}^{\theta_H} u(t(\theta; m, \underline{\theta}), \theta; m, \underline{\theta} - c(m)) \cdot dF(\theta)$. The social welfare is then the sum of three components: the firm's profit, the aggregate worker surplus, and the aggregate customer surplus.

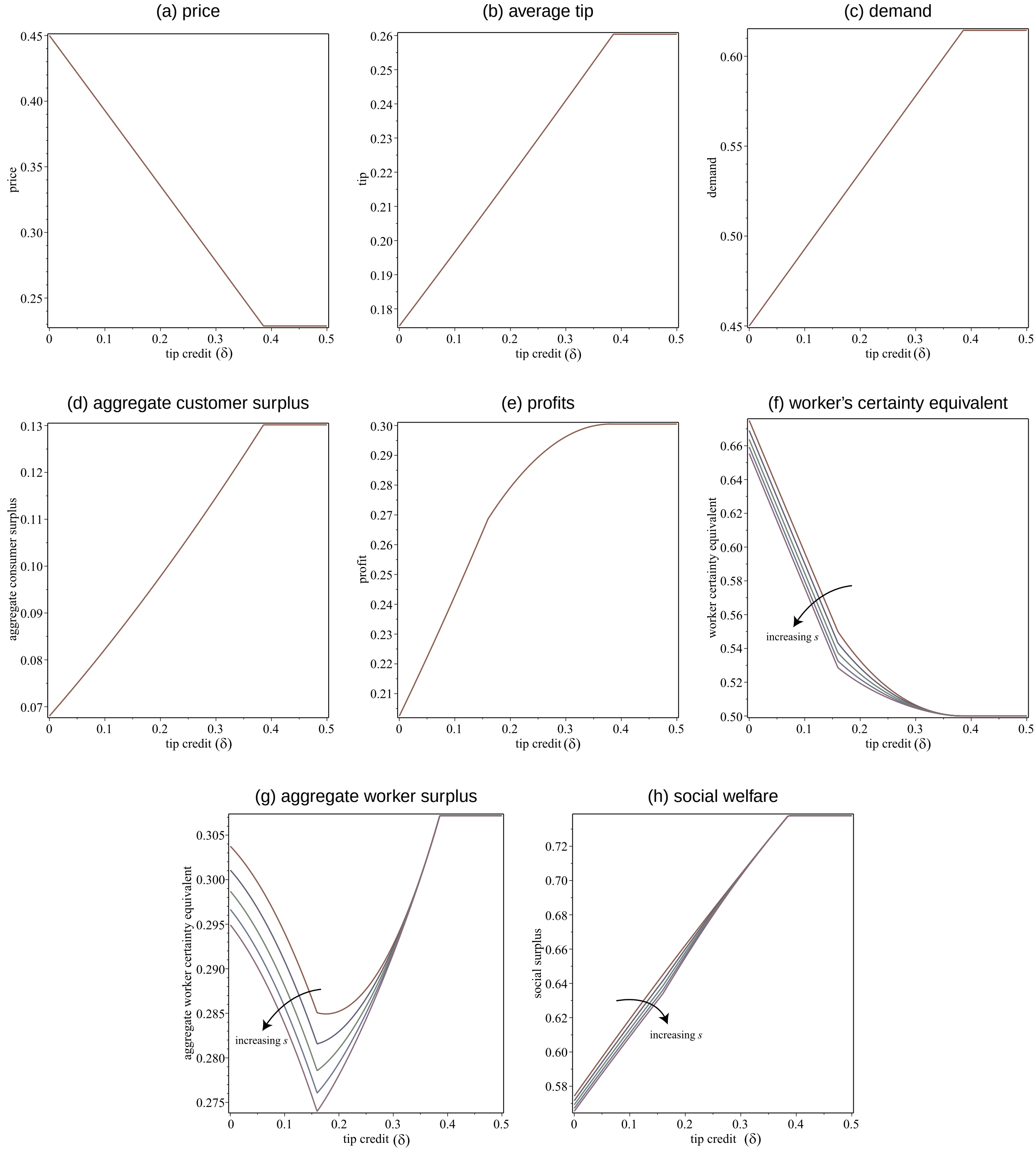
Figure 3 depicts results from numerical analysis of how the tip credit impacts the equilibrium price, tip, demand, aggregate customer surplus, firm's profit, worker utility, aggregate worker surplus, and social welfare. In our experiments, $\sigma = 1/4$, types are uniformly distributed over [0, 1], and the social-pressure cost is quadratic ($K(d) = \kappa \cdot (1/2)([d]^+)^2$) with $\kappa = 20$. The minimum wage is fixed at $\underline{w} = 3/10$.

Noticeably, in all panels, the curves are "kinked" such that for $\delta > 0.2$, they become constant. This is because for any tip credit larger than the maximal tip, the firm always tops up the worker's tipped income to make sure they meet the minimum wage $\underline{w}$. Each worker earns exactly $\underline{w}$ per service irrespective of the tip, and the firm's per-service average labor cost is $\underline{w} - m^*$. Focusing on tip credits below the maximal tip, we observe that increasing the tip credit reduces the labor cost to the firm, allowing it to reduce its price (Figure 3(a)), as hypothesized by Lynn (2017). This increases the surplus for the customers, increasing appreciation tips as well as the average tip (Figure 3(b)). Naturally, the reduction in price boosts the demand for service (Figure 3(c)), as well as the aggregate customer surplus (Figure 3(d)). Overall, we observe that the firm's profit increases[14] (Figure 3(e)).

Regarding the workers, increasing the tip credit impacts them in several opposing aspects. As discussed, a higher tip credit increases the average tip, which is part of the worker's income. It also benefits the worker, especially when the uncertainty-aversion parameter s is large, by reducing their income variability. However, these two positive effects are not enough to outweigh the worker's loss associated with reduced cash wage, and as a result, both the worker's expected utility and their certainty equivalent drop (Figure 3(f)). Interestingly, although the expected utility per worker decreases, the aggregate worker surplus may increase because the increased demand causes the firm to hire more workers. Consistent with empirical observations (see Even and Macpherson 2014), Figure 3(g) indeed shows that the aggregate worker surplus increases for intermediate levels of tip credit.

One frequent practice in the restaurant business is *tip-pooling* arrangements. At the end of each shift, all tips are gathered and divided equally between servers of that shift. With tip pooling, the variability of the worker's income is mitigated because each individual worker does not depend on the tip of a random customer they are assigned to but, rather, on the average tip. However, in the presence of a tip credit, tip pooling facilitates tip appropriation by the firm; hence, it may hurt the worker's income. To see why, assume, for example, that the minimum wage and tip credit are both \$10, and at the end of a shift, one worker receives in tips \$8, whereas the other receives \$12. Without tip pooling, the firm will then top up the former's wage

Figure 3. (Color online) The Impact of the Tip Credit on Various Economic Quantities, with Uniformly Distributed Types over [0, 1], Quadratic Social-Pressure Cost $K(d) = (\kappa/2)([d]^+)^2$, $\kappa = 10$, $\sigma = 1/4$, and $\underline{w} = 0.5$ for $U(\text{income}) = 1 - \exp(-s \times \text{income})$ Values of $s = 0, 5, 10, 15, 20$



by \$2, and the average income will be \$11. With tip pooling, each worker will receive exactly \$10 so that the firm will not have to pay them extra; thus, workers will be worse off. We will relate back to tip pooling in the following subsection and see how it can backfire on the firm and customers.

Based on our analysis, tip credits positively affect social welfare (Figure 3(h)), potentially making the tipping mechanism more efficient. This aligns with analytic findings presented in Azar (2012). However, we stress that this result is mainly because of the firm in our model being a local monopolist with abundant labor supply, in which case the tip credit functions as a form of per-unit subsidy, thereby increasing the output in the market (Robinson 1969, chapter 13). Still, more research is required to understand the implications of tip credits in competitive markets with mobile workers. In Section 3.3, we suggest an approach to model worker mobility and utilize it to study tip abolishing.

3.2. Tipping and Quality

One conventional argument among economists views tipping as an incentive mechanism for workers to put effort into delivering high-quality service to customers (Jacob and Page 1980). The rationale behind this claim is that the tip is paid by the customer, who, unlike the manager, directly experiences the service, can effortlessly measure its quality, and, in turn, rewards a positive service experience with a generous tip (Zeithaml et al. 1988). Hence, it is natural to expect that servers strategically adjust quality provision to increase tip revenue (Barkan and Israeli 2004). In the literature, however, only weak empirical evidence is found for this view (Lynn and McCall 2000). We investigate this argument through the lens of our framework. In a closely related theoretical model, Azar (2008) endogenizes the service quality in a game between a service provider (determining effort for service) and a customer (determining the tip), in which the social norm is exogenous. In this section, we discuss an extension of our framework that, similarly to Azar (2008), allows the service provider to choose the quality, yet unlike Azar (2008), the price and the social norm of tipping are endogenous and contingent on the service quality.

In what follows, we introduce the service quality level q. Similarly to Mussa and Rosen (1978), for customer type θ, a service quality of level q yields a gross utility of $q\theta$. The service quality level, and therefore the customer's willingness to pay, both depend on the worker's effort but also on factors that are beyond the worker's control. Let q_0 be the base quality offered to the customer by the firm if the worker puts the least acceptable effort in service. We consider q_0 as an exogenous parameter that is determined by factors such as infrastructure (location of the facility, accessibility, etc.), the service production system (food quality, menu offerings, etc.), and so on. The worker can improve the base service quality by putting "extra" effort in the form of friendliness, courtesy, consideration, or other aspects of customer service orientation alike. To increase the service quality by some amount $q_e \geq 0$, namely, from level q_0 to a total of $q_0 + q_e$, the worker has to exert additional effort at a cost of $c^{\text{effort}}(q_e) = (\gamma/2)(q_e)^2$, where $\gamma > 0$ is an exogenous parameter. In Section 2, we have assumed that $q_0 = 1$ and $q_e = 0$; now, we will consider q_e as the worker's decision variable. The firm cannot force the employee to exert additional effort in their work, yet the tipping mechanism might encourage them to do so.

Extending upon the equilibrium definition in Section 2.2, we study a rational-expectations equilibrium problem: the firm sets the price, and customers tip rationally, based on their believed service quality, whereas the worker exerts an optimal effort, holding fixed their belief about the firm and customers' actions. In equilibrium, these beliefs must coincide with the rational actions taken by all players (the worker, the firm, and the customers).

Let m, $\underline{\theta}$, and $q = q_0 + q_e$ be the customary tip, the cutoff type, and the (total) service quality, respectively. Similarly to Proposition 1, in equilibrium, the price p satisfies $p = q\underline{\theta} - c(m)$, and for any type $\theta \in [\underline{\theta}, \theta_H]$, the tip is given by

$$t(\theta; m, \underline{\theta}, q) = \begin{cases} m - \underline{d}, & \text{if } \theta \leq \hat{\theta}(m, \underline{\theta}, q), \\ (q(\theta - \underline{\theta}) + c(m))/2, & \text{otherwise}, \end{cases}$$

where $\hat{\theta}(m, \underline{\theta}, q) = \underline{\theta} - c(m)/q + 2(m - \underline{d})/q$. As before, $\hat{\theta}$ is interpreted as the type of customer who is indifferent between the conformity and appreciation tip. Similarly to Equation (5), the firm's optimal cutoff type in equilibrium is given by

$$\underline{\theta}^{\text{opt.}}(m, q) \in \arg \max_{\underline{\theta} \in [\theta_L, \theta_H]} (q\underline{\theta} - c(m)) \cdot \int_{\underline{\theta}}^{\theta_H} dF(\theta),$$

and the worker's optimal quality level is determined by the expected tip and the cost of effort:

$$q^{\text{opt.}}(m, \underline{\theta}) \in \arg \max_{q \geq q_0} ET(m, \underline{\theta}, q) - c^{\text{effort}}(q - q_0),$$

where

$$ET(m, \underline{\theta}, q) = \frac{\int_{\underline{\theta}}^{\theta_H} t(\theta; m, \underline{\theta}, q) dF(\theta)}{\int_{\underline{\theta}}^{\theta_H} dF(\theta)}.$$

Hence, in equilibrium, the triplet $(m, \underline{\theta}, q)$ must satisfy $m = ET(m, \underline{\theta}, q)$, $\underline{\theta} = \underline{\theta}^{\text{opt.}}(m, q)$, and $q = q^{\text{opt.}}(m, \underline{\theta})$.

From the formulation of the worker's decision problem, it is clear that the worker trades off higher tips for effort cost. For fixed m, $\underline{\theta}$, and q, the worker's marginal return to effort is given by

$$\frac{\partial}{\partial q} ET(m, \underline{\theta}, q) = \frac{1}{2} \cdot \frac{\int_{\hat{\theta}(m, \underline{\theta}, q)}^{\theta_H} (\theta - \underline{\theta}) dF(\theta)}{\int_{\underline{\theta}}^{\theta_H} dF(\theta)} \geq 0.$$

This expression represents the marginal increase, with respect to q, in the share of the average tip income that is accrued from appreciation tips. This is intuitive because when the customary tip (m) is fixed, a higher-quality service increases the customer surplus and thereby the customer's appreciation tip, whereas the conformity tip remains unchanged. Recall that, according to Proposition 3, in equilibrium under strong social norm (namely $\kappa \to +\infty$), $\hat{\theta}$ approaches θ_H, and the volume of appreciation tips dwindles to zero. Therefore, the worker's marginal return to effort drops to zero as well: $(\partial/\partial q)ET \to 0$. In this case, no effort will be exerted, and the equilibrium quality will remain at q_0. Thus, there is little incentive for the worker to increase quality when the social pressure is high.

In light of the above, it is natural to expect that tipping provides stronger incentives for servers to put effort when the social-pressure level is low; hence, we focus next on the weak-social-norm regime ($\kappa \to 0$). Recall that under weak social norm, in equilibrium, all customers tip their appreciation tips (see Proposition 3). Assume for simplicity that types are uniformly distributed; hence, $(\partial/\partial q)ET = (\theta_H - \underline{\theta})/4$, which is half of the average surplus in the consumption set. The optimal cutoff type (for every q and m, as in Proposition 4) is $\theta_H/2$. By equating the marginal cost of effort, $\gamma(q - q_0)$, to the marginal return, we find the employee's equilibrium effort level, which is $\theta_H/(8\gamma)$ (see Appendix , section 1.9 for details). To get an assessment of how tips perform as a quality control mechanism, we compare this effort level with the effort that maximizes the firm's profit if the latter were able to fully control the worker's effort and incur the cost of it. In this case, the firm's optimal solution is found via solving

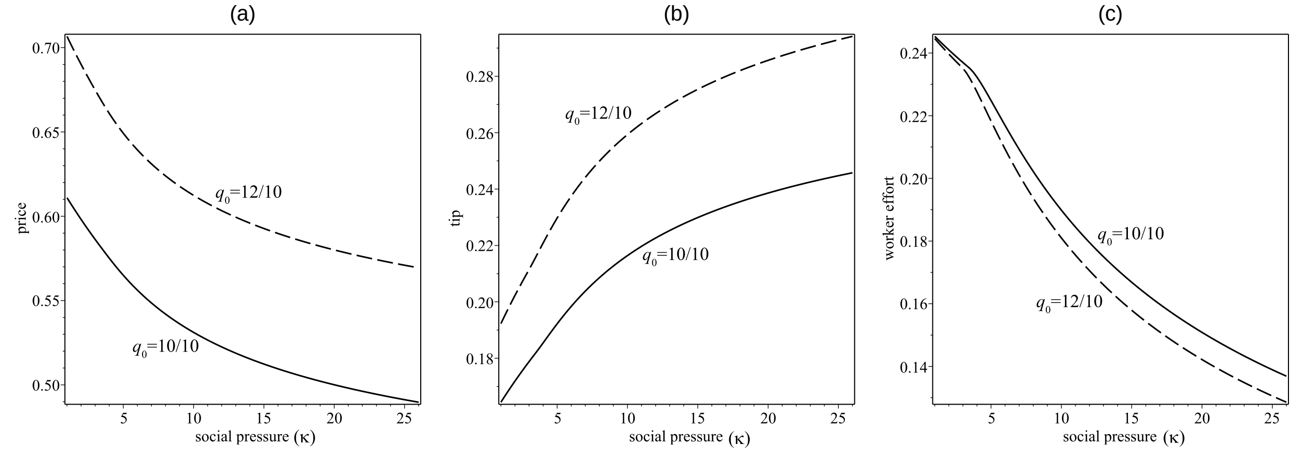
$$\max_{q \geq q_0, \underline{\theta} \in [\theta_L, \theta_H]} (q\underline{\theta} - c^{\text{effort}}(q - q_0)) \cdot (\underline{\theta} - \theta_L)/(\theta_H - \theta_L);$$

the marginal return to an increase in quality is simply $\underline{\theta}$, and the firm's optimal effort equals $\underline{\theta}^*/\gamma$, where the optimal cutoff, $\underline{\theta}^*$, satisfies $\underline{\theta}^* \geq \theta_H/2$ (see Appendix, section 1.9). In other words, the firm would have set the effort level above $\theta_H/(2\gamma)$, that is, more than four times larger than the equilibrium level of effort set by the tipped worker (which is $\theta_H/(8\gamma)$). Thus, even in the absence of social pressure, when all customers tip for appreciation, tips are hardly enough to incentivize the worker to invest even a quarter of the effort the firm could hope for, let alone when social pressure is present.

Overall, this suggests that tipping, which is driven by social pressure and appreciation, is not a very efficient mechanism for quality improvement. This resonates with empirically observed evidence that tips are relatively "blunt" instruments for motivating workers to increase quality (see Lynn and McCall 2000; Azar 2020, p. 225). In the case of a strong social norm, social pressure inflicts costs on customers that force the price low and the average tip high, yet the majority of customers tip only to conform with the norm, making their tips insensitive to quality improvement. Without social pressure, customers tips are sensitive to quality because higher quality increases customer surplus and therefore their appreciation tip, yet the firm, striving to appropriate as much surplus as possible through pricing, leaves the customer only a little surplus to tip from. In general, the worker's marginal return to effort, which is due to appreciation tips, only captures a limited fragment of the aggregate increase in customer surplus and therefore provides the worker with only little incentive to contribute to service quality.

We plot in Figure 4 the equilibrium price, tip, and worker effort as functions of the social-pressure level κ. Not surprisingly, greater social pressure reduces the effort, as well as the price, but increases the tip. Therefore, the tip-to-price ratio increases too. This

Figure 4. Equilibrium Price, Tip, and Worker Effort as Functions of the Social Pressure, κ, for $\sigma = 1/4$, $q_0 = 10/10$ (Solid Line), and $q_0 = 12/10$ (Dashed Line), with Uniformly Distributed Types over [0, 1], Quadratic Cost of Effort $c^{\text{effort}}(e) = \frac{\gamma}{2}e^2$ with $\gamma = 1/2$, and Quadratic Social Pressure $K(d) = (\kappa/2)(\{d\}^+)^2$



Notes. (a) Equilibrium price p. (b) Tip m. (c) Worker effort e.

shows that our findings in Proposition 3 are robust and further apply when the quality is endogenous. Noticeably, when the base quality is higher (dashed line), the price and average tip are higher, but the worker effort level is lower. This is because in our model, q_0 and q_e, which, respectively, are the firm and worker's contributions to quality, are perfect substitutes, and the worker exerts effort at level $\{q^{\text{opt.}}(m,\underline{\theta}) - q_0\}^+$, which is decreasing with q_0. The latter point implies that in terms of service quality, restaurants that already provide high service quality to start with benefit less from the tipping mechanism compared with restaurants with poor initial quality. Hence, if such a restaurant decides to do away with tipping, its service quality is not likely to deteriorate significantly. This supports empirical results found by Lynn and Brewster (2018), showing that high-end restaurants see a smaller drop in quality than low-end restaurants when they eliminate tipping. Yet our model offers an explanation that is different than that provided in Lynn and Brewster (2018). Whereas Lynn and Brewster (2018) hypothesize that upscale restaurants, being more resourceful, can easily replace tipping with different means of service quality control, our analysis suggests instead that tips in upscale restaurants reflect the customer's enjoyment from the overall experience, to which the server's contribution is relatively small compared with factors directly controlled by the manager such as the food quality or the ambience.

Finally, we turn back to the concept of tip pooling previously discussed in Section 3.1. Here, in Section 3.2, we assumed that each worker individually reaps the returns of their effort in the form of the tip they receive from the customer they have served. However, with tip pooling, the returns to the individual workers' effort are reduced, as each worker's income now depends on the effort of all other workers as well. In the extreme, when many workers pool tips together, the impact of the individual worker on the average tip becomes negligible. This encourages workers to freeload such that in equilibrium, none of them will exert any effort beyond what is minimally required. This observation is consistent with Wilson (2019, p. 678), who finds that tip pooling can create conflicts among workers, especially when one is not "pulling their weight" as much as the others. Hence, tip pooling can damage the quality of service and, as a result, also the customer experience and firm's profit.

3.3. Abolishing Tipping

As soon as tipping emerged in the United States, efforts have been made to abolish it. For example, in early 1900s, tip-ban laws were passed in Washington (Washington State Legislature 1909), only to be repealed later. Recently, some restaurant chains announced abolishing tipping, but reverted later back to tipping (Moskin 2020). One might think that tips are equivalent to service fees (or increased menu prices) and therefore could be easily replaced by simply adding the average tip amount to the current price. Presumably, this will ensure the firm could pay the workers a fixed wage that matches the tipped income without sacrificing its profits, resulting in both the firm and the worker being exactly equally off as with tips. It could even be claimed that such a move benefits the customers for not having to endure the social pressure for tipping. However, in arguing so, one clearly ignores the different objectives, alternatives, and incentives of the different players. In particular, tips, unlike fixed wages, are not subject to the employer's decision and differ from one customer to another. To adequately address the issue of tip abolishing, it is necessary to consider the tip-abolishing firm's optimal price and offered worker wage in the presence of outside labor opportunities for workers. Below, we do so through a simple extension of our base model.

Consider the firm as a local monopolist embedded in an industry in which the average tip is m^*. With tipping, the demand (hence, the required workforce) is given by $\bar{F}(\underline{\theta}^*)$, and the per-service tip income, ex ante, is m^*. The firm may choose to abolish tipping, thus sparing its customers the cost associated with tipping. However, because workers do not receive tips, the firm needs to pay them a fixed wage w per unit service.[15] This wage w and the service price are the firm's decision variables. To capture the mobility of workers in the labor market, we assume the potential labor supply for a fixed wage w is $S(w) = \bar{F}(\underline{\theta}^*)\left(\frac{w}{m^*}\right)^{\beta}$, where $\beta > 0$. Thus, when the wage is equal to (or larger than) the expected tip income, the firm can retain (or expand) its existing workforce; otherwise, it has to shrink it. The parameter β measures the level of worker mobility. Setting $\beta \gg 1$ corresponds with a labor market with many outside opportunities and low job requirements, so workers can easily switch jobs if their wages are low. By contrast, $\beta \ll 1$ models a monopsonistic labor market with specialized labor and nontransferrable skills—switching between jobs is costly, and workers may compromise on their wage. We next study how labor mobility affects the firm's incentive to abolish tipping.

When customers do not tip and workers are paid directly by the firm, the price coincides with the service valuation of the cutoff customer, and the firm's net revenue per service is $\underline{\theta} - w$. The firm's profits are then expressed as

$$\max_{w \geq 0,\, \underline{\theta} \in [\theta_L, \theta_H]} (\underline{\theta} - w)\min\{\bar{F}(\underline{\theta}), S(w)\}. \tag{9}$$

Clearly, the firm never sets w such that $S(w) > \bar{F}(\underline{\theta})$; otherwise, it could cut down the wage while still satisfying all the demand. Similarly, setting $\underline{\theta}$ such that

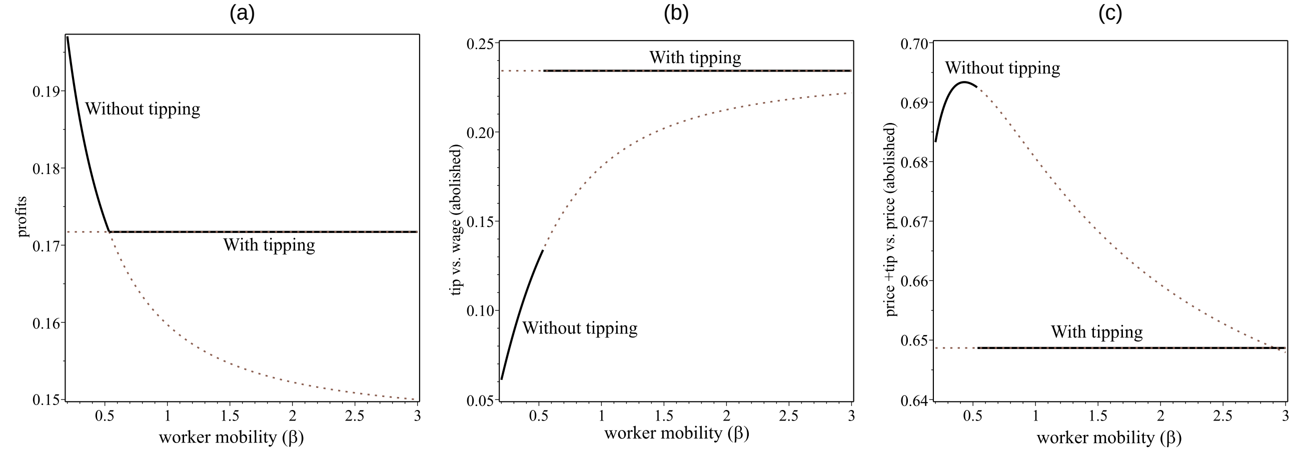
$S(w) < \bar{F}(\underline{\theta})$ is suboptimal because, then, the firm can increase the price while keeping the demand for service greater than the supply. Thus, the firm's optimal solution satisfies $S(w) = \bar{F}(\underline{\theta})$, or equivalently, $w = S^{-1}(\bar{F}(\underline{\theta}))$, and the optimization problem in (9) can be rewritten and solved via $\max_{\underline{\theta}\in[\theta_L,\theta_H]} (\underline{\theta} - S^{-1}(\bar{F}(\underline{\theta}))\bar{F}(\underline{\theta})$.

When workers are perfectly mobile, that is, $\beta \to +\infty$ (equivalently, labor supply elasticity, βm^*, is infinite), any decrease in the wage below the average tip m^* results in all workers leaving the firm, whereas setting the wage at (or above) m^*, the firm can employ as many workers as needed. Thus, the firm must set a wage $w = m^*$, and the optimal cutoff (i.e., price) is found solving $\max_{\underline{\theta}\in[\theta_L,\theta_H]} (\underline{\theta} - m^*)\bar{F}(\underline{\theta})$. Recall from Equation (5) that with tipping, the optimal cutoff solves $\max_{\underline{\theta}\in[\theta_L,\theta_H]} (\underline{\theta} - c(m^*))\bar{F}(\underline{\theta})$, and by Proposition 1, $c(m^*) < m^*$. Therefore, when labor is perfectly mobile, retaining a worker with a fixed wage is more costly to the firm than allowing them to collect tips. The intuition is the following. When tipping, customers of higher types contribute more to the worker's income than those of the lower types. Thus, as suggested by Lynn (2017), tipping acts as a natural mechanism for *price discrimination*, namely, encouraging customers of higher types to spend more on the same service than those of lower types. However, when tipping is abolished and workers are paid a wage $w = m^*$, a cost m^* is deducted from the profit for every unit of service. This means that the cost incurred by the cutoff customer increases (from $c(m^*)$ to m^*), and the firm can extract less surplus from the customers so that abolishing tipping becomes unprofitable. On the other hand, when there are few outside opportunities for workers ($\beta \to 0$), the workers are "trapped" with the firm. Abolishing tipping allows the firm to reduce the worker's wage to zero while still catering to all customers, making it more profitable to abolish tips than to allow them.

In Figure 5, we plot the firm's profit, worker wage, and service price as functions of β in two cases: when tipping is abolished and when it is allowed (in which case, the outcome is invariant of β). In each panel, the solid curve corresponds to the more profitable choice between the two. In Figure 5(a), the decreasing curve depicts the firm's profits when abolishing; thus, the horizontal line depicts the firm's profit when customers tip, and the solid curve indicates the maximum profit. As expected, when β is low, abolishing tips is more profitable, and the advantage of abolishing shrinks as β increases until it is no longer favorable. In Figure 5(b), we plot the worker's income, which is the optimal fixed wage w in the case that the firm abolishes tipping; otherwise, it is the expected tip, m^*. It can be seen that the individual worker is always better off with tips and that abolishing tips is more attractive when the potential reduction in labor cost is substantial. In Figure 5(c), we plot the optimal price when the firm abolishes tipping and compare it to the price plus the (average) tip when tips are accepted. Interestingly, we observe that in the tip-abolishing firm, the price is higher than its counterpart plus the tip, especially when β is such that abolishing tips is the profitable option. Hence, keeping all else equal, a revenue-maximizing firm that abolishes tipping will increase the price by more than the expected tip.

In this subsection, we focused on the trade-offs faced by the firm in the context of retaining labor and show that in a market with a social norm of tipping and high labor mobility, the firm will favor tips over fixed service charges. It is natural to expect that if the firm can rely on tips to finance the worker's wage

Figure 5. (Color online) Abolishing Tipping: Profits, Wages, and Prices When Abolishing vs. Not Abolishing Tipping When the Supply of Workers Is Determined by $S(w) = \bar{F}(\underline{\theta}^*)\left(\frac{w}{m^*}\right)^\beta$



Notes. (a) Profits. (b) Wages. (c) Prices. The solid lines in all panels represent the profits, wages, and prices, respectively, at the firm's profit-maximizing decision (abolishing vs. not abolishing tipping), and the dotted lines represent the (less profitable) alternative.

(e.g., via a tip credit), its incentive to allow tipping will further grow. Azar (2004c) points out that another reason for firms to stick to tipping is the possible positive effect on service quality (which we analyze separately in Section 3.2), though, as mentioned, research on the causal effect of tipping on service quality is inconclusive. Lastly, we acknowledge a drawback in that we single out one firm in an industry and ignore how its decision may affect the tipping norm and the decisions made by other firms. Yet, given the high mobility of workers and the well-established norm of tipping in U.S. restaurants, our analysis suggests that even in the absence of tip credits, tipping can stably prevail. Thus, an industry-wide change in tipping norms is not likely to arise spontaneously following the decisions of only a few businesses.

4. Concluding Remarks

In previous theoretical literature on tipping, the majority of papers have considered the norm of tipping as an exogenous economic force. To the best of our knowledge, Azar (2004a) is the only paper prior to ours that studies the endogenous evolution of a tipping norm. Our framework builds on Azar (2004a) by further incorporating the firm's pricing decisions, thereby forming the norm of tipping endogenously through sequential dynamic adjustments of both the price and the tipping behavior. This dynamic process formulation promotes a better understanding of the long-run implications of policy changes on tipping by allowing us to differentiate short-term (intertemporal) effects from long-term (equilibrium) effects of policies and managerial interventions. We build on this formulation in Section 3 to discuss the long-run effects of changing the tip credit, monitoring quality through tips, and abolishing tipping. Our modeling assumptions, as well as our applications, can be revisited and tailored to address different specific research questions. Below, we outline several potential directions and motivations for future research based on our framework.

4.1. Heterogeneity and Inequity

In our model, customers are only differentiated with respect to their willingness to pay. In reality, they also differ in their sensitivity to social pressure and their feelings of appreciation toward the server. For example, social pressure can vary among customers who dine alone versus those in groups of friends, potentially leading to higher tips. Other sources of heterogeneity that appear to be correlated with tipping include the time of day the customer frequents the facility and even sociopolitical affairs—for instance, an empirical work by Zhang et al. (2023) indicates that customers' empathy effects toward gig workers during times of political protests affect their tipping behavior. Evidence shows that workers are indeed aware of customers' heterogeneity and leverage this knowledge when planning their shift schedules or choosing which tables to wait on (Wilson 2019). Rather than allocating workers randomly to customers, as assumed in Section 3.2, managers can deploy methodical guidelines for matching workers with customers to increase efficiency. Furthermore, workers often use customers' visible attributes, such as race or gender, to speculate on how much tip they will receive and determine their service effort accordingly. This may motivate providers to discriminate against customers with certain attributes (e.g., by providing them with poorer service), as suggested in Lynn (2017) and Azar (2020). Analyzing the negative impact of such discrimination by introducing attributes that are imperfectly correlated with the customer's willingness to pay can add clarity to the discussion about abolishing tipping.

4.2. Cognitive Limitations and Tipping

Restaurant tipping data suggest that customers are more sensitive to prices than tips, as cognitive limitations make it hard for customers to account correctly for the full cost before making a purchasing decision (Greenleaf et al. 2016, Lynn 2017). This is relevant especially for businesses that consider replacing tips with fixed service charges or increased menu prices. In addition, cognitive limitations make customers determine the tip based on a (rounded) percentage of the price. Sometimes, customers' perception of the customary or average tip can be inaccurate. New technologies, such as mobile apps and Square displays in coffee shops, already take advantage of such cognitive limitations in making tip suggestions (Donkor 2019). These technologies not only convey suggested tip amounts to customers, but also allow firms to influence customers' tipping behavior via optimizing the offered menu of choices (Alexander et al. 2021). Studying efficient designs of tip-choice menus can be both an intriguing and practical research avenues to pursue.

4.3. Tip Pooling

Tip pooling encourages free riding among workers and reduces their incentives to provide effort. In the presence of tip credits, tip pooling may also facilitate tip appropriation by the employer (Section 3.1). However, it mitigates the impact of income variability on workers' income (Section 3.2). This duality makes it challenging to determine who benefits from the practice and to what extent. Moreover, criticism often arises against tip pooling for being unfair because the service of a single customer usually involves the cooperation of a chain of workers, many of whom are excluded from the tip pool. For example, in a restaurant, cooks, bussers, waiters, and managers work together to create value for the customer, yet only the customer-facing workers receive tips. Some voluntary tip-pooling arrangements (referred to as "tip outs") allow the non-customer-facing, lesser-paid workers to receive a limited share, usually smaller than a waiter's share, of the tip pool (Lynn 2017). A more detailed model of the service delivery process would help better understand how tip pooling impacts service quality, worker welfare, and equity among workers.

To conclude, no single model can fully account for a phenomenon as complex as tipping. Yet we believe that parsimonious modeling provides crisp and intuitive insights by disentangling and isolating the different economic forces at play. We hope our work will inspire further exploration of the operational and managerial aspects of tipping and help address unsettled questions in the ongoing public discussion surrounding this significant, compelling phenomenon.

Endnotes

[1] In our model, we set the parameter q in Charness and Rabin (2002) to zero, as the customer always attributes their obtained surplus $\theta - p$ to the server. The results can be easily generalized to any distributional preference function taking the value $\sigma_1\pi_C + (1-\sigma_1)\pi_W$ if $\pi_W \le \rho\pi_C$ and $\sigma_2\pi_C + (1-\sigma_2)\pi_W$ otherwise, assuming $0<\sigma_1 < 1/2 < \sigma_2 < 1$ and $\rho > 0$. To simplify the exposition, we set $\sigma_2 = 1 - \sigma_1$ and $\rho = 1$.

[2] Our results can be easily generalized to incorporate positive utility (i.e., negative cost) when tipping more than the customary tip via $K(d) - \lambda d^+$, where $\lambda > 0$.

[3] The additive form of the utility with respect to the reward and social-pressure components is consistent with the framework laid out by Helbing (2010, section 2.2) and the utility model of Levitt and List (2007).

[4] To keep focused, in the statements of all the results in this section, we will implicitly assume that the rational demand is nonzero, $\int_{\theta\in C^*} dF(\theta) > 0$.

[5] When v takes the generalized form $\sigma_1\pi_C + (1-\sigma_1)\pi_W$ if $\pi_W \le \rho\pi_C$ and $\sigma_2\pi_C + (1-\sigma_2)\pi_W$ otherwise (where $0 < \sigma_1 < 1/2 < \sigma_2 < 1$ and $\rho > 0$), then the minimal tip (hereafter, the "appreciation tip") becomes $\rho(\theta - p)/(1+\rho)$. As explained, this does not affect the paper's qualitative results.

[6] Note that when the price is set by a profit-maximizing firm (as we shall assume later on), we can exclude the case where all customers in $[\theta_L, \theta_H]$ receive strictly positive net utility.

[7] We assume that the maximizer in (5) is unique; otherwise, we follow convention and take the minimum over all optimal solutions for (5). We later impose Assumption 2, which implies uniqueness of the maximizer.

[8] Symmetrically, for $m_0 > m^*$, the average-tip sequence is decreasing, and the price sequence is increasing.

[9] This is not the case for any finite κ because for proper primitives, in equilibrium, some customers tip for appreciation.

[10] For types uniformly distributed over $[\theta_L, \theta_H]$, under Assumption 1, it must hold that $\theta_L < \theta_H/2$.

[11] More accurately, the type distribution F increases in the first-order stochastic-dominance sense. The same conclusion holds also under right shifting, that is, replacing $[\theta_L, \theta_H]$ by $[\theta_L + \epsilon, \theta_H + \epsilon]$ for $\epsilon > 0$.

[12] Though there is evidence indicating that the likelihood of taxi passengers leaving tips in the Chicago taxi market was declining during the COVID-19 pandemic (Conlisk 2022).

[13] In this formulation, the minimum wage, $\underline{w}$, does not impact the customer distributional preference v in a direct way.

[14] There are small regions in which the firm's profit slightly decreases because of customers' increased cost of tipping that comes along with raising the tip credit.

[15] Unlike Section 3.1, we assume here that the minimum wage and tip credit are zero and workers are risk neutral.

References

Alexander D, Boone C, Lynn M (2021) The effects of tip recommendations on customer tipping, satisfaction, repatronage, and spending. *Management Sci.* 67(1):146–165.
Azar OH (2004a) What sustains social norms and how they evolve? The case of tipping. *J. Econom. Behav. Organ.* 54:49–64.
Azar OH (2004b) The history of tipping—From sixteenth-century England to United States in the 1910s. *J. Socio-Econom.* 33(6): 745–764.
Azar OH (2004c) Optimal monitoring with external incentives: The case of tipping. *Southern Econom. J.* 71(1):170–181.
Azar OH (2005) The social norm of tipping: Does it improve social welfare? *J. Econom.* 85:141–173.
Azar OH (2007) Why pay extra? Tipping and the importance of social norms and feelings in economic theory. *J. Socio-Econom.* 36(2):250–265.
Azar OH (2008) Strategic behavior and social norms in tipped service industries. *B.E. J. Econom. Anal. Policy* 8(1):1–18.
Azar OH (2010) Tipping motivations and behavior in the U.S. and Israel. *J. Appl. Soc. Pysch.* 40(2):421–457.
Azar OH (2011) Business strategy and the social norm of tipping. *J. Econom. Psych.* 32(3):515–525.
Azar OH (2012) The effect of the minimum wage for tipped workers on firm strategy, employees and social welfare. *Labour Econom.* 19(5):748–755.
Azar OH (2020) The economics of tipping. *J. Econom. Perspect.* 34(2):215–236.
Barkan R, Israeli A (2004) Testing servers' roles as experts and managers of tipping behaviour. *Service Indust. J.* 24(6):91–108.
Ben-Zion U, Karni E (1977) Tip payments and the quality of service. Ashenfelter OC, Oates WE, eds. *Essays in Labor Market Analysis* (John Wiley & Sons, New York), 37–44.
Bodvarsson OB, Gibson WA (1997) Economics and restaurant gratuities: Determining tip rates. *Amer. J. Econom. Sociol.* 56(2):187–203.
Camerer CF (2011) *Behavioral Game Theory: Experiments in Strategic Interaction* (Princeton University Press, Princeton, NJ).
Charness G, Rabin M (2002) Understanding social preferences with simple tests. *Quart. J. Econom.* 117(3):817–869.
Clifton SM, Herbers E, Chen J, Abrams DM (2018) The tipping point: A mathematical model for the profit-driven abandonment of restaurant tipping. *Chaos* 28(2):023109.
Conlisk S (2022) Tipping in crises: Evidence from Chicago taxi passengers during COVID-19. *J. Econom. Psych.* 89:102475.
Donkor K (2019) The economic value of norm conformity and menu opt-out costs: A field study in the taxi industry. Preprint, submitted October 1, https://doi.org/10.2139/ssrn.3955553.
Even WE, Macpherson DA (2014) The effect of the tipped minimum wage on employees in the us restaurant industry. *Southern Econom. J.* 80(3):633–655.
Fehr E, Gächter S (2000) Fairness and retaliation: The economics of reciprocity. *J. Econom. Perspect.* 14(3):159–182.
Fehr E, Schmidt KM (1999) A theory of fairness, competition, and cooperation. *Quart. J. Econom.* 114(3):817–868.
Festinger L (1957) *A Theory of Cognitive Dissonance*, vol. 2 (Stanford University Press, Stanford, CA).
Forsythe R, Horowitz JL, Savin NE, Sefton M (1994) Fairness in simple bargaining experiments. *Games Econom. Behav.* 6(3):347–369.

Greenleaf EA, Johnson EJ, Morwitz VG, Shalev E (2016) The price does not include additional taxes, fees, and surcharges: A review of research on partitioned pricing. *J. Consumer Psych.* 26(1):105–124.

Heider F (1946) Attitudes and cognitive organization. *J. Psych.* 21(1): 107–112.

Helbing D (2010) *Quantitative Sociodynamics: Stochastic Methods and Models of Social Interaction Processes* (Springer, Berlin, Heidelberg).

Jacob N, Page A (1980) Production, information costs, and economic organization: The buyer monitoring case. *Amer. Econom. Rev.* 70(3):476–478.

Jones MR (2016) Measuring the effects of the tipped minimum wage using W-2 data. CARRA Working Paper Series No. 2016-03, Center for Administrative Records Research and Applications, U.S. Census Bureau, Washington, DC.

Kahneman D, Knetsch JL, Thaler R (1986) Fairness as a constraint on profit seeking: Entitlements in the market. *Amer. Econom. Rev.* 76(4):728–741.

Kostami V, Kostamis D, Ziya S (2017) Pricing and capacity allocation for shared services. *Manufacturing Service Oper. Management* 19(2):230–245.

Levitt SD, List JA (2007) What do laboratory experiments measuring social preferences reveal about the real world? *J. Econom. Perspect.* 21(2):153–174.

List JA (2007) On the interpretation of giving in dictator games. *J. Political Econom.* 115(3):482–493.

Lynn M (2006) Tipping in restaurants and around the globe: An interdisciplinary review. Altman M, ed. *Handbook of Contemporary Behavioral Economics: Foundations and Developments* (Routledge, New York), 626–643.

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

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

Lynn M, Brewster ZW (2018) A within-restaurant analysis of changes in customer satisfaction following the introduction of service inclusive pricing or automatic service charges. *Internat. J. Hospitality Management* 70:9–15.

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

Mai Y, Hu B (2022) Optimizing free-to-play multiplayer games with premium subscription. *Management Sci.* 69(9):3437–3456.

Margalioth Y (2006) The case against tipping. *Univ. Pennsylvania J. Labor Employment Law* 9(1):117–146.

Margalioth Y (2010) The social norm of tipping, its correlation with inequality, and differences in tax treatment across countries. *Theoret. Inquiries Law* 11(2):561–588.

Momot R, Belavina E, Girotra K (2020) The use and value of social information in selective selling of exclusive products. *Management Sci.* 66(6):2291–2799.

Moskin J (2020) Danny Meyer's restaurants will end their no-tipping policy. *New York Times* (July 20), https://www.nytimes.com/2020/07/20/dining/danny-meyer-no-tips.html.

Mussa M, Rosen S (1978) Monopoly and product quality. *J. Econom. Theory* 18(2):301–317.

Muth JF (1961) Rational expectations and the theory of price movements. *Econometrica* 29(3):315–335.

Pencavel JH (2015) Work effort, on-the-job screening, and alternative methods of remuneration. Polachek SW, Tatsiramos K, eds. *35th Anniversary Retrospective*, Research in Labor Economics, vol. 35 (Emerald Group Publishing Limited, Leeds, UK), 537–570.

Post L, Senning DP (2022) *Emily Post's Etiquette* (Ten Speed Press, Berkeley, CA).

Robinson J (1969) *The Economics of Imperfect Competition* (Palgrave Macmillan, London).

Roels G, Su X (2014) Optimal design of social comparison effects: Setting reference groups and reference points. *Management Sci.* 60(3):606–627.

Ruffle BJ (1998) More is better, but fair is fair: Tipping in dictator and ultimatum games. *Games Econom. Behav.* 23(2):247–265.

Schein JE, Jablonski EF, Wohlfahrt BR (1989) *The Art of Tipping: Customs and Controversies* (Tippers International, Oshkosh, WI).

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

Star N (1988) *The International Guide to Tipping* (Berkley Publishing Group, New York).

Tan T, Netessine S (2019) When you work with a superman, will you also fly? An empirical study of the impact of coworkers on performance. *Management Sci.* 65(8):3495–3517.

Tereyağoğlu N, Veeraraghavan S (2012) Selling to conspicuous consumers: Pricing, production, and sourcing decisions. *Management Sci.* 58(12):2168–2189.

U.S. Department of Labor, Wage and Hour Division (2011) The Fair Labor Standards Act of 1938, as amended, sec. 203(m). Accessed March 1, 2024, https://www.dol.gov/sites/dolgov/files/WHD/legacy/files/FairLaborStandAct.pdf.

U.S. Department of Labor, Wage and Hour Division (2023) Minimum wages for tipped employees. Accessed September 12, 2023, https://www.dol.gov/agencies/whd/state/minimum-wage/tipped.

Washington State Legislature (1909) Criminal code, chapter 249, sections 439, 440. Accessed June 30, 2025, https://archive.org/details/walaw-session-laws-1909/1909v1/page/1028/mode/2up?view=theater.

Wilson ER (2019) Tip work: Examining the relational dynamics of tipping beyond the service counter. *Symbolic Interaction* 42(4):669–690.

Zeithaml VA, Berry LL, Parasuraman A (1988) Communication and control processes in the delivery of service quality. *J. Marketing* 52(2):35–48.

Zhang ZJ, Deng W, Hu M (2023) Riding through rallies: Will you tip more? Preprint, submitted August 28, https://doi.org/10.2139/ssrn.4553134.