NBER WORKING PAPER SERIES

THE DRIVERS OF SOCIAL PREFERENCES:
EVIDENCE FROM A NATIONWIDE TIPPING FIELD EXPERIMENT

Bharat Chandar
Uri Gneezy
John A. List
Ian Muir

Working Paper 26380
http://www.nber.org/papers/w26380

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2019

The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w26380.ack

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2019 by Bharat Chandar, Uri Gneezy, John A. List, and Ian Muir. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

The Drivers of Social Preferences: Evidence from a Nationwide Tipping Field Experiment
Bharat Chandar, Uri Gneezy, John A. List, and Ian Muir
NBER Working Paper No. 26380
October 2019
JEL No. C93,D63,D64

ABSTRACT

Even though social preferences affect nearly every facet of life, there exist many open questions on the economics of social preferences in markets. We leverage a unique opportunity to generate a large data set to inform the who's, what's, where's, and when's of social preferences through the lens of a nationwide tipping field experiment on the Uber platform. Our field experiment generates data from more than 40 million trips, allowing an exploration of social preferences in the ride sharing market using big data. Combining experimental and natural variation in the data, we are able to establish tipping facts as well as provide insights into the underlying motives for tipping. Interestingly, even though tips are made privately, and without external social benefits or pressure, more than 15% of trips are tipped. Yet, nearly 60% of people never tip, and only 1% of people always tip. Overall, the demand-side explains much more of the observed tipping variation than the supply-side.

Bharat Chandar
Stanford University
579 Serra Mall
Stanford, CA
chandarbharatk@gmail.com

Uri Gneezy
Rady School of Management
University of California - San Diego
Otterson Hall, Room 4S136
9500 Gilman Drive #0553
La Jolla, CA 92093-0553
ugneezy@ucsd.edu

John A. List
Department of Economics
University of Chicago
1126 East 59th
Chicago, IL 60637
and NBER
jlist@uchicago.edu

Ian Muir
Lyft
muir.ian.m@gmail.com

"We pay that tax knowing it to be unjust and an extortion; yet we go away with a pain at the heart if we think we have been stingy with the poor fellows." ~Mark Twain

Introduction

Tipping has a long and storied history in modern economies. The practice of tipping is commonly believed to have started in the 17th century in Tudor, England, where overnight guests in private homes tipped the host's servants for excellent service.[1] The act of tipping soon spread to London, where commercial establishments, such as barbershops, smokehouses, and coffeehouses, adopted tipping. Tipping subsequently reached the United States, but met strong social resistance, as the epigraph suggests. A New York Times 1899 editorial noted that tipping was the "vilest of imported vices," arguing that customers were willing "to reward the servility" of servants, and those accepting the tips were deemed as "men among us servile enough to accept their earnings in this form." The Washington Post echoed similar sentiments, noting that tipping is "one of the most insidious and one of the most malignant evils" in today's world. Perhaps best highlighting the scrutiny of tipping during this era was Presidential hopeful William Taft, who was a well-known non-tipper, and ran for president in 1908 as "the patron saint of the anti-tip crusade."

Even though the anti-tipping movement reached its height soon after in 1915, when three states (Iowa, South Carolina, and Tennessee) joined three other states (Washington, Mississippi, and Arkansas) in abolishing tipping, the practice has now expanded well beyond waiters, porters, barbers, and bellhops. Today, it is not atypical for the barista, the sandwich maker, the driver, the shoe shiner, and the dry cleaner to receive tips. Around every corner it seems someone is asking

[1] As Azar (2004) notes, there are several different versions of the origination of tipping. Hemenway (1993), for example, cites the Roman era; Schein et al. (1984) asserts that the beginnings happened in the time of feudal lords; Segrave (1998) traces tipping back to the Middle Ages. The list goes on and on.

for and receiving a tip (see, e.g., Lynn et al. (1993), who consider 33 tipped service professions). In aggregate, some estimates suggest that in the United States alone $36.4 billion was given in tips in 2017 (Shierholz, et al. 2017).

While tipping has played an important social and economic role for centuries and remains a hallmark of modern service economies, economists' contributions in the area are limited because of data paucity. In this paper, we leverage a unique opportunity at Uber wherein we helped to develop and implement in-app customer tipping on Uber's ridesharing app. Prior to June 2017, Uber did not have a tipping feature on its app. We layered several experiments during the introduction of the new feature from June 20 to July 17, 2017. The release to customers was done in a rolling spatial and temporal fashion across the United States, and we ran several field experiments within the app itself to deepen our scientific understanding of tipping behaviors and social preferences more generally.

Our data set includes over 40 million observations of people who are engaged in routine tipping behavior. The data allow us to shed light on many open questions regarding social preferences and tipping. Beyond establishing facts around tipping, explaining what motivates people as social creatures is important for understanding how the economy functions. Social preferences such as trust, altruism, and reciprocity have each served critical roles creating wealth in lieu of institutional rules and regulations.

For their part, economists have recently begun to include formally such factors in their economic models and empirical exercises. Extant literature studying these theoretical models use experimental games in areas such as gift exchange (Fehr et al. (1993); Charness (2004); Gneezy and List (2006)) and trust games (Joyce et al. 1995). An early account of this literature can be found in Fehr and Gächter (2000), while Charness and Kuhn (2011) offers a more recent summary.

The empirical literature has been quite useful in developing our understanding of certain comparative statics, such as when participants are willing to pay money to punish what they consider an unfair act, and to distinguish between trust and reciprocity (Guth et al. (1982); Fehr and Gachter (2002); Cox (2004)).

A major obstacle in studying social preferences and parsing their underlying determinants outside of the laboratory setting has been the quality and depth of available data. In an effort to extend the social preference insights beyond the lab, researchers have recently moved the empirical approach from the lab to the field (see, e.g., Gneezy and List (2006) and List (2006); and the charitable giving literature summarized in Andreoni, et al. (2017)). Our study represents a combined field experiment and big data approach to deepen our understanding of the facts around tipping, and the underlying motivation for tipping.

We find several interesting insights from the data. For example, roughly 16% of Uber rides are tipped. Yet, most riders (60%) never tip over our four weeks of data collection. Of those who do tip, very few (1%) tip on every trip. The remainder of people only tip on about 25% of trips. This data pattern suggests a different picture of generosity preferences than one would anticipate from the extant literature wherein 60% of individuals transfer money to an anonymous stranger in dictator games (see List (2007); Engel (2011)). Yet, we find it noteworthy that even though the act of tipping is made privately, and without external social benefits or social pressure, roughly $0.50 is tipped on the average trip and when a tip is made more than $3 is tipped (26% of the fare).

Exploring the underlying variables that correlate with tipping provides a deeper level of tipping behaviors. For instance, when parsing the various explanations of tipping, we find that rider effects account for about three times more of the observed tipping variation than driver effects. In this spirit, rider ratings represent a key demand-side explanatory variable - we find that, for example, riders who have a 5-star rating tip more than twice as often as those with a 4.75 rating, and when they do tip they tip nearly 14% more. An interesting parallel to this finding is its link to the charity literature, where the individual donor characteristics are found to be much more important than the features of the charitable organization receiving the donation (List 2011). This insight reveals a new connection between tipping and charitable acts.

Other demand-side variables, including the place of residence, lifetime number of Uber trips, and gender, are important explanatory variables. When considering gender, for example, we find differences both in giving and receiving tips. Male riders tip 23% more than female riders, a result largely driven by the fact that men are more likely to tip than women (approximately 19% more often). Further, female drivers are tipped more than male drivers—a fact that is true regardless of rider gender: men (women) tip female drivers nearly 12% (11%) more than they tip male drivers. However, the tip premium that male riders pay to female drivers falls with the driver's age and disappears by the age of 65. This tipping difference results in a $0.05 per trip driver gender tip gap.[2] Our results add clarity to the extant literature, which identifies differences in the economic preferences of men and women (see Croson and Gneezy (2009) for a survey), with mixed results with respect to social preferences.

Although the demand side explains more of the overall variance, we find a number of supply-side variables that show a significant correlation with tipping, including driver age, experience, the app language used, the driver's ZIP code, and the driver rating. In terms of magnitudes, drivers with a 5-star rating are tipped close to 50% more often than those with a 4.75 rating, and when they do receive tips they are nearly 5% higher.

[2] In this manner, the gender pay gap that Cook et al. (2018) report (men earn $0.30 more per trip than women) is closed by 13% when accounting for tips, which were not part of the platform at the time of their analysis.

Beyond demand and supply-side determinants, we find that trip-related characteristics influence tipping patterns. Leveraging telematics data from drivers' phones, we find that quick accelerations, hard brakes, and speeding are all associated with lower tip levels. Similarly, not meeting estimated pick-up times lowers tips. Beyond quality of the trip, we also find that other features of the trip are correlated with tipping: i) tipping levels are concave in fares (measured at the means, a 10% increase in fare is associated with a 2.5% increase in tips); ii) tips tend to be highest for airport and business trips, and iii) trips that take place in small cities receive more tips than those in large cities.

While the tipping facts provide evidence on the where's, when's, and who's of tipping, interpreting the data patterns within the extant social preference literature is difficult. We combine experimental variation with non-experimental variation and use two approaches to this problem. First, we examine individuals' behavior over repeated interactions. We find that when the same rider matches with the same driver multiple times, they tip more on average each successive encounter. For example, when a rider matches with the same driver twice, they tip 27% more the second time than they do the first time. This clarifies that repeated interaction is an important input to an individual's tipping decisions. Importantly, we show that this result is not due to strategic reciprocity—driven by updated perceptions of the probability of meeting the same driver again—rather it is consonant with the repeat interaction building a greater social connection between the rider and driver.

Our second approach to understanding the underlying mechanism at work is through a complementary field experiment where we explore the effects of defaults. Within our nationwide field experiment, we varied the defaults the riders received when asked for a tip. Such defaults have been shown to be quite influential in a variety of domains (see, e.g., Johnson and Goldstein

(2003); Choi et al. (2003); Thaler and Benartzi (2004); Haggag and Paci (2014)). In our default field experiment, which includes more than 10 million observations, riders are allocated to different treatments in which they are exposed to a different preset of tip options. Our results are somewhat surprising given the literature, as we find only a modest effect of defaults: average tips as a percent of the fare increase by 2.5% between the lowest and highest default options. This finding is in contrast to existing literature, where Haggag and Paci (2014) find that higher defaults lead to a greater than 10% increase in tips.

The difference in magnitude in our results and those in Haggag and Paci (2014) is in line with the notion that defaults are less influential when the tipping decision is made privately. Our preferred interpretation of this result is that while social norms can be clearly influential (see, e.g,, Benabou and Tirole (2006)), their level of import relies critically on both the strength of the norm and whether the action is public in nature. A general insight naturally follows: norms and the monitoring of behavior are *complements*. As such, the strength of social nudges, or any particular course of action brought about through social norms, is moderated without public verification of an action. Alternatively, the effects of social nudges are enhanced as the veil of choice anonymity is removed.

The remainder of our paper is structured as follows. Section 2 provides details about the Uber tipping feature and a summary of our methodology. Section 3 outlines our key findings about tipping behavior, and examines the potential drivers of this behavior, focusing on demand, supply, and trip-related determinants of tipping. Section 4 discusses our analysis of the effect of norms and anchoring on tipping behavior. Section 5 concludes.

2. Overview of the Tipping Feature and Field Experiment Roll-Out

Prior to June 2017 Uber did not have in-app tipping on its platform. Passengers were free to provide drivers with a cash tip, but there was no way to add a tip for a driver through the Uber platform. We worked with product teams and top management at Uber to help conceive of, design, implement, and research the tipping product on the platform—three of us (List, Muir, and Chandar) in our capacities as employees of the company at the time (none of us currently work at Uber).

A key issue was changing the culture around the nature of tipping as a useful economic tool. After several months of meetings, the company announced that it would introduce the feature in June of 2017 as part of its "180 Days of Change" campaign geared towards improving the driver experience. Upon the introduction of tipping, the core user experience for the rider was as follows: once a trip was completed and the driver rated the rider, the rider would be prompted to return to the Uber app via a notification on their device asking them to rate the trip and notifying them of the option to tip. They would not be told how the driver rated them.

In practice, if passengers either clicked on the phone notification or returned to the Uber app, then they would be taken to the screen in which they had historically been given the option to rate the trip on a five-star scale. On this screen, they were now also invited to provide the driver an optional tip (see Figure 1 for the screen in question). In this case, the passenger was presented with three preset options, e.g. \$2, \$4, \$6 as in Figure 1, and then had the option to enter a custom amount (capped at the minimum of two times the fare or \$100 to reduce fraud). The rider was not required to provide a rating in order to tip and tipping itself was entirely optional.

Importantly, a driver could see the amount of a tip on a given trip in a list of historical trips in the app, but the option to tip was only presented to the passenger after the driver had rated the rider. This form of tipping differs importantly from tipping in other contexts in that the tipping decision is made without any social pressure from the driver.

The rollout was staggered across cities in part to ensure there were no bugs in the product and in part as an experiment at the city level to understand tipping's impact on the marketplace (see Chandar et al. (2019) for results from this market-level field experiment). We randomized three cities to receive tipping on June 20, 2017 (internally called the alpha launch), followed by half of the operational markets in the United States and Canada on July 6, 2017 (internally called the beta launch). Remaining operational markets in the United States and Canada launched the tipping feature on July 17, 2017 (the full roll-out).

In addition to this market-level experiment, the roll-out had several rider-level experiments. First, riders were randomized into one of three groups: (a) no tipping; (b) tipping in the standard user flow, where they could request another trip without needing to rate the previous trip; (c) tipping before returning to the standard user flow, where passengers had to consider rating and tipping of the previous trip before they were allowed to request their next trip. Riders who were eligible to tip were then in several experiments varying the preset options. In the first such experiment, which took place during the alpha and beta launches, the preset options (a), (b), and (c) shown to riders were randomized. In the second field experiment, carried out three weeks after all cities had launched tipping, the set of preset options shown to riders was randomized again, now with potentially different suggested tip values based on the price of the trip.

In total, our dataset is comprised of over 40 million UberX trips from across the United States from August 18, 2017 through September 14, 2017. For legal reasons, the driver needed to opt in to be eligible for receiving tips. Moreover, riders needed to have a recent version of the Uber app installed to be shown the tipping feature. We only include trips in which both the rider and the driver met these eligibility standards, leaving us with a sample of close to 90% of overall trips. We also exclude data from New York City, as it was treated independently from other cities in our experiment for business and legal reasons.

3. Tipping Facts

Our general approach in this section is to summarize the tipping facts as succinctly as possible. Given the breadth of our data and results, there are many interesting outcomes and approaches to the data analysis. We focus our discussion on robust results and follow the general rule of first presenting results from the raw data and then conditioning on various control variables that potentially impact the outcome variable. In terms of conditioning, we use the following regression model:

$$y_i = \beta X_i + \varepsilon_i$$

where the covariates for a trip $X_i = (\Lambda_i, \Gamma_{r(i)}, \Theta_{d(i)}, K_{l(i)})$ comprise trip-specific controls, Λ_i; rider-specific controls, $\Gamma_{r(i)}$; driver-specific controls, $\Theta_{d(i)}$; and fixed effects for the time and location of the trip, $K_{l(i)}$. We report the covariates included in Λ_i, $\Gamma_{r(i)}$, and $\Theta_{d(i)}$ in Appendix Section 1.1, 1.2 and 1.3 respectively. $K_{l(i)}$ includes fixed effects for hour of week, date of trip, starting location, and ending location.

Overall, we find that roughly 16% of trips on Uber are tipped. Conditional on a positive tip, \$3.11 is tipped on average, which corresponds to about 26% of the fare. Putting these two together yields an average tip of \$0.50 per trip. While these aggregate "what is tipped" numbers are interesting, exploring the when's and where's of tipping provides a first result:

Result 1: There is substantial temporal and spatial heterogeneity in tipping

Empirical support for Result 1 is contained in Figures 2-4, which provide a visual depiction of when throughout the week tips occur by hour. Tips tend to be highest during the very early morning hours, between 3:00 am and 5:00 am, when more than 17% of trips are tipped and the average tip is nearly $3.70. A disproportionate percentage of airport and business trips takes place during these hours. Passengers taking these trips may get reimbursed for travel expenses by their company. As a result, the observed tipping pattern could reflect either an income effect, or increased generosity when dealing with other's money. Tips also tend to be high on Friday and Saturday evenings around 6:00 pm. On the flip side, tips tend to be lowest each day around midnight, when only 13% of trips are tipped and the average tip amounts to less than $3. Performing standard tests of means, these results are significant whether we consider the percent of trips tipped (Figure 2), the average conditional tip (Figure 3), or the average tip (Figure 4).

Turning to spatial variation, we can explore tipping in several dimensions. We chose two: across cities and within a given city. Both reveal substantial heterogeneity. Average tips tend to be lower in large cities compared to small cities, as seen in Figure 5. Moreover, states in the middle of the country tend to have higher tip rates than the Northeast and California. These spatial differences are significant at conventional levels (city as the observation) using an F-test of spatial homogeneity.

To explore variation within specific cities, or neighborhood characteristic effects, it is important to recognize that Uber does not collect demographic information, such as household income, race, or education about riders or drivers. Yet, to form tests of spatial differences we join drivers' and riders' home ZIP codes with ZIP code-level demographic data from the U.S. Census. We consider median income, the percentage of black residents, the percentage of Hispanic residents, and the percentage of people with a Bachelor's degree or higher. Riders' ZIP codes are inferred from their credit card on file, while drivers' ZIP codes come from documents they filled out upon sign up. ZIP codes are not granular enough to inform us about riders' or drivers' individual characteristics, but they can inform us of the area makeup where they are from.

In the end, we were able to obtain the driver's ZIP code for 87.3% of trips in our data and the rider's ZIP code for 82.9% of trips. To explore spatial differences within cities we first discretize the observed demographic values into quintiles across trips. We then compare mean tip amounts across the demographic quintiles. Means within each quintile for the different demographic indicators we consider are in Table 1. As is clear from the top and bottom panels of Table 1, there is significant heterogeneity in race, income, and education across the quintiles for both drivers and riders.

Table 2 complements Table 1 by showing the average tip amounts across the different driver and rider ZIP demographic quintiles. In the raw data, we find that drivers from ZIP codes with a higher percentage of black and Hispanic residents tend to get tipped less. People from higher income and more educated ZIP codes tend to get tipped more, though individuals from the most educated ZIP codes get tipped slightly less than individuals from the third or fourth quintile ZIP codes. Considering the rider side (lower panel of Table 2), we find that riders from ZIP codes with a higher percentage of black and Hispanic residents tend to tip less. Mean tip levels by education level follow an inverted U shape, with the least and most educated ZIP codes associated with lower tip levels. Higher income ZIP codes are associated with higher tips.

Below in our exploration of the interaction of the demand and supply side determinants of tipping we will complete more formal analysis of these data in the spirit of Equation 1 conditioning on various variables that might cause such effects to exist, but the general result will maintain: there is significant temporal and spatial variation in tipping within and across individual U.S. cities.

3.1 Demand-Side Determinants

The aggregate data summarized above helps to shed light on the what's, when's, and where's of tipping, but much underlying variation exists beyond these macro metrics. Upon digging deeper into the demand side correlates, we find a second result:

> *Result 2: Demand side factors such as rider gender, rider rating, and their previous experience with Uber are each important in explaining the variation in tipping*

Rider Gender

To provide support for Result 2 we begin by exploring rider gender. Uber does not collect riders' gender as part of its sign-up process so imputation is necessary. We impute rider gender using rider first names and public birth certificate records from the U.S. Social Security Administration (SSA); see Appendix Section 3 for details. We find that 93.3% of trips have a matched rider name. Riders with uncommon first names who do not match any records in the SSA data set are marked as Unmatched. Names that are not matched tend to be nicknames, false names, or names characteristic of cultures less represented in the SSA data.

Using this imputation method, we summarize differences in tip rates by rider gender in Table 3. Table 3 shows that there are more trips taken by men (52.5%) than by women (40.7%). More importantly for our purposes, in the raw data, there is a substantial difference in tipping behavior by rider gender. Women are significantly less likely to tip than men (14.3% versus 17.0% of trips, respectively); they tip slightly less conditional on tipping ($3.067 versus $3.129); and consequently tip about 9.2 cents (17%) less than men on average ($0.439 versus $0.531).

Since Result 1 teaches us that the when's and where's are important in explaining tipping patterns, the raw data in Table 3 are difficult to interpret. To condition on these factors, we regress the tip amount on gender as well as temporal and spatial features of the trip in the spirit of Equation

1. Regression results controlling for date, hour of week, and location are given in Table 4; estimates are clustered at the rider level.

Column 1 of Table 4 presents estimates of the effect of gender without additional controls. The estimate in this model suggests a $0.09 difference in tips across men and women, or roughly a 17% difference. In Column 2 we control for the date, coded as a factor variable. This barely changes the R-squared of the regression and leads to no change in the estimate for the difference in tip rates, suggesting that there are not large day-to-day fluctuations in tip amounts. In Column 3, we add 168 indicator variables for each hour of the week. This absorbs more variation, but the R-Squared remains small. The estimate for the difference in tip levels by rider gender is again roughly unchanged, at 17%.

In Column 4 of Table 4, we add "level 5" pick-up location geohash[3] measures to control for the date of the trip (not including hour of week controls anymore). Geohashes are a hierarchical spatial data structure that subdivide space into grid-shaped buckets. Adding information for where riders begin trips explains more variation and shrinks the differences in tip rates across rider genders. The gender estimate falls to about 6.8 cents on average. Adding hour of week controls to this model leaves the estimates reported in Column 5 unchanged. In Column 6, we add geohash controls for where the trip ended. This absorbs more variation, with the R-squared increasing to 0.028. The difference in average tip between women and men is 6 cents in this model.[4] As a whole, the data suggest a first demand-side result:

---

[3] For each city, we added indicator variables for the top level 5 geohashes in which trips get requested until more than 90% of the trips in the city have been accounted for. The remaining geohashes are grouped into a city-specific "other" category.

[4] Below we explore the effects of driver and trip characteristics (such as fare, trip distance, and other trip features), but when we additionally control for trip, driver, and other rider characteristics (see Appendix Table 1), we still find that women tip about 6 cents (11%) less than men on average.

Result 2a: Men tip 12%-17% more than women

This result is in accord with the Mturk survey of Lynn (2016), who reports that men are better tippers than women (see also Lynn et al. (1993) and the cites therein). The literature on who tips and who is tipped has argued broadly that demographics such as gender matter, and past work documents differences in the economic preferences and outcomes for men and women (see Croson and Gneezy (2009) for a survey), but there are mixed results with respect to social preference differences across men and women. In our data, the gender results are robust.

Rider Rating

Next, we consider how tipping varies with the rider's rating at the time of the trip. Recall that the rider is rated by the driver with a discrete choice of stars (from 1-5) after each trip is completed. Importantly, however, the rider is not asked by the app for a tipping decision until *after* the driver gives the rider a rating. Hence, any correlation we find between tipping and rider rating cannot be higher tips inducing drivers to give higher ratings. We operationalize our analysis by computing the rider's rating as the average of their past 500 rated trips on the platform, or if they have taken fewer than 500 rated trips, the average across all of their past ratings. Riders with no prior rated trips are coded as having a missing rating for that trip (6.2% of trips in our data have a missing rider rating). Conditional on having a rating, riders have a perfect 5.0 lifetime rating for 24% of trips and less than 4.4 stars on 3.0% of trips. Figure 6 shows the distribution across trips of riders' ratings.

Figure 7 through Figure 9 show the distribution of how often trips are tipped, how much is tipped, and the average tip across trips across the rider rating scale. The figures reveal that ratings are positively associated with higher propensities to tip, higher tip magnitudes when the trip is tipped, and higher overall mean tips. For example, riders with a 5 star rating tip more than twice as often as those with a 4.75 rating, and when they tip they tip nearly 14% more. This leads to the average tip being more than 100% greater (more than \$0.70 per trip versus less than \$0.30).

To test whether the results in the raw data are robust to conditioning variables, we complete the regression analysis in Equation 1 (regressing the various tip outcomes on rider rating, gender, spatial and temporal features of the trip, driver and other trip characteristics such as fare, trip distance, and other trip features). Upon doing so, we find that a one standard deviation (0.24 rating point) increase in the pre-period rider rating is associated with a 1.2 percentage point increase in the probability of a trip being tipped, a \$0.021 increase in the tip when the trip is tipped, and a \$0.039 increase in the mean tip on a given trip. And, in each case the rider rating is significant using standard t-tests at conventional levels ($p < .05$). The regression results are in Appendix Tables 2 through 4. As a whole, the data suggest a second demand-side result:

Result 2b: Rider ratings are positively associated with tipping

This result suggests that the individual features associated with generating a high star rider rating—punctuality, congeniality, niceness—are correlated with providing more in tips. Interestingly, the literature in general has had difficulties finding generosity correlated across domains (see, e.g., Levitt and List (2007)), but this result suggests that there is an important shared component between generating higher rider ratings and tipping. In this way, a person's tendency to act pro-socially may not be entirely situation-dependent.

Rider's Experience with Uber

To examine how consumption exposures influence tipping behavior, we explore whether tipping is associated with the number of trips a rider has completed (including the current trip). The red dots in Figure 10 through Figure 12 show the average tipping frequency, the average tipped amounts, and the average tip overall across the number of trips a passenger has taken with

Uber. Interestingly, we find a large negative correlation between ride exposure and tips. For example, in the raw data, trips in the first quintile for passenger lifetime trip count (i.e. that are one of a rider's first 15 trips) are tipped 23.4% of the time, tippers tip $3.34 on average, and the average trip is tipped at a $0.783 rate. In comparison, trips from the most experienced riders (i.e. a rider's 275th trip or more) are tipped only 8.1% of the time, $2.69 on average conditional on being tipped, and $0.218 on average overall.

To ensure that other variables are not causing this correlation, we follow our approach in Equation 1 and regress the tip amount on trip count, and control for all of the other rider, driver, and trip features. We find that a one standard deviation (260 trips) increase in the number of trips a rider has completed is associated with: i) a 1.60 percentage point decrease in the probability of the trip being tipped, ii) a decrease in the mean conditional tip of $0.089, and iii) the overall mean tip dropping by $0.052. In each case, the number of previous trips is significant at conventional levels ($p < .05$).

There are several interpretations of this surprising insight. For example, this trend could be due to a selection effect, whereby riders who take a large number of trips are simply lower tippers than those riders who take fewer trips. Alternatively, perhaps there is a treatment effect, whereby riders become accustomed to the notion that tipping as a social norm is not as strongly ingrained in the ride-sharing industry as it is for taxis or restaurants. To parse the explanations, we examine a data set of riders who took their first trip within a six-week period from August 7-September 18, 2017 and compare two cohorts. The first cohort, labelled "All riders", includes all riders who took their first trip during this period. By contrast, the "Fixed cohort" includes all riders who not only took their first trip in this period, but also completed exactly 20 trips by its end.

The frequency of tipping, average tip amounts and average tip amounts conditional on tipping for the two cohorts are depicted in Figure 10 through Figure 12. We find that while the effect on the probability of a trip being tipped is more pronounced when considering all riders (Figure 10), there remains a decrease in the probability of a trip being tipped for the cohort that reaches twenty trips. This shows that riders tip less often as they take more trips. However, when examining the tip amount conditional on the trip being tipped, the fixed cohort of riders who eventually complete 20 trips during the period does not change the amount they tip when they do tip (Figure 11), while the set of all riders do decrease the amount tipped, suggesting selection is driving a large amount of the observed effect. The combination of these two trends shown in Figure 12 results in a pronounced decay in mean tip amount with experience both for the population of all riders and for the fixed cohort. The downward trend for the fixed cohort is more muted. Together, these lead to a third demand-side insight:

Result 2c: Riders tip less as they take more Uber trips, with both selection and treatment playing a role

We have been unable to find an antecedent for this result in the tipping literature, which is dominated by studies using responses to hypothetical surveys (see, e.g., Lynn et al., (1993) and the cites therein) showing various interesting correlations but little relevance to Result 2c.

3.2 Supply-Side Determinants

The data examined in 3.1 highlights the role that certain demand-side variables play in the tipping decision to show the who's of tipping, but we purposely ignored the supply side in much of the analysis. When examining data focusing on supply-side variables, we find a new set of insights:

Result 3: Supply-side factors such as driver gender, age, rating, and experience as well as trip features explain substantial variation in tipping

Driver Gender and Age

Unlike the demand-side, Uber collects the driver's gender and age as part of its sign-up process. Concerning gender, Cook, et al. (2018) use this information to decompose the gender pay gap on Uber. They show that women earn about 30 cents less per trip (5% less per hour) than men in Chicago. This wage gap is explained by differences in experience, where and when men and women work, and driving speed. Here, we explore differences in tip outcomes between male and female drivers. Trip statistics across driver genders are reported in Table 5. There are more male drivers than female drivers, and male drivers take more trips on average. In the raw means, we find that women get tipped 5.7 cents (12%) more on average. This difference is driven by a higher percentage of trips tipped—women receive tips nearly 10% more often than men.

We test the robustness of these results via regression analysis, successively adding controls for the pick-up location, date of the trip, hour of the week, and drop-off location. Empirical estimates are presented in Table 6 and are clustered by driver. The results in Table 6 paint a picture similar to the raw data: female drivers receive considerably more tips than men, roughly 4.8 to 5.7 cents more on the average trip. In total, this observed tipping difference yields a roughly 5 cents difference in the gender pay per trip. While we observe a five cent tip gap favoring women, the implied change in the gender wage gap accounting for tips is only a reduction of approximately 13% of the gap reported in Cook, et al. (2018).

Of course, differences in tipping may still be driven by variation in trip distance, fare, or characteristics that differ between men and women, such as age, experience, or rating. To test whether differences in tip levels can be explained by other observable characteristics aside from driver gender we add the full set of controls that we used above when exploring rider gender and tips. Even after account for these factors, the gender tip gap remains statistically significant (see Appendix Table 5). In sum, we report a first supply side result:

Result 3a: Female drivers are tipped 10%-12% more than male drivers

The tipping literature largely is in accord with this result, as Lynn and Simons (2000) and Hornik (1993) report that women earn more in tips compared to men. Likewise, Lynn, et al. (1993) show that server gender is important in restaurants to the tip amount. In a related literature, Landry, et al. (2006) find that female solicitors receive more charitable contributions than male solicitors, with physical attractiveness also related to contributions (see also Hornik, 1993). This provides an interesting link between the motivators of charitable giving and tipping based on receiver gender.

Next, we turn to assessing whether a driver's age is related to their tip amount. Age information is gathered at the time of sign-up for all drivers—they must be 21 years of age or older to be an Uber driver. To show the data patterns, we split drivers into six groups based on their age on September 18, 2017. The youngest group, aged 21 to 26, is denoted group 1, while the oldest group, aged 65 to 90, is denoted group 5. We code drivers who are missing date of birth, are listed as under 21 years of age, or are listed as over 90 years old as "Missing". Our results are not sensitive to the nature of our binning.

Summary statistics of the raw data suggest that drivers who are 65+ years old are tipped more often and consequently more on average. However, the results are not robust to regression specifications that control for where and when the trips happen. Nonetheless, we do find a robust and significant effect of the gender-age interaction in our regression analysis. Namely, the female-favoring gap in tips reported above shrinks with age and disappears by the age of 65.

To derive this result, we regress tip amount on the interaction between driver gender and age, with controls for date, hour of the week, pick-up location, drop-off location, and other trip, driver, and rider characteristics. For ease of interpretation, in Figure 13 we report fitted values for the different driver gender cross age pairs. Results are relative to men aged 21 through 25. The figure shows a stark pattern: older men are tipped about 2 cents less than younger men, while older women are tipped 4.5 cents less on average than younger women. This tipping pattern across ages yields an interesting age and gender interaction result:

> *Result 3b: While younger female drivers receive more tips than comparably-aged male drivers, this disparity shrinks over time and disappears completely by age 65*

This result follows the spirit of the insights in Lynn and Simons (2000) and Hornik (1993), and has implications for work on gender and age discrimination. Below we explore how this general data pattern co-varies with rider gender and show that male riders are the key reason for this tip disparity shrinkage.

Driver Rating

Similar to our analysis on rider rating, we can explore how tipping varies with the driver's rating on the platform at the time of the trip. One of the most prevalent reasons for tipping in the literature is service quality, as studies often argue that an important predictor of the tip is service quality (Lynn and Simons (2000); Hornik (1993); Lynn et al. (1993)). Thus, provided the driver star system, which is a discrete choice of stars (from 1-5) that the rider gives the driver after each trip is completed, captures quality, we expect a positive correlation with tips. We calculate driver ratings by taking an average over the previous trips. Yet, not all of the trips are rated: drivers with no prior rated trips are coded as having a missing rating for that trip. 0.5% of trips in our data have a missing driver rating. Conditional on having a rating, drivers have a perfect 5.0 lifetime rating for 2.9% of trips and less than 4.4 stars on 1.4% of trips. Figure 14 shows the distribution across trips of drivers' lifetime ratings.

Figure 15 through Figure 17 show averages of how often trips are tipped, how much is tipped, and the average tip across the driver rating scale. The figures reveal that ratings are positively associated with higher propensities to receive a tip, higher tip magnitudes when the trip is tipped, and higher overall mean tips. For example, drivers with a 5 star rating, are tipped close to 50% more often as those with a 4.75 rating, and when they do receive tips they are nearly 5% higher. This leads to the average tip being roughly 50% higher (more than $0.60 per trip versus just over $0.40).

To test whether the results in the raw data are robust to conditioning variables, we estimate Equation 1, and find that a one standard deviation (0.42 rating point) increase in a driver's rating is associated with a 1.40 percentage point increase in the probability a trip is tipped, a $0.137 increase in tips for tipped trips, and a $0.046 increase in tips per trip. In sum, the data suggest a third supply-side result:

Result 3c: Driver ratings are positively associated with tipping

This result suggests that the individual features associated with generating a high star driver rating are correlated, or similar to those characteristics that attract tips. The literature (see, e.g,, Zeigler-Hill et al. (2015)) finds that self-reported conscientiousness is positively correlated with performance reviews and tips. In addition, their survey evidence finds that servers had the highest tips when they had high levels of extraversion. If a similar phenomenon is happening in our data, then the extroversion of drivers might be garnering tips while also generating higher ratings. This is speculative, and more research is necessary to draw firm conclusions.

Driver's Experience with Uber

Similar to the data analysis on rider experience with Uber, we examine the number of trips a driver has completed (including the current trip). The red dots in Figure 18 through Figure 20 show averages of how often trips are tipped, how much is tipped, and the overall tip over the number of trips the driver has taken in their lifetime. Interestingly, much like the case for riders, we find a negative correlation between trips completed and tips. For example, in the raw data, trips in the newest quintile (i.e. driver's first through 299th trip) are tipped 16.7% of the time, are tipped $3.14 when tipped, and are tipped $0.526 on average (compared to 13.9% of the time, $3.04 when tipped, and $0.422 on average for the trips from the most experienced drivers, i.e. drivers' 3,827th trip or more). When controlling for other trip features in a regression analysis the differences remain, but do get smaller: a standard deviation (2,871 trips) increase in the number of trips provided is associated with only a 0.20 percentage point decrease in the probability a trip is tipped, a $0.010 decrease in the amount tipped when the trip is tipped, and a $0.005 decrease in tips on average.

We use six weeks' of data to understand the relative impact of the selection and treatment effects. In this context, the selection effect would correspond to the drivers less likely to receive tips working more frequently; the treatment effect would be reflected in drivers losing their tip-earning aptitude as they gain experience, e.g., due to increased indifference about their riders' trip satisfaction. Figure 18 through Figure 20 show the percentage of trips tipped, mean tip amount conditional on the trip being tipped, and mean overall tip for two cohorts: 1) all drivers who started working for Uber during the period, and 2) a fixed cohort of drivers who started working in the period and completed 20 trips during the period. Compared to riders, we do not see as much separation between the two cohorts, with the fixed cohort having a slightly less pronounced decay in the amount their trips are tipped as they gain experience. These data patterns suggest our final supply-side result:

> *Result 3d: Drivers receive less in tips as their number of trips increases, due to a lower likelihood of receiving a tip on any given trip; an effect largely driven by treatment*

While this result is at odds with our ex ante expectations, it is consonant with the notion that the monotony of the work and the realization that tips are not generating extraordinary returns might lead to this perverse result. Indeed, personality and self-reported conscientiousness has been found to relate to tipping (Zeigler-Hill, et al. (2015)), and perhaps these decrease over time as the work becomes monotonous.

3.3 Trip Features

Our data thus far suggest the import of the demand and supply sides, but those two factors are not entirely encompassing. There are other trip features that might impact tipping behavior that relate to both the demand and supply sides. In this section we outline our main conclusions focusing on these other trip variables. A first insight follows:

> *Result 4: Features of the trip, including the rider/driver gender match, the fare level, and quality of the service are correlated with tipping*

Interaction between rider and driver gender

Above we established two gender-related tipping behavioral patterns: i) male riders tip more than female riders and ii) female drivers receive more in tips than male drivers. We now consider the interaction between rider and driver gender. Table 7 shows summary statistics for each interaction. In the raw data, men tip men 8.9 cents more than women tip men. In addition, men tip women about 6.8 cents more than they tip men, and women tip women about 4.8 cents more than they tip men. These differences are driven by both variation in the percentage of trips tipped and the mean tip conditional on tipping. Table 7 shows that unmatched riders tip far less overall and tip women more than men.

As a whole, the raw data show that men tip female drivers nearly 12% more than they tip male drivers, while females tip female drivers roughly 11% more than they tip male drivers. When estimating Equation 1, we find similar results: men and women both tip female drivers more than they tip male drivers (see Appendix Table 6). Given that both male and female riders are responsible for the higher level of tips for female drivers, it is instructive to explore if they are also both causing the interaction effect observed in Result 3b.

Following Equation 1, we regress tip amount on the interaction between driver gender, rider gender, and age, with controls for date, hour of the week, pick-up location, drop-off location, and other trip, driver, and rider characteristics. When considering the fitted values from Equation 1, we find that tips for female drivers are more steeply decreasing with age when the rider is male (see Appendix Figure 1). Indeed, by the time the driver reaches 65 years of age, men tip male and female drivers identically. When the rider is female, tips for female drivers are much less correlated with age. In this way, the effect of driver gender and age on the tip amount depends on the gender of the rider (male riders have different age effects compared to females at $p < .05$), leading to our next result:

> *Result 4a: Both women and men tip female drivers more than male drivers. For male riders this gender gap disappears as the driver ages due to decreased tipping of older female drivers.*

To the extent that age negatively correlates with physical attractiveness, this result appears consonant with studies that find a positive correlation between physical attractiveness and tip amounts (see, e.g., Lynn et al (1993); Lynn and Simons (2000); and Hornik (1993)). Alternatively, the pattern could be driven by differences in perceived need of the driver, and social norms and social pressure play a role in the temporal tip patterns. Our data are not rich enough to parse these alternative interpretations.

Fare Level

Another feature of the trip that the literature suggests potentially impacts the tip level is the level of fare (see Lynn et al. (1993)). In particular, norms for expected tips appear to be closely tied to bill size (Azar 2003). We find that riders are more likely to tip as the fare of the trip increases, but at a decreasing rate. Appendix Figure 2 through Appendix Figure 4 show the relationship between the fare and our main tipping outcomes: the probability of a trip being tipped, the mean tip when the trip is tipped, and the mean tip overall. Overall, each of the three outcome metrics reveals a consistent story with the literature: tips are positively associated with fare level.

Differences in tipping may be driven by variation in time, location, and other factors that may differ across fares. To test whether differences in tip levels can be explained by other observable characteristics aside from fare, we use Equation 1 and regress the amount tipped against fare and the other control variables. The model yields a significant coefficient that suggests a 10% increase in fare is associated with a \$0.013 increase in expected tip, or around a 2.5% increase (See Appendix Tables 2 through 4). This leads to our next result:

> *Result 4b: The level of tip and the fare are positively associated in a concave manner. On average, a 10% increase in fare is associated with a 2.5% increase in tip.*

The novel aspect of this result is the concave nature of the relationship. While intuitive, we are not aware of robust results that show tips increase at a decreasing rate in the amount of the bill.

Trip Quality

We suspect that most of our readership has experienced rideshare, either through Lyft or Uber. If you think back about your last trip, you will note that several distinct features arose. For our purposes, important features include how long Uber estimated it would take for your ride to arrive and to reach your destination, and whether that promise was met. Likewise, the experience on the trip might affect the level of tip. In terms of longer than expected trips, there is related evidence that Uber customers react negatively to trips that take longer than expected (Halperin, et al. 2018). In this section, we consider two measures of delays: time from request to pick-up and time from pick-up to drop-off. The delay in picking up passengers ranges roughly from being 2.7 minutes early in the first percentile to 7.8 minutes late in the 99th percentile, while the delay in time from pick-up to drop-off varies more, from 8.2 minutes early to 22.6 minutes late.

In terms of the on-trip experience, Uber collects telematics information from drivers' phones during trips to better understand car movement. These data include estimates for the number of "hard accelerations," "hard brakes," and speeding during the trip.[5] In addition, Uber logs the vehicle model year of the car used during each trip. Finally, drivers can change the default language in their apps from English—which approximately 5% of drivers choose to do. This presents a rough measure of potential interaction during the trip.

To explore how these trip features impact the tipping decision, we use Equation 1 to examine how the tip amount is affected by these variables. We find that each of these variables has an impact on the tipping decision (see Appendix Tables 2 through 4). First, a one standard deviation increase in the actual time to pick-up minus the expected time to pick-up leads to a \$0.007 decrease in tips. On the other hand, a one standard deviation increase in the delay from pick-up to drop-off is correlated with a \$0.036 increase in tips. That is, while being late to pick up a rider has a small negative effect on average tip, longer than expected rides have the opposite (and larger) effect. It could be that drivers get tipped more for longer than expected trips because

[5] Hard brakes are defined as acceleration events that are less than -11 (Km/h/s), while hard accelerations are accelerations that exceed 11 (Km/h/s).

they request stops along the way (keep the driver waiting during the pick-up, or make other requests that lead to a detour that are not observed directly by Uber).

Second, there is a small negative effect of hard acceleration or braking events. For hard accelerations, there is a statistically significant 10 basis point decrease in the probability a trip is tipped, a statistically significant $0.018 decrease in the tip amount for tipped trips, and a statistically significant $0.007 decrease in the average tip overall. For hard braking, the effect is even larger: a 10 basis point decrease in the probability of being tipped, a $0.049 decrease in the tip amount when tipped, and a $0.013 decrease in the average overall tip. Finally, trips with speeding episodes are 50 basis points less likely to be tipped, are tipped $0.044 less when they are tipped, and have an average tip $0.029 lower than other trips.

Third, concerning vehicle age, we find that trips in cars from before 2009 are tipped $0.013 less compared to trips in newer cars. Fourth, drivers who chose to change their app language from English get tipped nearly 30% less than those drivers who do not change their app language. In sum, these insights lead to our next result:

> *Result 4c: The size of tip is correlated with the quality of the trip. This is reflected in both the probability of a trip being tipped and in the tip size.*

There is a long literature that this result conforms to (see citations above and the work of Michael Lynn, Ofer Azar, and colleagues more generally), and we view these insights as simply confirming that service quality and tip are inextricably linked.

Finally, another trip feature is when and where the trip occurs. Above, we showed that the where's and when's are important in our within and between city analysis. Following Equation 1, we explore the when's and where's controlling for our set of conditioning variables. The estimation shows that Result 1 continues to hold: tips are spatially correlated to where both the rider and driver are from (ZIP code level), and the time of day is important as well. Appendix

Figure 5 and Appendix Figure 6 provide a flavor of the fitted tip amounts in a few representative model runs.

3.4 Explaining Overall Tipping Variation

In an effort to provide an integration of our results, we decompose formally the tip variation across rider, driver, location, and time for several cities in the data. This exercise helps to determine the importance of each input in explaining tip outcomes. Following the procedure used in Athey, et al. (2019), we regress:

$$y_{ijkt} = \alpha + \mu_i + \nu_j + \gamma_k + \varepsilon_{ijkt} \quad (2)$$

This regression decomposes the tip outcome y_{ijkt} on a given trip between driver effects μ_i, rider effects υ_j, (location cross time) effects γ_k, and a residual term ε_{ijkt}. We compare the standard deviation of each estimated effect type to assess how much variation is explained by each component. We define time by location (time cross location) pairs as level 5 geohashes crossed with the hour of the week. Since effects estimated on few trips will have higher variance, we restrict the sample only to include riders, drivers, and time by location (time cross location) pairs that have at least 10 UberX trips between August 18 and September 14, 2017.

To aid in interpretation, we restrict the analysis to six cities, selected to be roughly reflective of different city types in the United States: a college town (Bloomington, IN), large cities on the East coast, West coast, and Midwest (Boston, San Francisco, and Chicago), and moderately sized cities in the South and Mountain West region (Asheville, NC and Salt Lake City). Our general results are not changed if we examine other cities. Table 8 shows the summary statistics within the data set for each city.

Table 9 presents the standard deviation of estimated effects from Equation 2 for the different sources of variation across trips in the data set. We exclude effects that are below the 2nd percentile and above the 98th percentile for a given effect type to ensure the estimated standard deviations are not driven by outliers. Results with tail values included reveal similar insights and are available in Appendix 4.1.

Interestingly, empirical results in Table 9 show that across all cities, the amount of variation due to rider effects is considerably larger than the magnitude of other sources of variation. Indeed, rider effect sizes are comparable to residual effect sizes, and about three times larger than driver effects across each city. Time by location effects are roughly half as important as driver effects in large cities, but comparably important to driver effects in smaller cities. To provide a visual depiction of the results, Figure 21 shows density plots for the different effects in each city. Effects have been demeaned. While the densities of the driver effects and time by location effects roughly peak around 0, the rider effect densities peak below 0 in each city and have long right tails. This means that in every city some riders tip significantly more than the median rider. In contrast, there are no groups of drivers or time by location blocks that receive disproportionately higher tips than everyone else.

To ensure that larger variance of rider effects is not simply a result of fewer trips per rider, we also explore the robustness of these results controlling for number of trips. The results, reproduced in Appendix 4.2, are qualitatively similar. In addition, we performed a similar analysis using driver ratings instead of tips as the outcome metric. We find that rider effects remain considerably more important than other factors, but their relative size decreases. The data point to our next result:

> *Result 5: Demand-side variables explain roughly three times more of the observed tipping variation than the supply side or features of the trip explain.*

We view this result as new to the literature and provides a new lens into tipping: it is more about who is giving rather than who is receiving or the quality of the service. In this manner, tipping outcomes reflect the personal characteristics of the consumer much more than those of the provider or even the service quality. An interesting parallel to this finding is its link to the economics of charity literature, where the individual donor characteristics are found to be much more important than the features of the recipient charitable organization (see List (2011)).

This exercise allows us to identify the social-preference types of different people, as classified by their tipping behavior. The fact that people are different in their prosocial behavior is well established in laboratory experiments, in which participants are classified into types based, for example, on their distributive preferences (e.g., Andreoni and Miller (2002); Blanco et al. (2011); Iriberri and Rey-Biel (2013)). Our data show that these heterogeneities extend beyond the lab, and our variance decomposition provides clear evidence from the field that there is substantial heterogeneity in social preference types in the domain of tipping.

4. Repeat Interaction and the Effect of Defaults

The variance decomposition exercise provides a crisp look into the relative importance of the various factors driving tipping behaviors, but a complementary deeper step can be taken to interpret the data patterns within the extant social preference literature. In this section, we combine experimental variation with non-experimental variation to explore the underlying motivations for why people tip and how social norms affect tipping. We use two distinct approaches to this problem. First, we examine individuals' behavior over repeated interactions. Second, to understand the underlying mechanism at work we use a complementary field experiment that examines the impact of defaults.

4.1 Repeat Interactions

One fact around the typical ride share experience is that it is uncommon for riders to interact with the same driver more than once in our time period. Of all the rider/driver matches that occurred on UberX trips between August 5 and August 13, 2017, only 1% appear again in the data at some point between August 13 to September 17, 2017. While matching again with a driver over a short time period is rare, there are enough trips in the data set that we can observe a substantial number of such occurrences. We again use the full data set of UberX trips from August 18, 2017 to September 15, 2017 to explore repeat interactions.

Table 10 shows a summary of the raw data on tip outcomes based on the number of times the rider and driver have matched on a trip. Interestingly, tips increase with the number of times the rider and driver have matched with each other: the mean tip on a trip increases from \$0.478 to \$0.563 to \$0.643 from the first to the third interaction, an increase of more than 35%.

This effect is intuitive, but could be driven completely by selection: drivers and riders who match with each other more than once could be systematically different from those who match only once. To account for selection effects, we construct fixed cohorts of (rider cross driver) pairs. Cohort 1 only includes (rider by driver) pairs that matched together exactly once in the data set. Similarly, Cohorts 2 and 3 only include (rider by driver) pairs that matched together exactly two and three times, respectively.

Using Equation 1 as our specification, we find that for both Cohort 2 and Cohort 3, average tips increase with the number of times the rider has matched with the same driver.[6] Cohort 2 tips 27% more on average during the second interaction compared to the first interaction. Cohort 3 tips

[6] We do not include rider- and driver-level controls because there is no variation in these variables within cohorts. We also do not include controls for when and where the trip happens because there are fewer observations available to estimate the fixed effects within cohorts.

7% more during the second interaction and 23% more during the third interaction compared to the first interaction (see Appendix Figure 7). These results complement the raw data insights, and provide further evidence that repeat interaction breeds more rider generosity via higher tips.

There are competing explanations that potentially explain the fact that riders tip more when they interact a second or third time with the same driver. For example, when riders match again with the same driver, they might update their views of the likelihood of meeting the same driver more than once on Uber—a form of strategic reciprocity, shown to be of great import in certain markets (see, e.g., List (2006); Al-Ubaydli, et al. (2010)). Under this model, the increased chance of future interactions compels riders to tip more on all future trips. We refer to this explanation as Model 1. If Model 1 is correct, then a stark prediction is that after updating their priors that future interaction is more likely, riders should tip *all* subsequent trips higher due to strategic reciprocity.

If riders already have a realistic perception of the likelihood of repeat interaction, then there must be an alternative reason for tipping higher in future interactions. A psychological alternative to strategic reciprocity is that repeat interaction builds a greater social connection between the rider and driver, which compels the rider to tip more, or perhaps induces greater guilt for not tipping (in the spirit of the literature on social connection and reciprocity, e.g. Chen and Li (2009)). We refer to this explanation as Model 2. If Model 2 is correct, then only those trips with the same previous driver will receive higher tips; tips on other trips will remain unaffected or decrease (given the previous results on riders "learning" to tip less (see Result 2c)).

To evaluate these models, we consider all riders who match exactly twice with the same driver. We exclude riders who match more than once with multiple different drivers. Let y_{rt} be the tip outcome on trip t for rider r, where t is ordered from the rider's first trip to their last trip in the sample. Let $d(r)$ be the driver that rider r matches twice with, $T_1(r)$ be the trip index for the first time the rider matches with $d(r)$, and $T_2(r)$ be the trip index for the second time the rider matches with $d(r)$. We separate the rider's trips into three distinct types: "Period 1" trips are all trips for rider r that occur prior to $T_1(r)$, "Period 2" trips are all trips that occur after $T_1(r)$ and before $T_2(r)$, and "Period 3" trips are all trips that occur after $T_2(r)$. If Model 1 is accurate, then riders should tip more for Period 3 trips than for Period 2 or Period 1 trips. Otherwise Model 2 offers a more reasonable explanation.

Table 11 provides summary regression estimates that compare tip levels across the three periods. Column 1 includes no controls whereas Column 2 includes trip controls. Both specifications show consistent empirical evidence: riders tip less in Period 3 than in Period 1, providing a refutation of the strategic reciprocity motive (Model 1). Since only those trips with the same previous drivers receive higher tips, the data are more consistent with the notion that social interaction leads to greater future generosity.

To dig a level deeper into what might induce such social interaction effects, we explore aspects of the nature of the verbal conversation. Of course, we cannot observe the interaction directly, but we can study cases where the rider or driver uses a default app language other than English. We assume that when either the rider or driver uses an app language other than English the pair becomes less likely to engage in verbal conversation. If social connection between the rider and driver leads to higher tip levels the second time they match, and if conversation breeds social connection, then we might expect the boost in tip levels to be smaller if the rider or driver is a non-native English speaker.

When estimating Equation 1, we find that in cases where only one of the rider or driver uses a default app language other than English there remains substantial increases in tip levels in 2^{nd} and 3^{rd} rider driver interactions (results are in Appendix Table 7). Indeed, the results are similar to the overall data patterns that we observe. This leads us to conclude that if the assumption holds that conversation is less likely when the rider or driver prefers a language other than English, then verbal communication is not the dominant mechanism through which greater social interaction leads to higher tips. This leads to our next result:

> *Result 6: Repeat rider/driver interactions increase tip levels, but the mechanism is not due to strategic reciprocity or conversation-based social interaction explanations; greater exposure itself seems to induce higher tips.*

Previous work highlights that social connections form easily and at times even arbitrarily between individuals (Billig and Tajfel (1973); Goette, et al. (2006); Chen and Li (2009)). In Billig and Tajfel (1973), participants were randomly allocated to distinct, arbitrary groups based on coin flips. There was no interaction between group members, and participants did not even know who else was in their group, yet they still rewarded individuals within their groups more than others. In an extension of this setting, Chen and Li (2009) show that individuals are more likely to reciprocate positively to in-group members and are more forgiving of bad behavior by them. Consistent with this literature, our results suggest that a level of social connection can form simply through a rider and driver interacting with each other more than once and regardless of the level of conversation. This connection can result in greater generosity. Our finding that repeated interaction increases generosity also aligns with experimental results from Wilder and Thompson (1980), who show that repeated interaction improves intergroup connection among female college students more than a single interaction does.

4.2 Effects of Defaults

While the above analysis teaches us some of the underpinnings of tipping behaviors by examining individuals' behavior over repeated interactions, we can shed light on social norms by using a complementary field experiment that examines the impact of defaults on tipping. Given that tipping was a new institution on Uber during our experimental roll out, these defaults may provide signals of what the social norms are for tipping on the Uber app. In this manner, the results from this field experiment help to determine what effect information and social cues can have on levels of tipping when the act is done privately.

In this section, we analyze experimental data (N=12,040,801 participants) from treatments completed during the August 18th, 2017 through September 14th, 2017 period. In the first experiment, riders were shown the same randomized preset for all trips during the observation window. The riders were randomized into one of the following eight preset options for all trips:

- $1 | $3 | $5
- $1 | $3 | $6
- $1 | $4 | $5
- $1 | $4 | $6
- $2 | $3 | $5
- $2 | $3 | $6
- $2 | $4 | $5
- $2 | $4 | $6

In the second experiment, riders saw the same presets as in the first experiment for trips under $20, but a different preset for trips over $20. These riders were randomized into one of the above eight options for their less expensive trips, and into one of these eight options for their more expensive trips:

- $3 | $5 | $8
- $3 | $5 | $10
- $3 | $6 | $8
- $3 | $6 | $10
- $4 | $5 | $8
- $4 | $5 | $10
- $4 | $6 | $8
- $4 | $6 | $10

Results from the second experiment are included in Appendix Section 5 as the patterns match those from the first experiment, which we discuss now.

Figure 22 through Figure 24 show averages of how often trips are tipped, how much is tipped when a tip is given, and the average tip across the default levels (we include errors bars, clustered by rider). As can be seen from the Figure 22, as the values in the presets increase, the proportion of trips that are tipped decreases. 16.2% of trips are tipped when riders have the lowest preset, $1 | $3 | $5, while trips under the highest preset, $2 | $4 | $6, are only tipped 14.9% of the time. As can be seen from the Figure 22, there is a notable drop when moving from presets starting with $1 to those starting with $2.

Conversely, as can be seen from Figure 23, the amount riders tip conditional on tipping increases as the preset values increase. For example, conditional on tipping, trips in which the rider has the lowest preset, [$1, $3, $5], are tipped $2.89 on average, whereas for the highest preset, [$2, $4, $6] the average tip is $3.23. We see again that the largest effect is when going from $1 to $2 in the first position, which increases the mean tip $0.25 (8.7%). Going from $3 to $4 in the middle position increases the mean tip by $0.03 (1.2%), and moving from $5 to $6 in the third position increases the average tip by $0.04 (1.3%).

Consequently, as can be seen in Figure 24, the average amount tipped (including $0 when there was no tip) is only marginally affected by presets, with a slight increase in the average tip as the values in the presets increases. The lowest preset generates an average tip of $0.467 per trip while the highest preset yields a $0.479 tip on average. Moving from $1 to $2 increases the average tip by only 0.7¢ (1.4%), and moving from $3 to $4 in the second position does not significantly increase the average tip at conventional levels. An increase in the third position leads to a 0.5¢ (1.1%) increase in the average tip on a trip.

Our experiment was designed so that we could understand the marginal impact of increasing the preset value in a given position on the page, e.g. the effect of the first digit in going from [$1, $3, $5] to [$2, $3, $5]. We run a regression of the following form:

$$outcome = 1_{\{first\ option\ was\ \$2\ instead\ of\ \$1\}} + 1_{\{second\ option\ was\ \$4\ instead\ of\ \$3\}} + 1_{\{third\ option\ was\ \$6\ instead\ of\ \$5\}}$$

where the different outcomes are the probability of the trip being tipped, the expected tip conditional on tipping, and the expected tip overall.

From this specification, we find that, on average, an increase in the first position from $1 to $2 decreases the probability of a trip being tipped by 1.07 percentage points (6.6%). A $1 increase in the second position decreased the probability of the trip being tipped by 16 basis points (1.0%), and an increase of $1 in the third position effectively did not change the probability of being tipped (results are in Appendix Table 8).

Overall, our results support the notion that defaults impact tipping levels. In particular, higher default options tend to lead to higher tip levels on average, but they also lead to a higher percentage of trips that do not get tipped at all. This result is consonant with the tipping literature that argues social norms drive tipping (see e.g., Conlin et al. (2003)). Yet, compared to previous work in the taxi cab industry, a stark difference surfaces. While our results suggest only a modest influence of defaults on tips—average tips as a percent of the fare increases by 2.5% between the lowest and highest default options—Haggag and Paci (2014) find that a similar increase in defaults lead to a greater than 10% increase in tips. A major difference between our study and Haggag and Paci (2014) is that tipping in their case is in the presence of the driver. A final insight follows:

> *Result 7: Defaults affect tip levels, but are much less powerful than comparable estimates from the literature exploring tips in the taxi cab industry.*

Our preferred interpretation of this result is that when the tipping decision is made privately defaults play a lesser role in the tip decision. Under this interpretation, a general insight is that norms and the monitoring of behavior are *complements*. As such, the strength of social nudges, or any particular course of action brought about through social norms, is diminished without public verification of an action. Alternatively, the effects of social nudges are enhanced as the veil of choice anonymity is removed. While many previous studies show that anonymity reduces generosity (Aplizar, et al. (2008); List (2006); List, et al. (2004); Andreoni and Petri (2004)), our findings are distinct from much of the literature in showing that anonymity reduces the influence of defaults. Future research should explore whether this insight extends broadly to other nudges.

5. Conclusions

In the past five years there is perhaps no consumer innovation that has had a greater impact on people's lives than ride-share. Certainly, in the transportation sector this is true. Indeed, jobs and hourly data highlight the effect on employment, as millions of drivers now serve the market, and debit and credit card data suggest that by 2018 as many as 43% of American adults tried ride-share at least once (Molla 2018).[7] We suspect that given the demographic make-up of our readership, most readers of this study have experimented with ride-share as a consumer.

Think back about your last Uber trip. The ride-share experience itself holds a number of interesting economic issues for the consumer. First is the decision of which app to open—many markets have multiple rideshare options to choose from. Second, setting a destination and receiving estimates of when you will be picked up, dropped off, and the exact price you will pay. Third, making the purchase decision. Maybe you should walk or take the train instead. Finally, if you decide to take the ride-share, after your trip you make an important decision far removed in space and time from the driver: whether, and how much to tip.

There are several interesting economic questions one can explore in this chain of events, and researchers have begun to examine the foundations of ride-share driver pay (see Cook, et al. (2018)), driving quality (Athey, et al., 2019), aspects of two sided markets (Castillo, et al., 2018),

[7] This estimate is from Second Measure, which "measures ride-share spending across a consistent panel of credit and debit card users who made at least one ride-hailing transaction, so user base skews more urban and affluent than the overall population."

and certain consumer behaviors (Halperin, et al. 2018). Each of these studies in their own way highlights details specific to the ride-share market and broader implications into the gig economy or behavioral features more generally.

In this study, we focus on an important economic and social component of the trip: tipping behavior. By doing so, this study sheds new light on the economics of tipping-related social preferences. We extend the investigation beyond the lab and experimentally study social preferences in the field across millions of observations and individuals. The data allow us to study demographic and environmental factors believed to affect generosity.

Our findings complement previous work on social preferences as they not only test these factors, but also deepen our understanding of their interactions, and clarify their boundaries. Specifically, our results indicate that gender matters when it comes to tipping. Women are tipped more than men, and men tip more than women do. Gender also interacts with age, with men tipping younger women more than they tip any other group. We also show that tipping varies based on the demographic characteristics of where the rider and driver are from. As expected, we find that the quality of the ride matters as well, with higher quality generating higher tips.

Our experiment also allows us to further explore the effect of defaults on tipping, as riders are exposed to different presets of tip options. Contrary to extant literature, we find only a modest effect of defaults on tipping behavior. Another important finding is the heterogeneity in people's behavior. Variation in tipping outcomes across riders is three times more important than variation across drivers, with some riders persistently tipping much more than others. Among riders with at least 10 trips in our data, we find that nearly 60% of them never tip, 1% always tip, and only 40% are affected by some aspects of their experience and sometimes tip, but not always. Our results also speak to the social preferences and norms that influence tipping. When there is repeated interaction between the rider and the driver, tipping is higher.

As we discussed above, while tipping is important in markets, our knowledge about such behavior is limited due to the quality of data in the literature. Tipping research tends to rely on small sample sizes, non-random samples, or self-reported tip outcomes. For instance, Hornik (1993) compares tip outcomes across only four waiters and four waitresses. Due to the disparities across these studies, researchers learn of piecemeal findings that are difficult to reconcile even in the best meta-analyses. Our nationwide tipping experiment provides an apples-to-apples comparison of demographic factors believed to affect tipping and social preferences but that have been difficult to test side by side across many observations. We hope that our findings will generate future research that will advance the understanding of tipping through models and testable predictions.

References

Akerlof, George A. 1982. "Labor Contracts as Partial Gift Exchange." *The Quarterly Journal of Economics* 543-569.

Akerlof, George A., and Janet L. Yellen. 1988. "Fairness and Unemployment." *The American Economic Review* 44-49.

Akerlof, George A., and Janet L. Yellen. 1990. "The Fair Wage-Effort Hypothesis and Unemployment." *The Quarterly Journal of Economics* 255-283.

Al-Ubaydli, Omar, Uri Gneezy, Min Sok Lee, and John A. List. 2010. "Toward an Understanding of the Relative Strenghts of Positive and Negative Reciprocity." *Judgment and Decision Making* 524-539.

Andreoni, James. 1998. "Toward a Theory of Charitable Fund-Raising." *Journal of Political Economy* 1186-1213.

Andreoni, James, and John Miller. 2002. "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism." *Econometrica* 737-753.

Andreoni, James, and Ragan Petrie. 2004. "Public goods experiments without confidentiality: a glimpse into fund-raising." *Journal of Public Economics* 1605-1623.

Andreoni, James, Justin M. Rao, and Hannah Trachtman. 2017. "Avoiding The Ask: A Field Experiment on Altruism, Empathy, and Charitable Giving." *Journal of Political Economy* 625-653.

Aplizar Rodriguez, Francisco, Fredrik Carlsson, and Olof Johansson-Stenman. 2008. "Anonymity, reciprocity, and conformity: Evidence from voluntary contributions to a national park in Costa Rica." *Journal of Public Economics* 1047-1060.

Athey, Susan, Juan Camilo Castillo, and Bharat Chandar. 2019. "Service Quality in the Gig Economy: Empirical Evidence about Driving Quality at Uber." *Work in Progress*.

Azar, Ofer H. 2003. "The Implications of Tipping for Economics and Management." *International Journal of Social Economics* 1084-1094.

Azar, Ofer H. 2004. "The History of Tipping - From Sixteenth-Century England to United States in the 1910s." *Journal of Behavioral and Experimental Economics* 745-764.

Benabou, Roland, and Jean Tirole. 2006. "Incentives and Prosocial Behavior." *American Economic Review* 1652-1678.

Billig, Michael, and Henri Tajfel. 1973. "Social categorization and familiarity in intergroup behavior." *European Journal of Social Psychology*.

Blanco, Mariana, Dirk Engelmann, and Hans Theo Normann. 2011. "A Within Subject Analysis of Other-Regarding Preferences." *Games and Economic Behavior* 321-338.

Bolton, Gary E., and Axel Ockenfels. 2000. "ERC: A Theory of Equity, Reciprocity, and Competition ." *The American Economic Review* 166-193.

Castillo, Juan Camilo, Daniel T. Knoepfle, and E. Glen Weyl. 2018. "Surge Pricing Solves the Wild Goose Chase." *Working Paper*.

Chandar, Bharat K., Ali Hortaçsu, John A. List, Ian Muir, and Jeffrey M. Wooldridge. 2019. “Design and Analysis of Cluster Randomized Field Experiments in Panel Data Settings,” NBER working paper.

Charness, Gary. 2004. "Attribution and Reciprocity in an Experimental Labor Market." *Journal of Labor Economics* 665-688.

Charness, Gary, and Matthew Rabin. 2002. "Understanding Social Preferences with Simple Tests." *The Quarterly Journal of Economics* 817-869.

Charness, Gary, and Peter Kuhn. 2011. "Lab Labor: What Can Labor Economists Learn from the Lab?" *Handbook of Labor Economics* 229-330.

Chen, Yan, and Sherry Xin Li. 2009. "Group Identity and Social Preferences." *American Economic Review* 431-457.

Choi, James J., David Laibson, Brigitte C. Madrian, and Andrew Metrick. 2003. "Optimal Defaults." *American Economic Review Papers and Proceedings.*

Conlin, Michael, Michael Lynn, and Ted O'Donoghue. 2003. "The Norm of Restaurant Tipping." *Journal of Economic Behavior and Organization* 297-321.

Cook, Cody, Rebecca Diamond, Jonathan Hall, John A. List, and Paul Oyer. 2018. “The Gender Earnings Gap in the Gig Economy: Evidence from over a Million Rideshare Drivers.” *No. w24732. National Bureau of Economic Research.*

Cox, James. 2004. "How to identify trust and reciprocity." *Games and Economic Behavior* (Games) 260-281.

Croson, Rachel, and Uri Gneezy. 2009. "Gender Differences in Preferences." *Journal of Economic Literature* 448-474.

Dufwenberg, Martin, and Georg Kirchsteiger. 2004. "A theory of sequential reciprocity." *Games and Economic Behavior* 268-298.

Engel, Christoph. 2011. "Dictator games: a meta study." *Experimental Economics* 583-610.

Fehr, Ernst, and Klaus M. Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *The Quarterly Journal of Economics* 817-868.

Fehr, Ernst, and Simon Gächter. 2000. "Fairness and Retaliation: The Economics of Reciprocity." *The Journal of Economic Perspectives* 159-181.

Fehr, Ernst, and Simon Gächter. 2002. "Altruistic Punishment in Humans." *Nature* 137-140.

Fehr, Ernst, Georg Kirchsteiger, and Arno Riedl. 1993. "Does Fairness Prevent Market Clearing? An Experimental Investigation." *The Quarterly Journal of Economics* 437-459.

Fehr, Ernst, Simon Gächter, and Georg Kirchsteiger. 1997. "Reciprocity as a Contract Enforcement Device: Experimental Evidence." *Econometrica* 833-860.

Gneezy, Uri, and John A. List. 2006. "Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments." *Econometrica* 1365-1384.

Goette, Lorenz, David Huffman, and Stephen Meier. 2006. "The Impact of Group Membership on Cooperation and Norm Enforcement: Evidence Using Random Assigment to Real Social Groups." *American Economic Review* 212-216.

Guth, Werner, Rolf Schmittberger, and Bernd Schwarze. 1982. "An Experimental Analysis of Ultimatum Bargaining." *Journal of Economic Behavior & Organization* 367-388.

Haggag, Kareem, and Giovanni Paci. 2014. "Default Tips." *American Economic Journal: Applied Economics* 1-19.

Halperin, Basil, Benjamin Ho, John A. List, and Ian Muir. 2018. "Towards an Understanding of the Economics of Apologies: Evidence from a Large-Scale Natural Field Experiment." *Natural Field Experiments.*

Hemenway, David. 1993. *Prices and Choices: Microeconomic Vignettes.* Lanham, MD: University Press of America.

Hornik, Jacob. 1993. "Tactile Stimulation and Consumer Response." *Journal of Consumer Research* 449-458.

Iriberri, Nagore, and Pedro Rey-Biel. 2013. "Elicited Beliefs and Social Information in Modified Dictator Games: What Do Dictators Believe Other Dictators Do?" *Quantitative Economics* 515-547.

Johnson, Eric J., and Daniel Goldstein. 2003. "Do Defaults Save Lives?" *Science* 1338-1339.

Joyce, Berg, John Dickhaut, and Kevin McCabe. 1995. "Trust, Reciprocity, and Social History." *Games and Economic Behavior* 122-142.

Juni, Samuel, Robert Brannon, and Michelle M. Roth. 1988. "Sexual and Racial Discrimination in Service-Seeking Interactions: A Field Study in Fast Food and Commercial Establishments." *Psychological Reports.*

Landry, Craig, Andreas Lange, John A. List, Michael Price, and Nicholas Rupp. 2006. "Towards an Understanding of the Economics of Charity: Evidence from a Field Experiment." *Quarterly Journal of Economics* 747-782.

Levitt, Steven D., and John A. List. 2007. "What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?" *Journal of Economic Perspectives* 153-174.

List, John A. 2006. "The Behavioralist Meets the Market: Measuring Social Preferences and Reputation Effects in Actual Transactions." *Journal of Political Economy* 1-37.

List, John A. 2007. "On the Interpretation of Giving in Dictator Games." *Journal of Political Economy* 482-493.

List, John A. 2011. "The Market for Charitable Giving." *Journal of Economic Perspectives* 157-180.

List, John A., Robert P. Berrens, Alok K. Bohara, and Joe Kerkvliet. 2004. "Examining the Role of Social Isolation on Stated Preferences." *American Economic Review* 741-752.

Lynn, Michael. 2016. "Why Are We More Likely to Tip Some Service Occupations than Others? Theory, Evidence, and Implications." *Journal of Economic Psychology* 134-150.

Lynn, Michael, and Tony Simons. 2000. "Predictors of Male and Female Servers' Average Tip Earnings." *Journal of Applied Social Psychology* 241-252.

Lynn, Michael, George M. Zinkhan, and Judy Harris. 1993. "Consumer Tipping: A Cross-Country Study." *Journal of Consumer Research* 478-488.

Molla, Rani. 2018. *Vox Recode.* 06 24. Accessed 06 05, 2019. https://www.vox.com/2018/6/24/17493338/ride-sharing-services-uber-lyft-how-many-people-use.

Rabin, Matthew. 1993. "Incorporating Fairness into Game Theory and Economics." *The American Economic Review* 1281-1302.

Schein, John E., Edward F. Jablonski, and Barbara R. Wohlfahrt. 1984. *The Art of Tipping: Customs & Controversies.* Wausau, WI: Tippers International.

Segrave, Kerry. 1998. *Tipping: An American Social History of Gratuities.* Jefferson, NC: McFarland & Company.

Shierholz, Heidi, David Cooper, Julia Wolfe, and Ben Zipperer. 2017. *Employers would pocket $5.8 billion of workers' tips under Trump administration's proposed 'tip stealing' rule.* Washington, DC: Economic Policy Institute.

Sobel, Joel. 2005. "Interdependent Preferences and Reciprocity." *Journal of Economic Literature* 392-436.

Thaler, Richard H., and Shlomo Benartzi. 2004. "Save More Tomorrow™: Using Behavioral Economics to Increase Employee Saving." *Journal of Political Economy* 164-187.

Wilder, David A., and John E. Thompson. 1980. "Intergroup contact with independent manipulations on in-group and out-group interaction." *Journal of Personality and Social Psychology* 589-603.

Zeigler-Hill, Virgil, Avi Besser, Jennifer Vrabel, and Amy Noser. 2015. "Would you like fries with that? The role of servers' personality traits and job performance in the tipping behavior of consumers." *Journal of Research in Personality.*

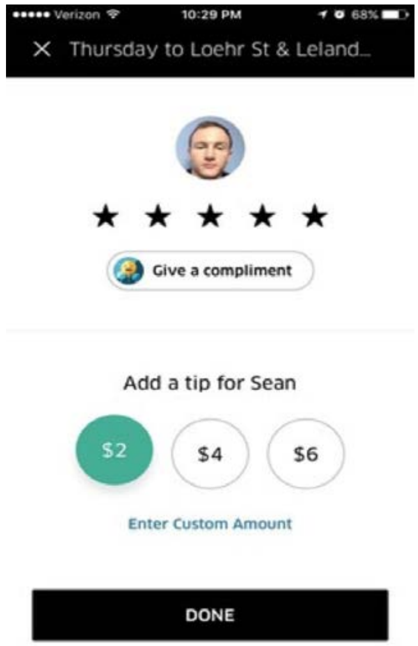


Figure 1: The figure shows an example of the screen riders are presented in the app upon completing a trip. Riders are only given the option to rate and tip after the trip is over and the driver has already rated them. At the time of our experiment, riders could choose from one of three default tip options, enter a custom amount, or enter no tip at all. In the above example, the default tip options shown are 2, 4, and 6.

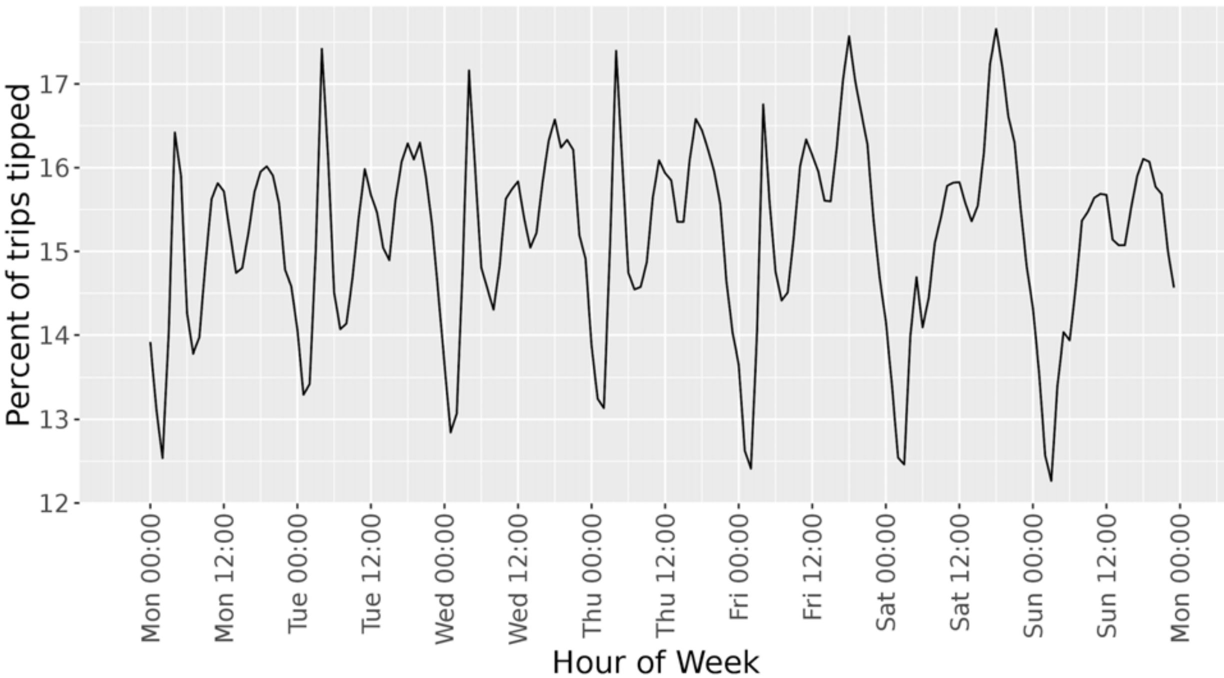


Figure 2: The figure shows the percent of trips tipped by hour of the week across the United States.

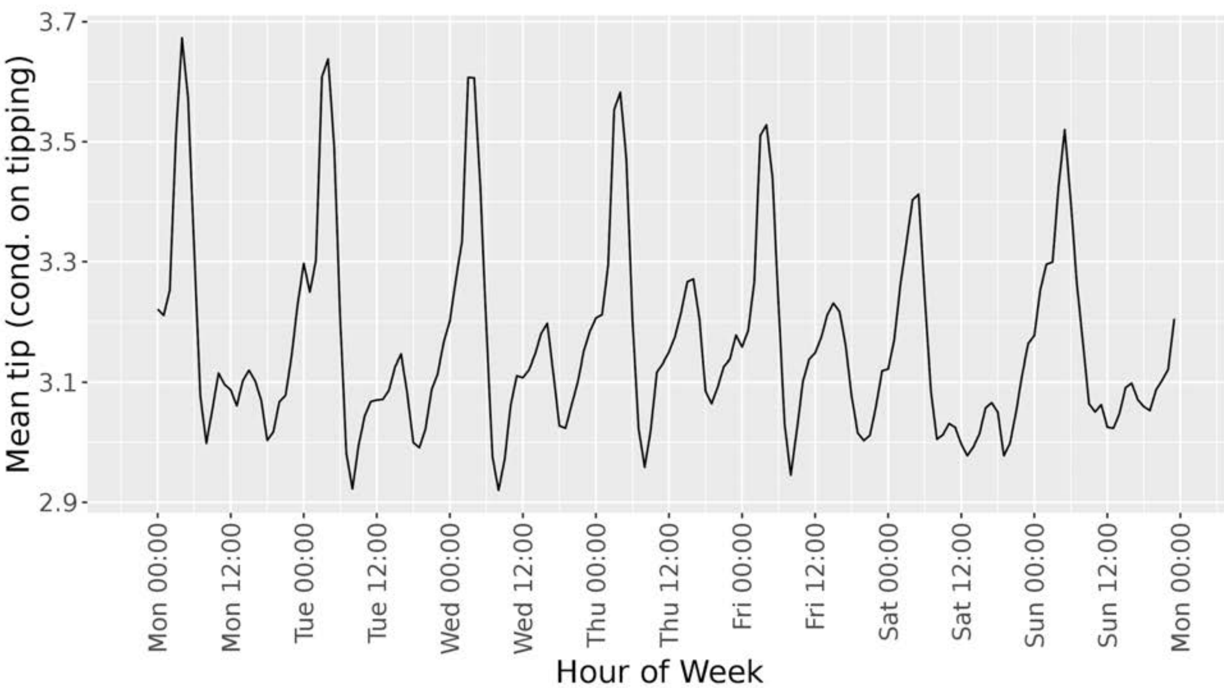


Figure 3: Average tip by hour of week across the United States, including only trips that were tipped.

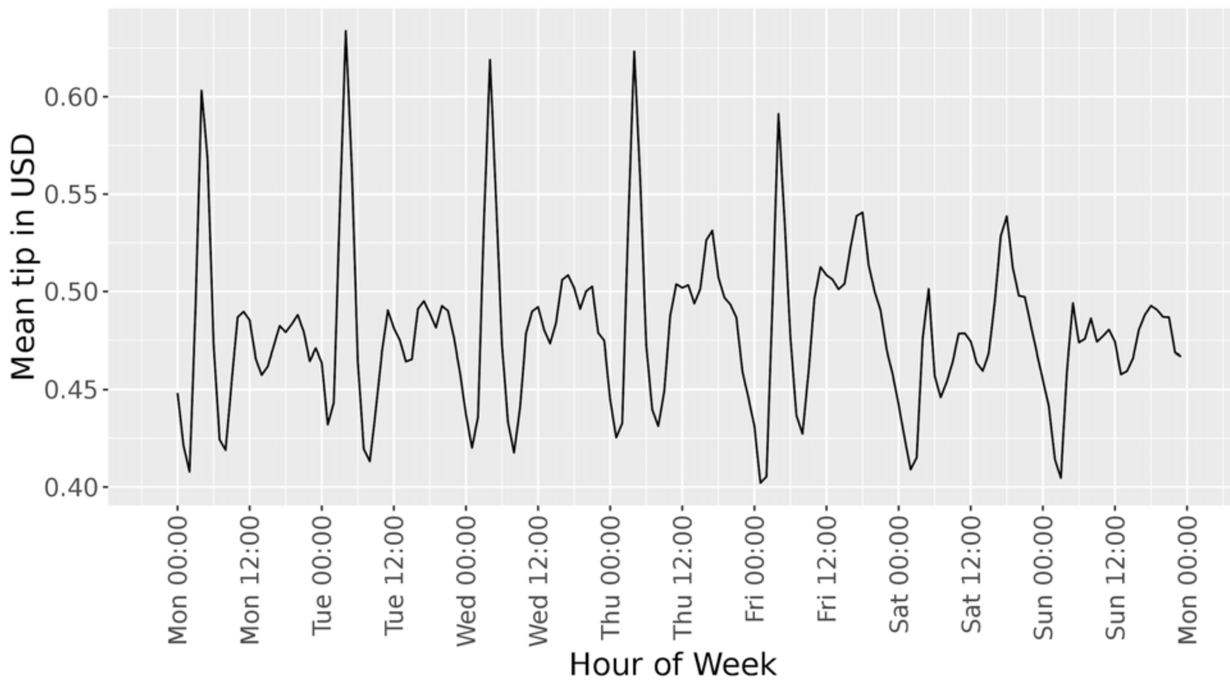


Figure 4: Average tip by hour of week across the United States, including trips that were not tipped (tipped $0).

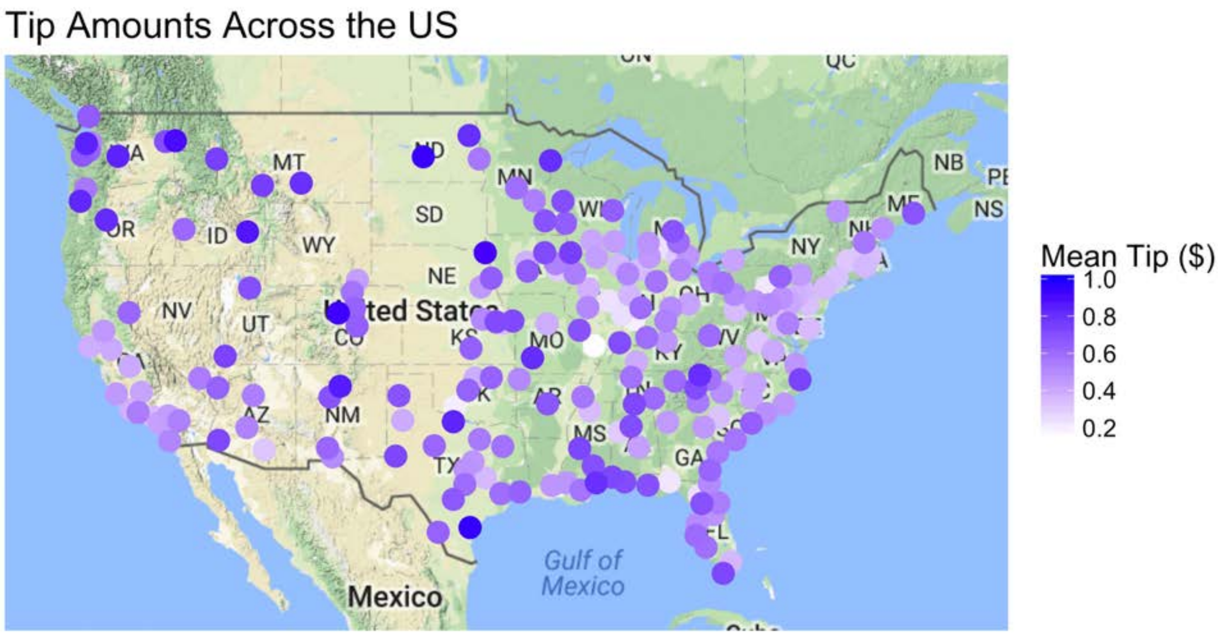


Figure 5: Average tips across cities in the United States. Tips tend to be highest in less dense areas in the middle of the country. They are lowest in many major cities along the Northeast and West Coast.

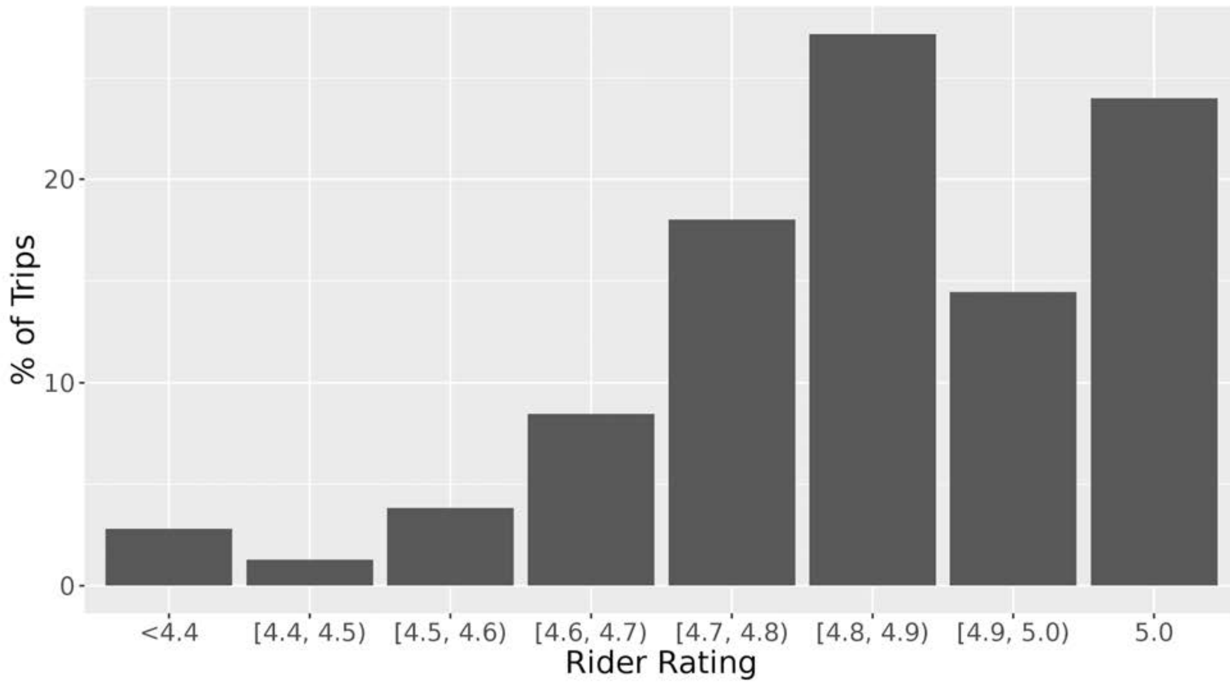


Figure 6: Distribution of rider lifetime ratings across trips, excluding missing ratings. A rider's lifetime rating is the average rating given to them by drivers over their past 500 trips. For riders that have taken fewer than 500 trips, their lifetime rating is their average rating over all of their past trips.

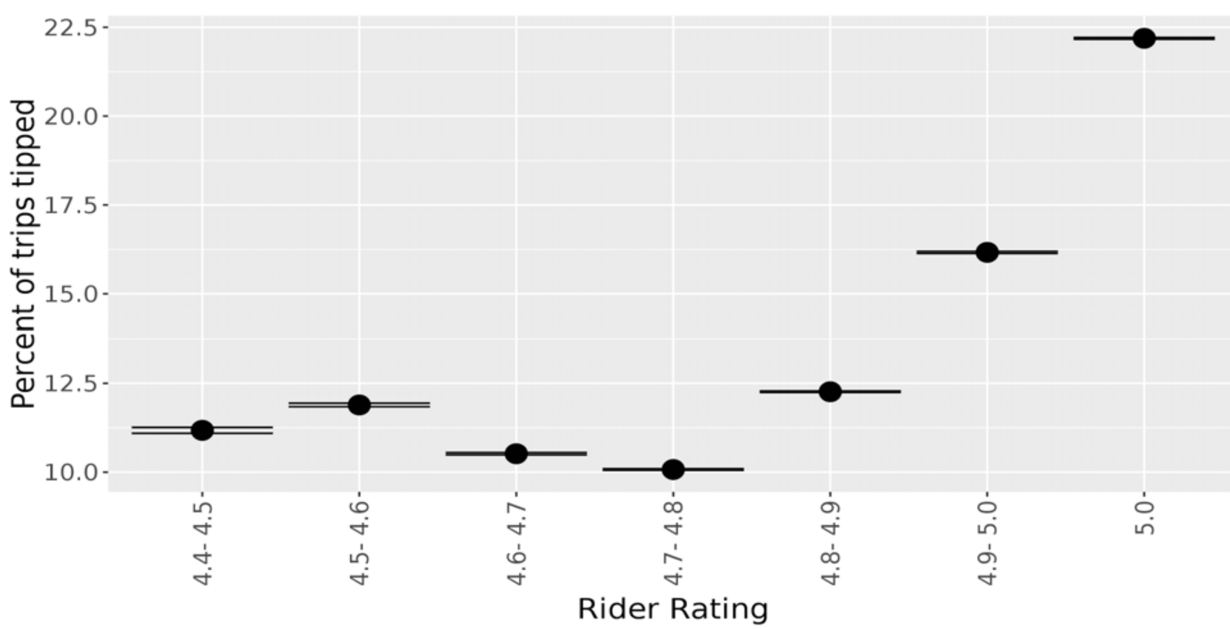


Figure 7: Percent of trips tipped by rider lifetime rating. A rider's lifetime rating is the average rating given to them by drivers over their past 500 trips. For riders that have taken fewer than 500 trips, their lifetime rating is their average rating over all of their past trips

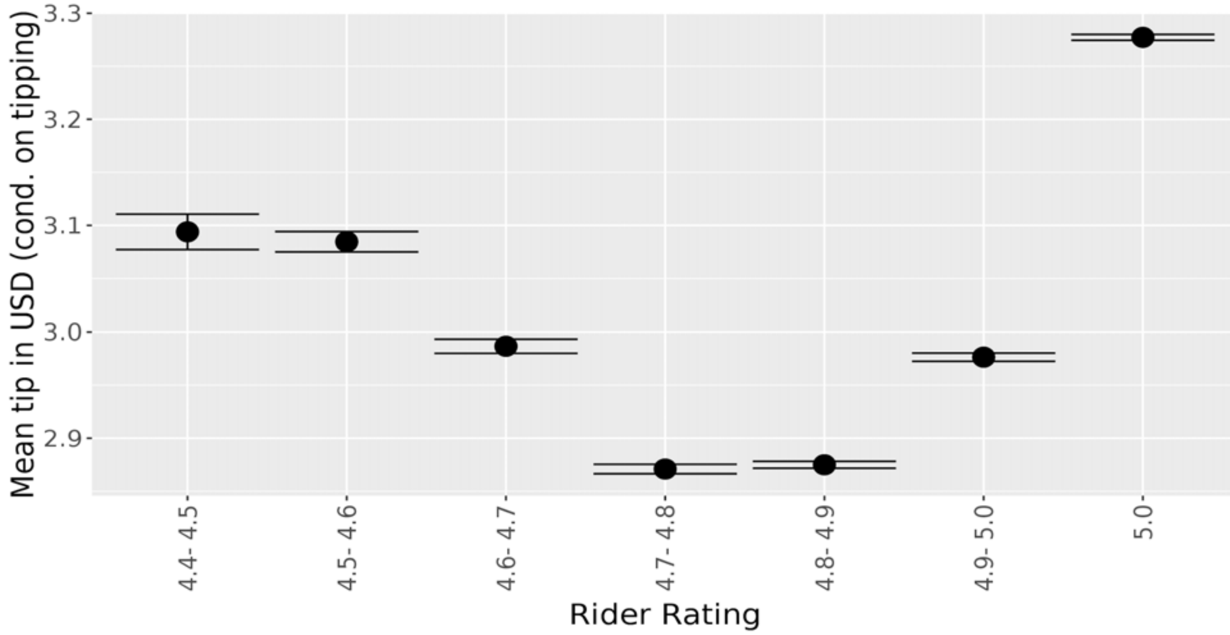


Figure 8: Average tip conditional on tipping by rider lifetime rating. A rider's lifetime rating is the average rating given to them by drivers over their past 500 trips. For riders that have taken fewer than 500 trips, their lifetime rating is their average rating over all of their past trips

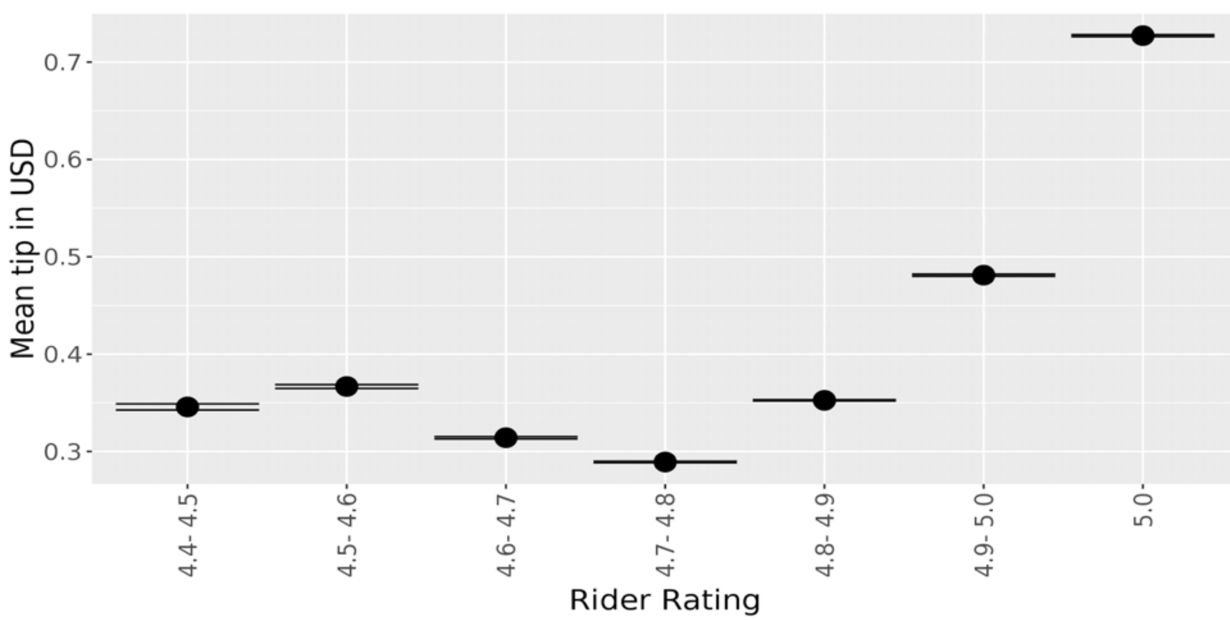


Figure 9: Average tip by rider lifetime rating. A rider's lifetime rating is the average rating given to them by drivers over their past 500 trips. For riders that have taken fewer than 500 trips, their lifetime rating is their average rating over all of their past trips

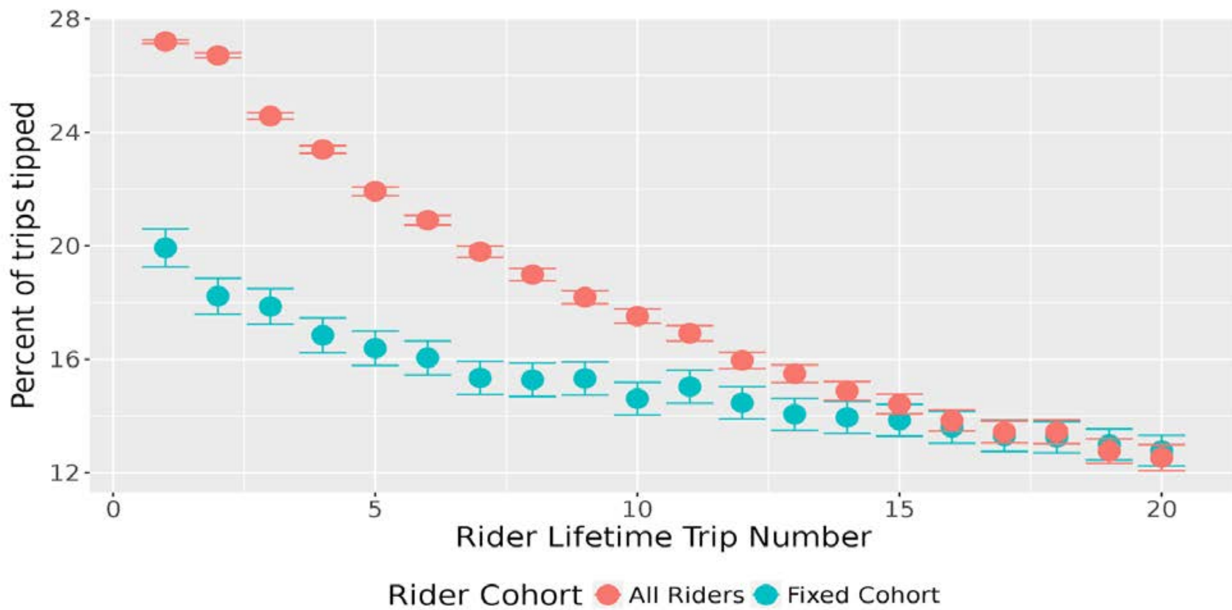


Figure 10: Probability of tipping by the number of trips a new rider has taken in their lifetime. The "All Riders" cohort includes all riders who took their first trip in the sample period. The "Fixed" cohort includes only riders who complete their first 20 trips in the sample period, so for that cohort each point includes the same set of riders.

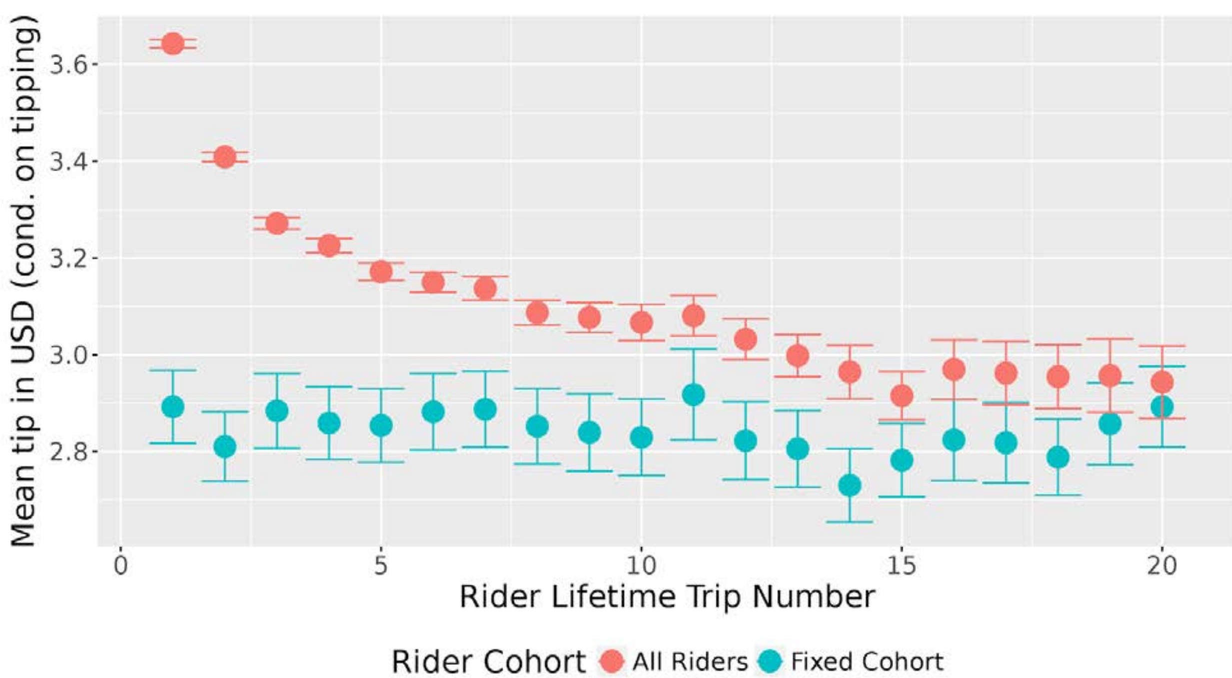


Figure 11: Average tip conditional on tipping by the number of trips a new rider has taken in their lifetime. The "All Riders" cohort includes all riders who took their first trip in the sample period. The "Fixed" cohort includes only riders who complete their first 20 trips in the sample period.

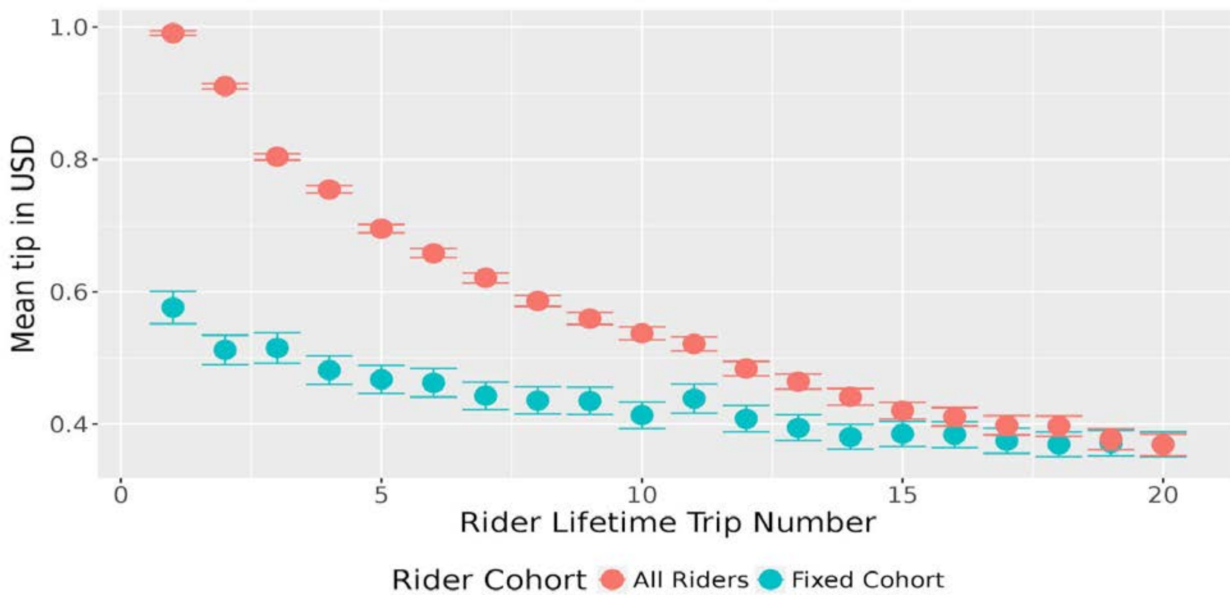


Figure 12: Average tip (including instances where the rider did not tip) by the number of trips a new rider has taken in their lifetime. The "All Riders" cohort includes all riders who took their first trip in the sample period. The "Fixed" cohort includes only riders who complete their first 20 trips in the sample period.

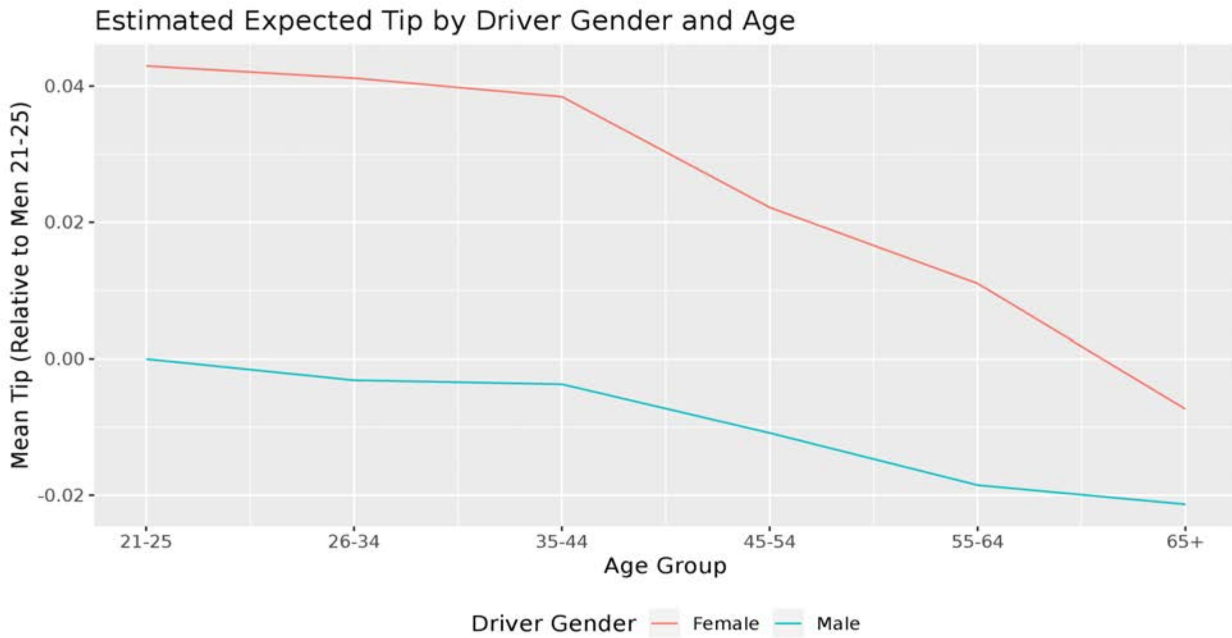


Figure 13: Fitted tip levels by the interaction of driver gender and age, controlling for time, location, and trip, rider, and driver covariates. Estimates are relative to male drivers between the ages of 21 and 25.

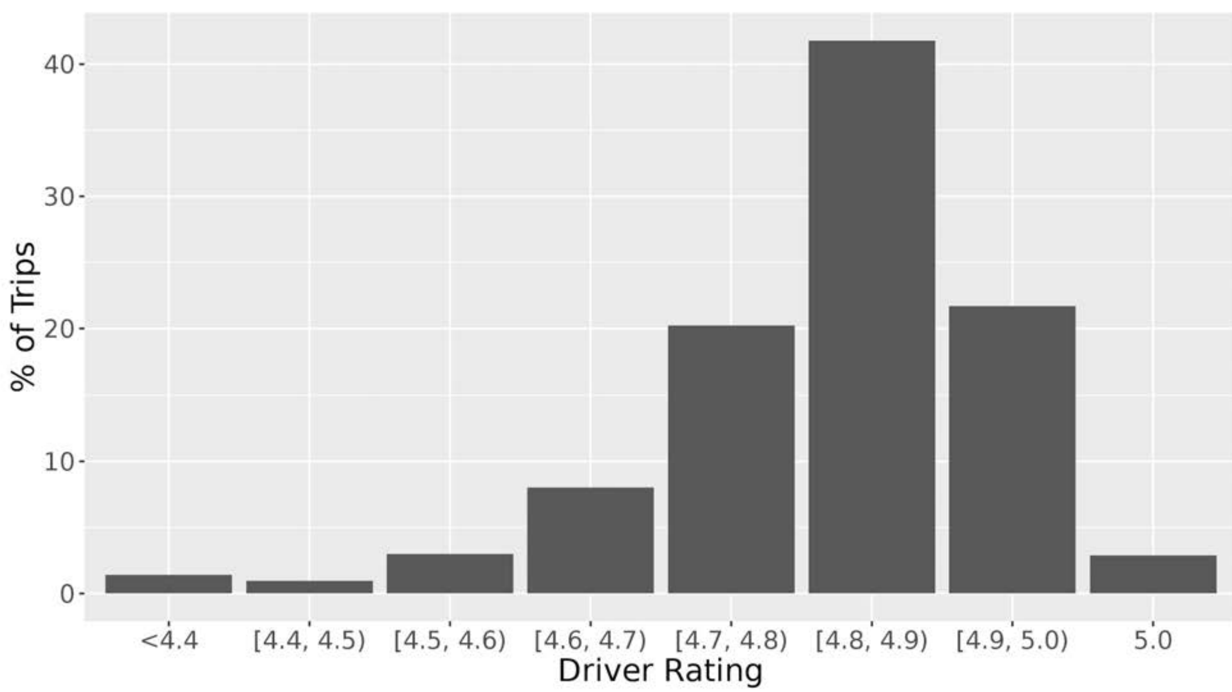


Figure 14: Distribution of driver lifetime ratings across trips, excluding missing ratings. A driver's lifetime rating is the average rating given to them by riders over their past 500 trips. For drivers that have taken fewer than 500 trips, their lifetime rating is their average rating over all of their past trips.

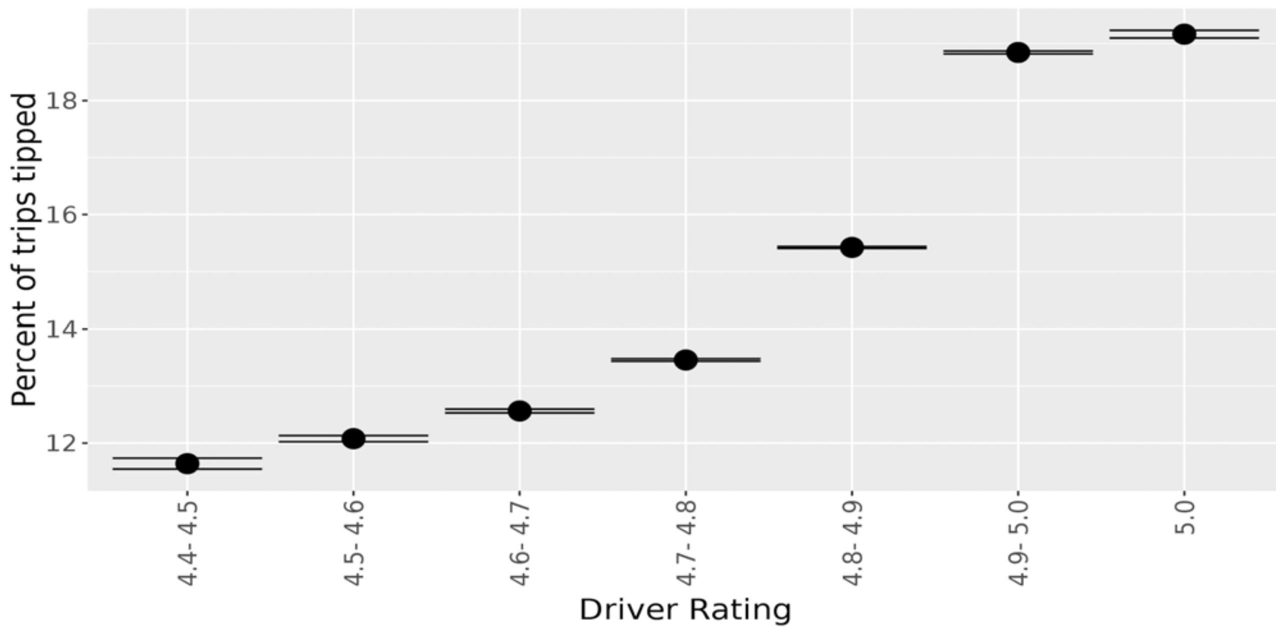


Figure 15: Percent of trips tipped by driver lifetime rating. A driver's lifetime rating is the average rating given to them by riders over their past 500 trips. For drivers that have taken fewer than 500 trips, their lifetime rating is their average rating over all of their past trips.

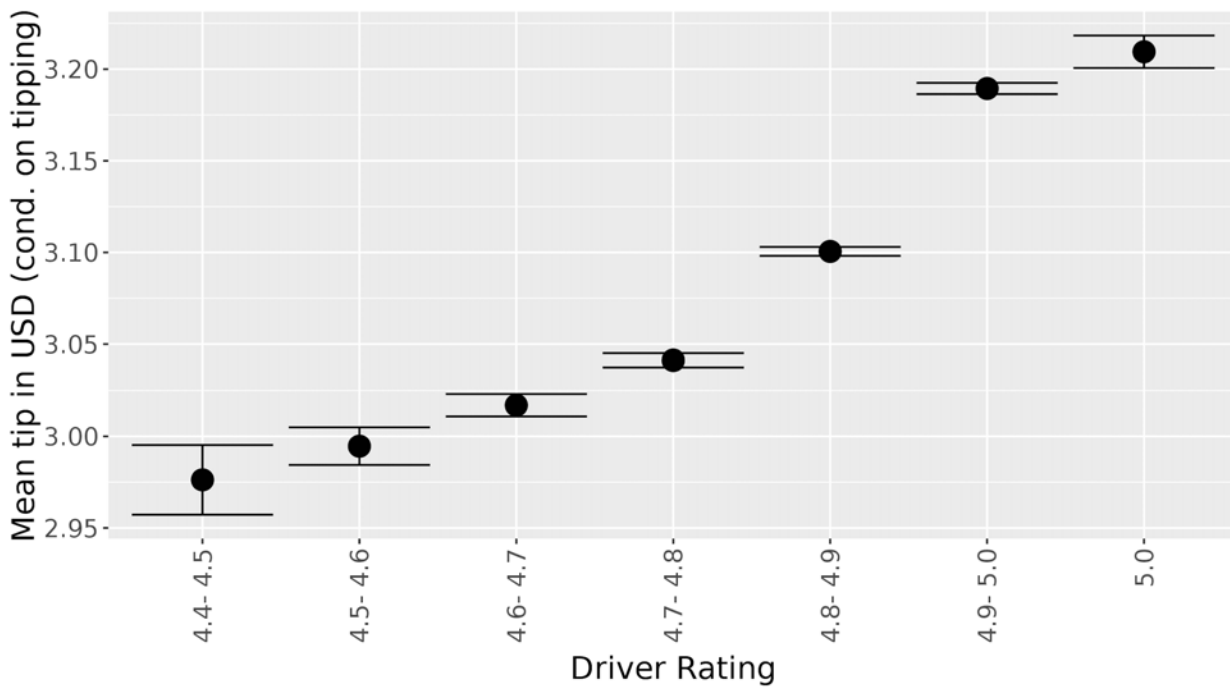


Figure 16: Average tip conditional on tipping by driver lifetime rating. A driver's lifetime rating is the average rating given to them by riders over their past 500 trips. For drivers that have taken fewer than 500 trips, their lifetime rating is their average rating over all of their past trips.

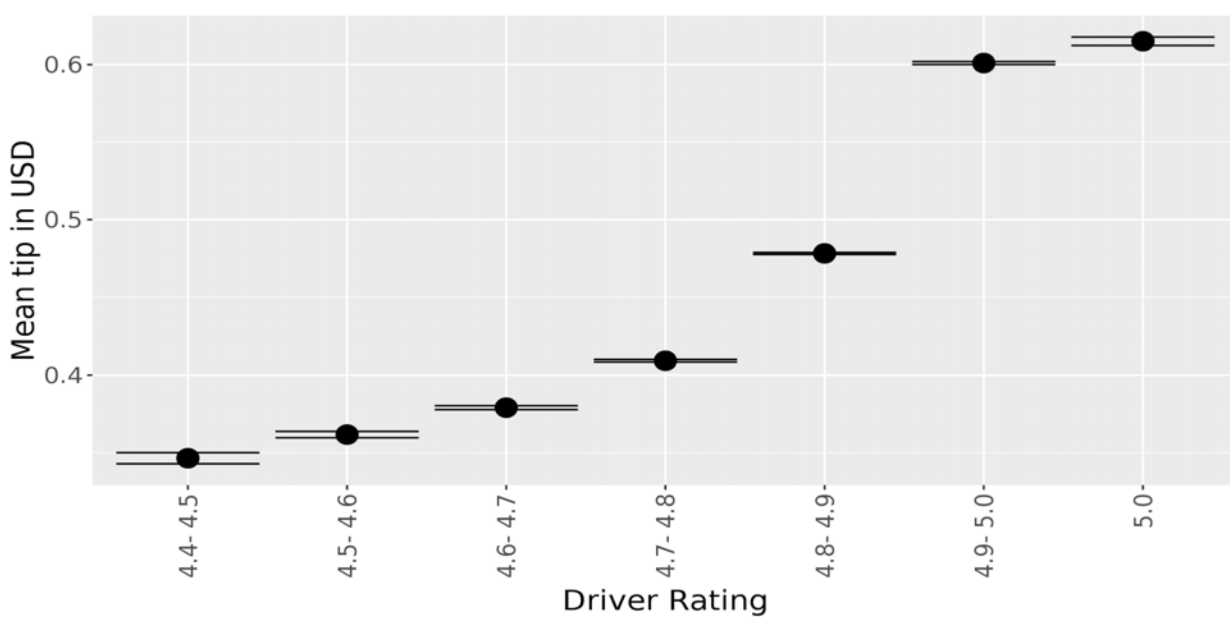


Figure 17: Average tip by driver lifetime rating. A driver's lifetime rating is the average rating given to them by riders over their past 500 trips. For drivers that have taken fewer than 500 trips, their lifetime rating is their average rating over all of their past trips.

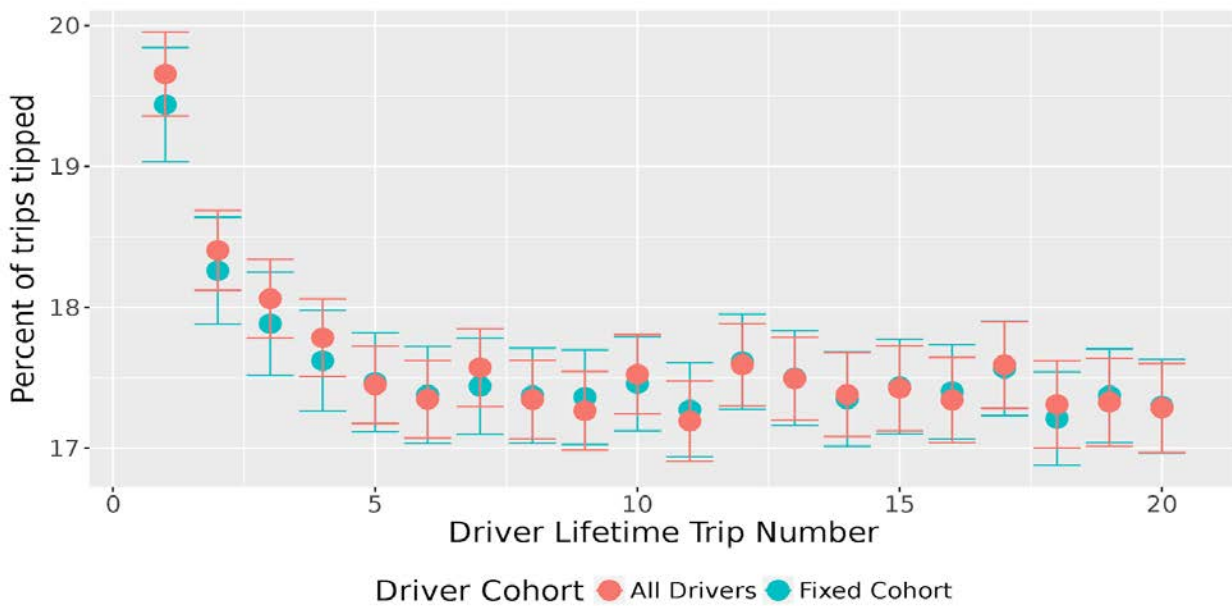


Figure 18: Probability of tipping by the number of trips a new driver has taken in their lifetime. Results are broken into two cohorts. The "All Drivers" cohort includes all drivers who took their first trip in the sample period. The "Fixed" cohort includes only drivers who complete their first 20 trips in the sample period, so for that cohort each point includes the same set of drivers.

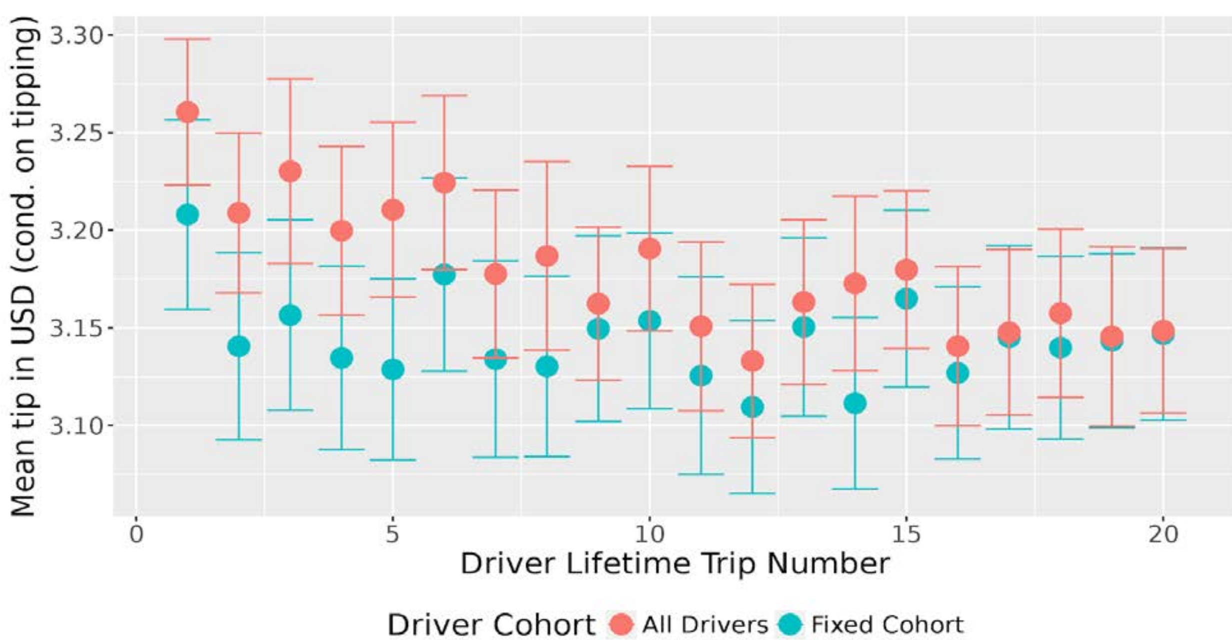


Figure 19: Average tip conditional on tipping by the number of trips a new driver has taken in their lifetime. The "All Drivers" cohort includes all drivers who took their first trip in the sample period. e. The "Fixed" cohort includes only drivers who complete their first 20 trips in the sample period, so for that cohort each point includes the same set of drivers.

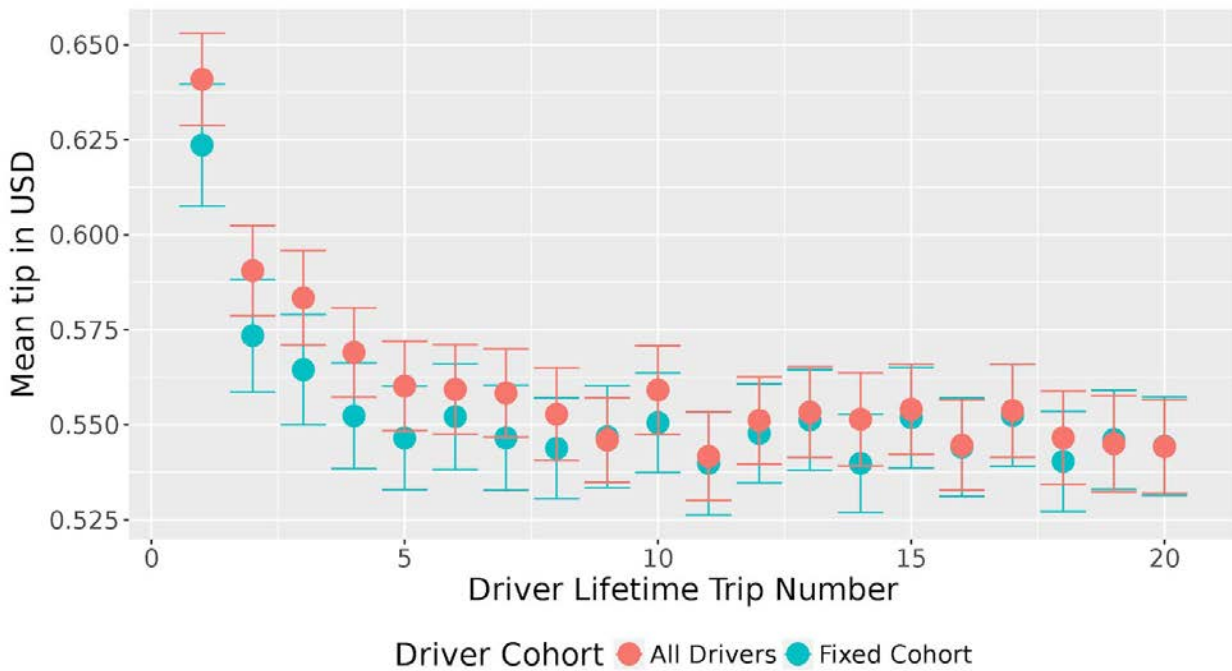


Figure 20: Average tip by the number of trips a new driver has taken in their lifetime. The "All Drivers" cohort includes all drivers who took their first trip in the sample period. e. The "Fixed" cohort includes only drivers who complete their first 20 trips in the sample period, so for that cohort each point includes the same set of drivers.

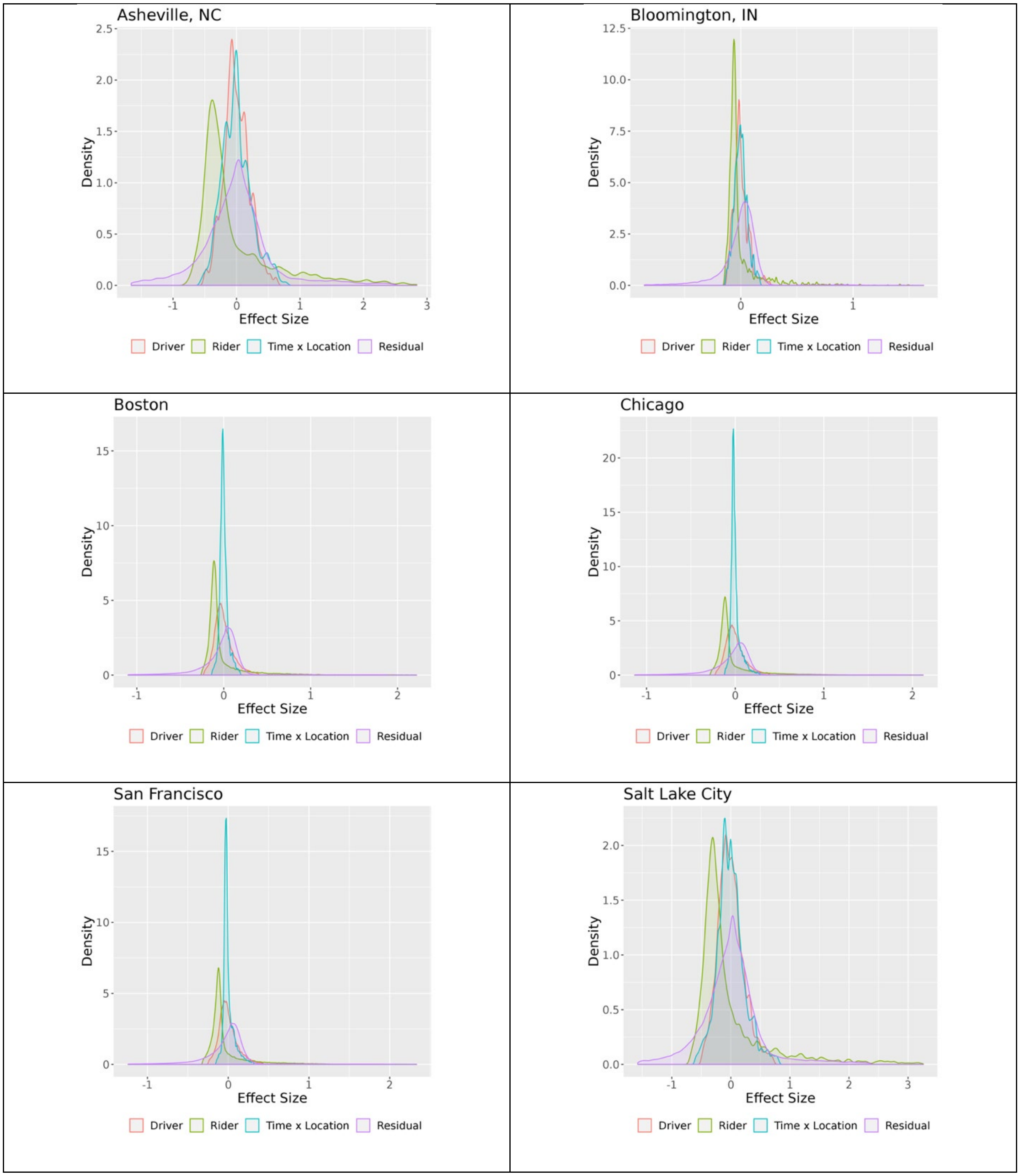


Figure 21: The figure above shows the densities for fixed effects estimated from the model in Equation 2. We include only riders, drivers, and (time x location) pairs with at least 10 trips in the sample period to protect against the results being driven by terms with few observations. We exclude estimated effects below the 2nd percentile and above the 98th percentile to ensure the densities are not dominated by outliers.

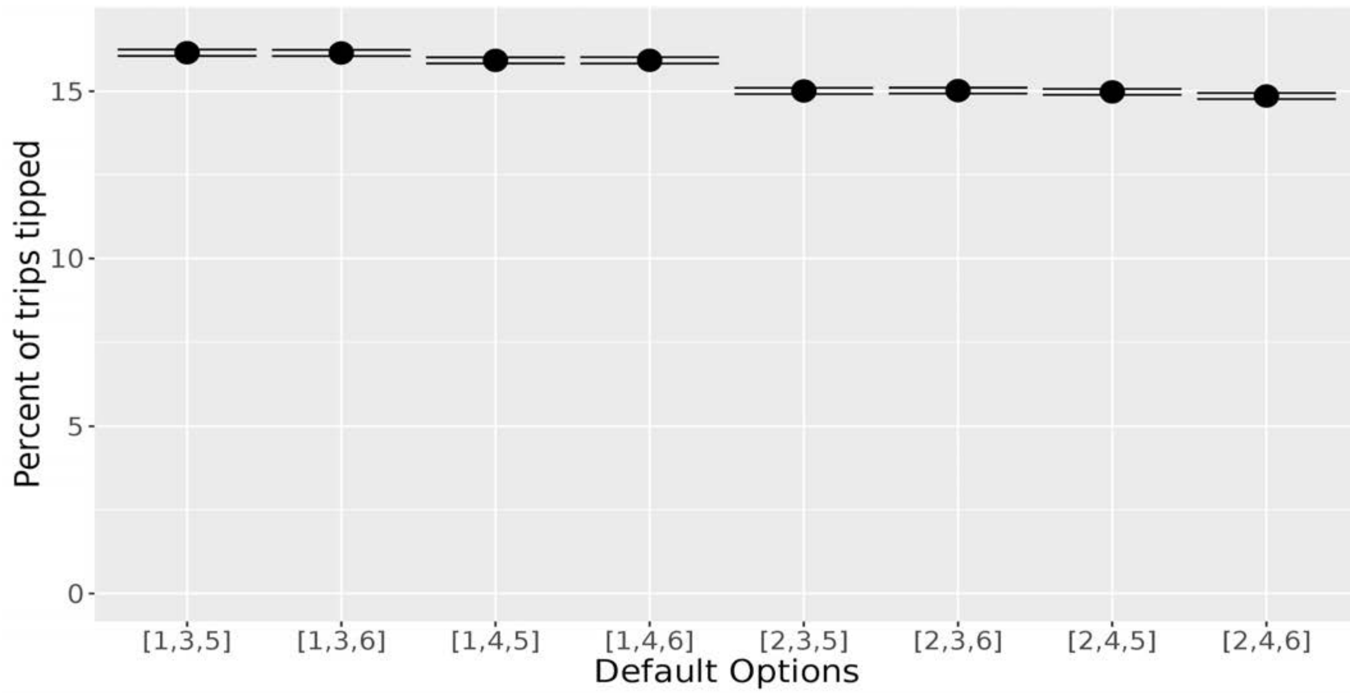


Figure 22: Percent of trips tipped by default options shown to the rider in the experiment. Estimates are clustered by rider.

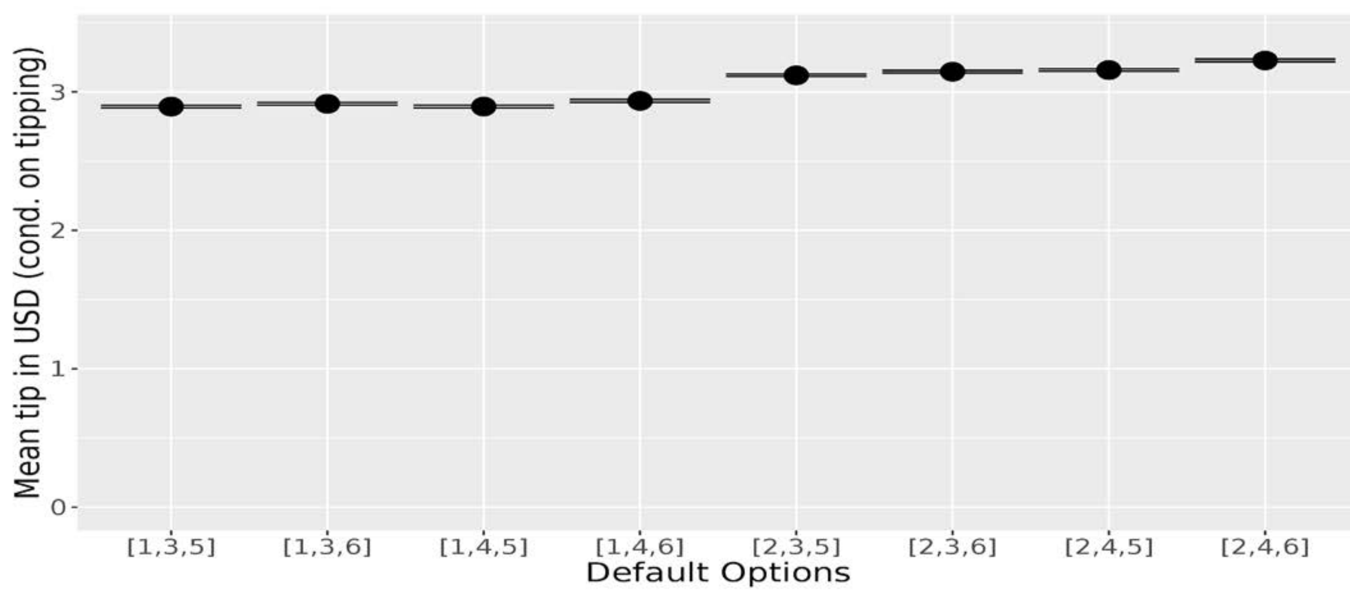


Figure 23: Average tip conditional on tipping by default options shown to the rider. Estimates are clustered by rider.

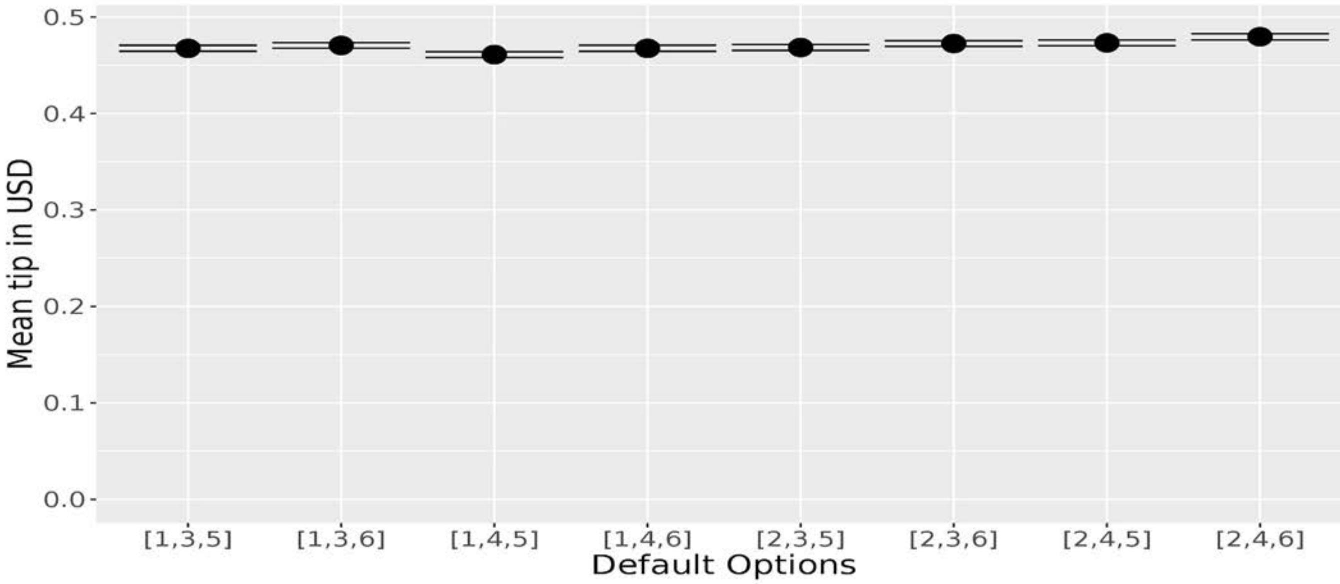


Figure 24: Average tip by default options shown to the rider. Estimates are clustered by rider.

a. Driver home ZIP demographic quintiles (across trips)

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Median Income	$32,921.76	$45,058.39	$54,533.35	$67,108.86	$92,749.37
% Black	1.28	4.17	8.84	19.23	56.06
% Hispanic	2.09	5.83	11.48	21.44	51.35
% Bachelor's+	13.32	22.84	31.41	41.89	59.80

b. Rider home ZIP demographic quintiles (across trips)

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Median Income	$34,344.42	$49,545.73	$62,725.80	$78,724.89	$111,990.37
% Black	0.87	2.88	5.77	12.52	43.71
% Hispanic	1.85	4.55	8.00	15.08	43.03
% Bachelor's+	16.95	31.15	43.82	57.24	74.50

Table 1: Driver and rider ZIP demographic quantiles (across trips). Uber has access to drivers' home ZIP codes through documents filled out upon sign up. Rider's home zip codes come from the billing ZIP codes on their credit or debit cards. We observe rider and driver home ZIP information for more than 80% of trips. In the table we match home zip codes to demographic data from the US census and report mean within each quintile for each of the demographic variables. Quintiles are computed across trips.

a: Average tip amounts by driver ZIP demographic quantile.

Demographic	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	Missing
Median Income	0.443	0.483	0.479	0.496	0.492	0.482
% Black	0.522	0.477	0.501	0.461	0.432	0.482
% Hispanic	0.510	0.507	0.49	0.461	0.425	0.482
% Bachelor's+	0.458	0.474	0.494	0.491	0.477	0.482

b: Mean tip amounts by rider ZIP code demographic quintiles.

Demographic	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	Missing
Median Income	0.388	0.471	0.507	0.498	0.509	0.499
% Black	0.566	0.489	0.478	0.471	0.369	0.500
% Hispanic	0.543	0.494	0.468	0.48	0.388	0.500
% Bachelor's+	0.421	0.504	0.519	0.489	0.440	0.500

Table 2: Mean tip amounts ($) by driver and rider ZIP code demographic quintiles.

	N. Riders	Pct. Riders	N. Trips	Pct. Trips	Pct. Tipped	Mean Tip |Tip	Mean Tip
Male	4,119,667	52.4	12,161,837	52.5	17.0	3.129	0.531
Female	3,248,882	41.3	9,429,153	40.7	14.3	3.067	0.439
Unmatched	489,655	6.2	1,555,177	6.7	11.1	2.893	0.321

Table 3: Summary statistics by imputed gender for riders on Uber. Uber does not collect the rider's gender. We impute the rider's gender using their first names by matching with name and gender data from the Social Security Administration. Because the SSA data only includes names given to at least five babies per year, uncommon names are not matched. More details are in Appendix Section 3.

	Dependent variable:					
	Tip Amount					
	(1)	(2)	(3)	(4)	(5)	(6)
Female Rider	-0.092^{***}	-0.092^{***}	-0.093^{***}	-0.068^{***}	-0.068^{***}	-0.060^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Unmatched Rider	-0.211^{***}	-0.210^{***}	-0.212^{***}	-0.179^{***}	-0.180^{***}	-0.171^{***}
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Constant	0.531^{***}					
	(0.001)					
Date		X	X	X	X	X
Hour of Week			X		X	X
Pick-up Geo				X	X	X
Drop-off Geo						X
Observations	23,146,167	23,146,167	23,146,167	23,146,167	23,146,167	23,146,167
R^2	0.002	0.002	0.003	0.021	0.021	0.029
Adjusted R^2	0.002	0.002	0.003	0.020	0.021	0.028
Residual Std. Error	1.387	1.387	1.387	1.374	1.374	1.369
df	23146164	23146151	23145990	23140105	23139944	23133161
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 4: Regression output for tip differences between male and female riders Time and location controls are included. The fist column includes no controls. Estimates are relative to male riders. The second column includes controls for the date of the trip. Column (3) includes controls for the date and the hour of the week of the trip. Column (4) includes controls for the pick-up location (coded as a level 5 geohash) and the date of the trip. Column (5) includes controls for the pick-up location, date of the trip, and hour of the week. Column (6) includes controls for the pick-up location, date of the trip, hour of the week, and drop-off location. Results when including additional trip, rider, and driver controls are in Appendix Table 1.

	N. Drivers	Pct. Drivers	N. Trips	Pct. Trips	Pct. Tipped	Mean Tip |Tip	Mean Tip
Male	513,410	78.0	19,302,308	83.4	15.3	3.083	0.470
Female	144,502	22.0	3,843,859	16.6	16.8	3.143	0.527

Table 5: Summary statistics by gender for drivers on Uber. A driver's gender is recorded by Uber as part of the sign-up process.

	Dependent variable:					
	Tip Amount					
	(1)	(2)	(3)	(4)	(5)	(6)
Female Driver	0.057***	0.057***	0.056***	0.047***	0.046***	0.048***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Constant	0.470***					
	(0.0005)					
Date		X	X	X	X	X
Hour of Week			X		X	X
Pick-up Geo				X	X	X
Drop-off Geo						X
Observations	23,146,167	23,146,167	23,146,167	23,146,167	23,146,167	23,146,167
R^2	0.0002	0.0003	0.001	0.020	0.020	0.028
Adjusted R^2	0.0002	0.0003	0.001	0.019	0.020	0.028
Residual Std. Error	1.388	1.388	1.388	1.375	1.375	1.369
df	23146165	23146152	23145991	23140106	23139945	23133162
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 6: Regression output for tip differences between male and female drivers. Time and location controls are included. The fist column includes no controls. Estimates are relative to male drivers.The second column includes controls for the date of the trip. Column (3) includes controls for the date and the hour of the week of the trip. Column (4) includes controls for the pick-up location (coded as a level 5 geohash) and the date of the trip. Column (5) includes controls for the pick-up location, date of the trip, and hour of the week. Column (6) includes controls for the pick-up location, date of the trip, hour of the week, and drop-

off location. Results when including additional trip, rider, and driver controls are in Appendix Table 2.

Rider Gender	Driver Gender	N. Trips	Pct. Trips	Pct. Tipped	Mean Tip |Tip	Mean Tip
Male	Male	10,156,152	43.9	16.7	3.116	0.520
Male	Female	2,005,685	8.7	18.5	3.188	0.588
Female	Male	7,852,109	33.9	14.1	3.060	0.431
Female	Female	1,577,044	6.8	15.5	3.100	0.479
Unmatched	Male	1,294,047	5.6	11.0	2.878	0.315
Unmatched	Female	261,130	1.1	11.8	2.958	0.349

Table 7: Summary statistics for the interaction between driver gender and imputed rider gender. The driver's gender is recorded by Uber as part of the sign-up process. The rider's gender is imputed using data from the US Social Security Administration.

City	Mean Tip ($)	# Trips	# Riders	# Drivers	# Time x Location
Asheville, NC	0.536	14032	1786	723	1025
Bloomington, IN	0.138	33703	2319	765	587
Boston	0.224	800711	66082	23664	6705
Chicago	0.23	1527234	117995	38824	11177
Salt Lake City	0.476	54202	6316	2950	4518
San Francisco	0.256	1467810	116894	37192	15075

Table 8: Summary statistics for tips across various cities. In the table above we filter to riders, drivers, and (time x location) pairs with at least 10 trips in the data.

Effect	Asheville, NC	Bloomington, IN	Boston	Chicago	Salt Lake City	San Francisco
Driver	0.200	0.066	0.107	0.108	0.215	0.115
Rider	0.654	0.182	0.307	0.311	0.640	0.348
Time x Location	0.235	0.057	0.047	0.056	0.230	0.068
Residual	0.569	0.192	0.293	0.311	0.510	0.334

Table 9: Standard deviation for the estimated fixed effects across the different cities from Equation 2 for the different sources of tip variation. In Equation 2, we regress tip outcomes on rider fixed effects, driver fixed effects, and time and location fixed effects for various cities in the United States. We use the standard deviation of the fixed effects as a measure of variation in tipping behavior across individuals, time, and space. We include only riders, drivers, and (time cross location) pairs with at least 10 trips in the sample period to protect against the results being driven by terms with few observations. We exclude estimated effects below the 2nd percentile and above the 98th percentile to ensure effects are not dominated by outliers.

Rider/Driver Match Occurrence	N. Trips	Pct. Tipped	Mean Tip |Tip	Mean Tip
First	46,017,895	0.154	3.103	0.478
Second	443,211	0.160	3.513	0.563
Third	39,954	0.168	3.828	0.643

Table 10: Summary statistics for tip outcomes by the number of times the rider and driver have matched with each other. Drivers and riders match with each other more than once infrequently, but we have sufficiently many observations that we still see many instances in which the rider and driver have seen each other more than once.

	Dependent variable:	
	Tip Amount	
	(1)	(2)
Period 2	-0.059^{***}	-0.050^{***}
	(0.003)	(0.003)
Period 3	-0.028^{***}	-0.032^{***}
	(0.003)	(0.003)
Constant	0.344^{***}	
	(0.002)	
Trip Controls		X
Observations	2,645,417	2,645,417
R^2	0.0004	0.019
Adjusted R^2	0.0004	0.019
Residual Std. Error	1.126 (df = 2645414)	1.116 (df = 2645378)
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 11: Regression results for tip levels when a rider matches with the same driver twice. The constant gives the expected tip amount for trips before the rider r matches with driver $d(r)$. The coefficient on Period 2 shows the change in tip amount for trips between the first match and the second match. The coefficient on Period 3 show the change in tip amount for trips after the second match.

APPENDIX:

1. Controls

The table below shows the controls we include in regressions when estimating Equation 1.

1.1 Trip Controls

Control Variable	**Explanation**
Duration	Log of trip duration in seconds
Distance	Log of trip distance in miles
Fare	Log of fare
Distance to pick up	Distance from the driver's dispatch location to rider's pick up location in miles
Is airport start	
Is airport destination	
Surge	The surge multiplier for the trip, discretized into a factor variable. Includes a factor level for no surge on the trip.
ATA - ETA	Actual time of arrival to pick up the rider minus expected time of arrival to pick up, in minutes
ATD - ETD	Actual time of arrival to the rider's destination minus expected time of arrival to the destination, in minutes
Is business trip	Whether the rider used a payment profile tied to an Uber for Business expense account
Any hard accelerations	Whether Uber estimates that there may have been a hard acceleration. Estimates are imperfect.
Any hard brakes	Whether Uber estimates there may have been a hard brake. Estimates are imperfect.
Did speed	Whether Uber estimates that there may have been speeding.

	Estimates are imperfect.
Average speed	Distance to destination divided by time to destination.
Is car from before 2010	

1.2 Rider Controls

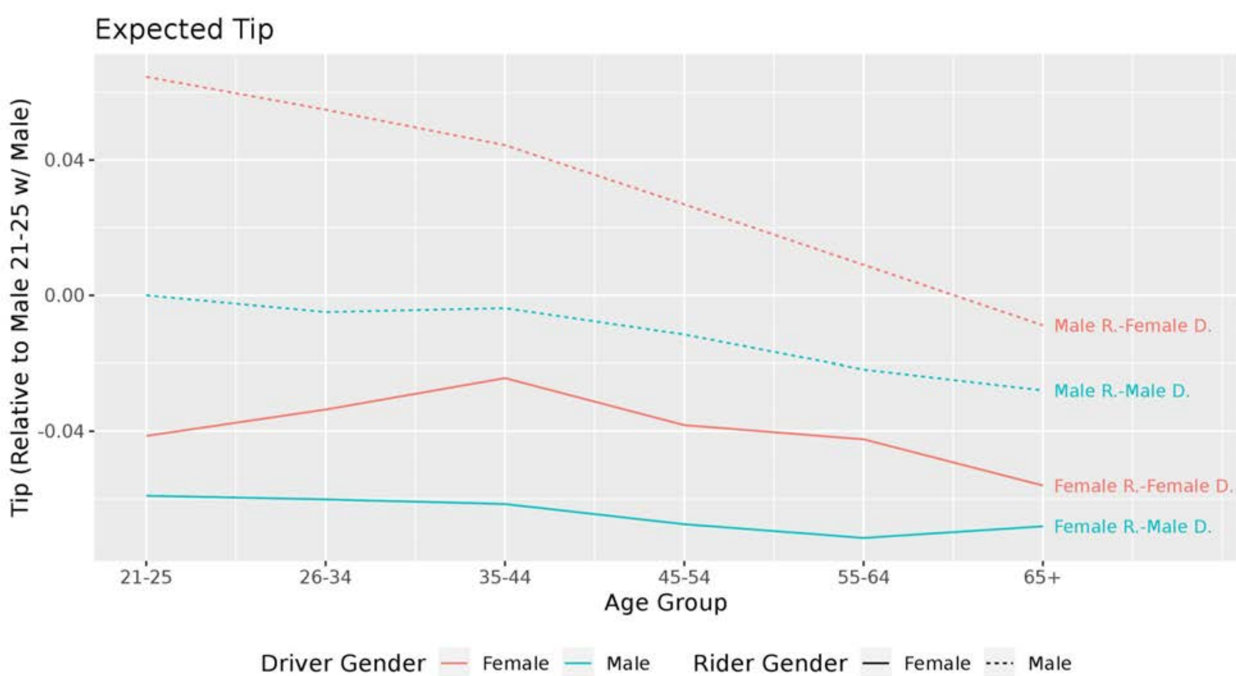
Control Variable	**Explanation**
Nudged rating screen	Rider's treatment status for the nudged rating screen experiment
Shown preset	The preset shown on the trip
Client OS	iOS or Android
Rider rating	Rescaled to be mean 0 and unit variance
Rider trip number	The number of trips the rider has taken, including the current trip. Rescaled to be mean 0 and unit variance.
Rider trips the month before	The number of trips the rider took in the month before the sample period
Rider gender (estimated)	
Rider home ZIP median income	Discretized by quintiles into a factor variable
Rider home ZIP % black	Discretized by quintiles into a factor variable
Rider home ZIP % Hispanic	Discretized by quintiles into a factor variable
Rider home ZIP % Bachelor's degree+	Discretized by quintiles into a factor variable

1.3 Driver Controls

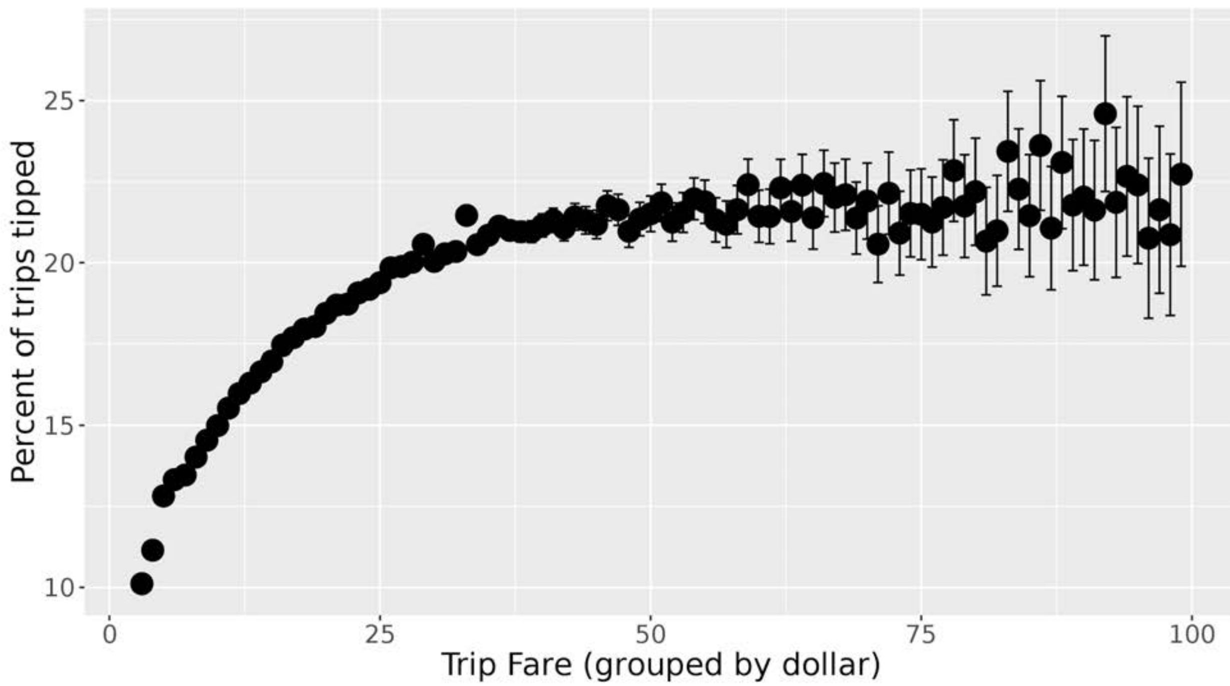
Control Variable	**Explanation**
Driver's age	Discretized into a factor variable with six levels
Is driver app in English	
Driver rating	Rescaled to be mean 0 and unit variance
Driver trip number	The number of trips the driver has taken,

	including the current trip. Rescaled to be mean 0 and unit variance.
Driver trips the month before	The number of trips the driver took in the month before the sample period
Driver gender	
Driver home ZIP median income	Discretized by quintiles into a factor variable
Driver home ZIP % black	Discretized by quintiles into a factor variable
Driver home ZIP % Hispanic	Discretized by quintiles into a factor variable
Driver home ZIP % Bachelor's degree+	Discretized by quintiles into a factor variable

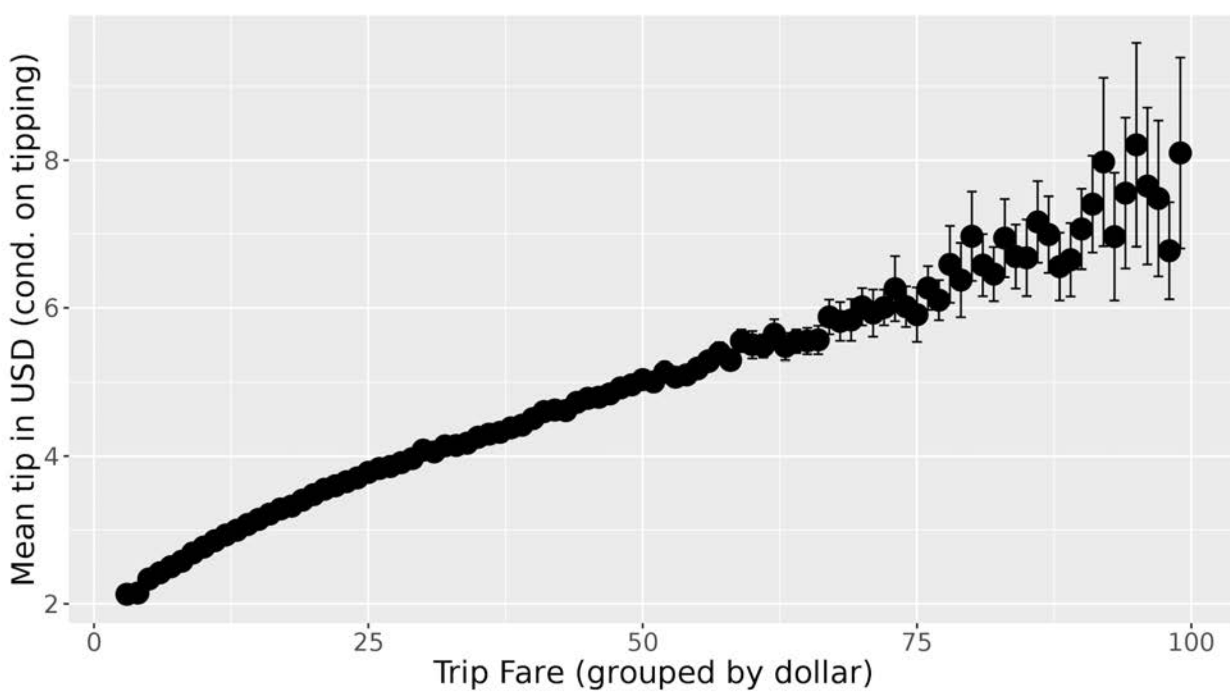
2. Supporting Results



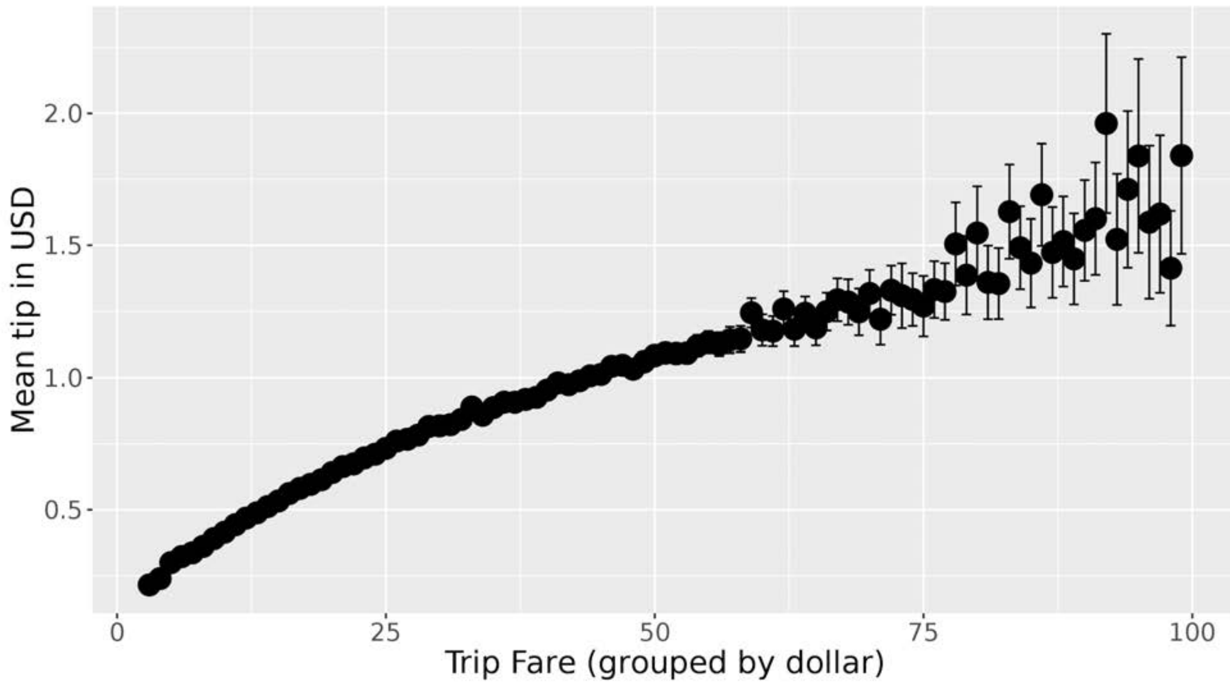
Appendix Figure 1: Fitted tip levels by the interaction of driver gender, rider gender, and age, controlling for time, location, and trip, rider, and driver covariates. Estimates are relative to male drivers between the ages of 21 and 25 matched with male riders.



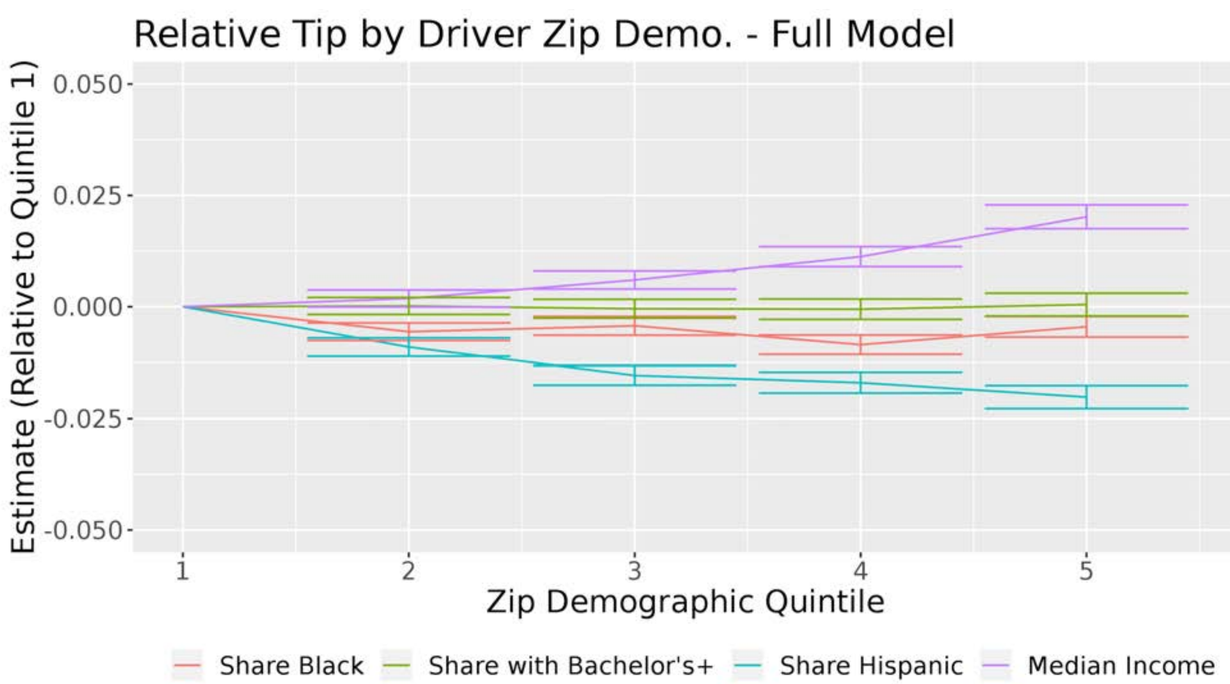
Appendix Figure 2: Percent of trips tipped by trip fare, rounded to the nearest dollar.



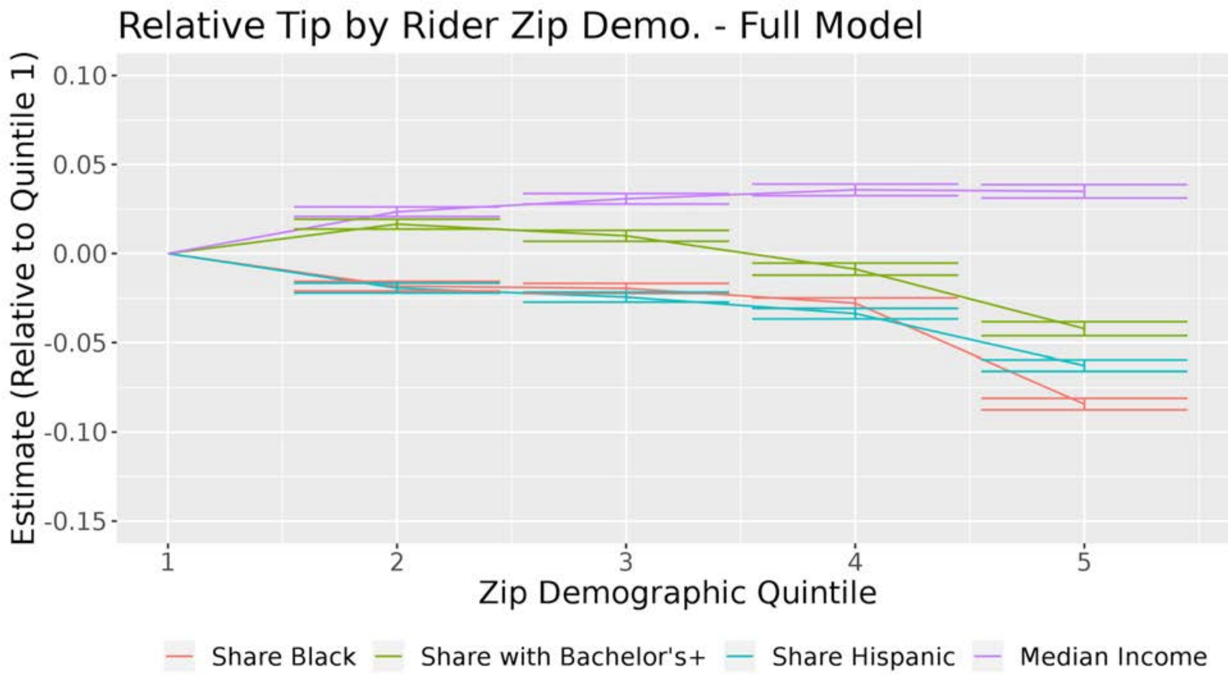
Appendix Figure 3: Average tip conditional on tipping by trip fare, rounded to the nearest dollar.



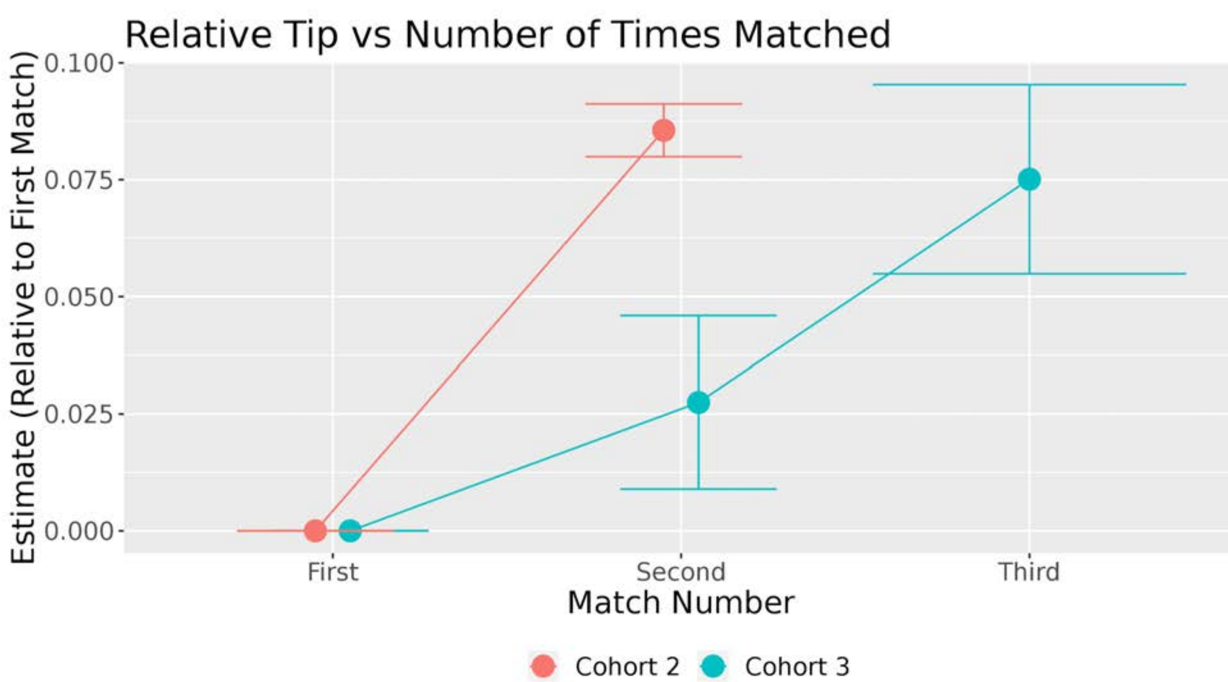
Appendix Figure 4: Average tip by trip fare, rounded to the nearest dollar.



Appendix Figure 5: Fitted tip amounts by driver ZIP demographic quintile. Controlling for where and when the trip happens as well as trip, rider, and driver covariates.



Appendix Figure 6: Fitted tip amount by rider ZIP demographic quintile. Controlling for where and when the trip happens as well as trip, rider, and driver covariates.



Appendix Figure 7: Fitted tip level against the number of times the rider and driver have matched with each other. Split by cohort of the number of times the rider and driver match with each other overall. Estimates are relative to the first match. Estimates control for trip characteristics included in Appendix 1.1.

	Dependent variable:		
	Tip Amount		
	(1)	(2)	(3)
Female Rider	−0.056***	−0.056***	−0.057***
	(0.001)	(0.001)	(0.001)
Unmatched Rider	−0.169***	−0.169***	−0.149***
	(0.002)	(0.002)	(0.002)
Date	X	X	X
Hour of Week	X	X	X
Pick-up Geo	X	X	X
Drop-off Geo	X	X	X
Trip Characteristics	X	X	X
Driver Characteristics		X	X
Rider Characteristics			X
Observations	23,146,167	23,146,167	23,146,167
R^2	0.042	0.043	0.055
Adjusted R^2	0.041	0.043	0.054
Residual Std. Error	1.360 (df = 23133125)	1.358 (df = 23133093)	1.350 (df = 23133052)
Note:			*p<0.1; **p<0.05; ***p<0.01

Appendix Table 1: Regression output for tip differences between male and female riders. Controlling for time, location, and trip, rider, and driver covariates.

	Dependent variable:
	Trip Tipped
Rider Rating (normalized)	0.012***
	(0.0001)
Rider Trips Taken (normalized)	−0.016***
	(0.0001)
Driver Rating (normalized)	0.014***
	(0.0001)
Driver Trips Taken (normalized)	−0.002***
	(0.0001)
log(Fare + 1)	0.006***
	(0.0002)
Delay in Pickup (normalized)	−0.002***
	(0.0001)
Delay in Dropoff (normalized)	0.003***
	(0.0001)
Any Hard Accelerations	−0.001***
	(0.0001)
Any Hard Brakes	−0.001***
	(0.0001)
Was Speeding	−0.005***
	(0.0001)
Model Year Before 2009	−0.003***
	(0.0001)
Driver Language Not English	−0.019***
	(0.0002)
Date	X
Hour of Week	X
Pick-up Geo	X
Drop-off Geo	X
Trip Characteristics	X
Driver Characteristics	X
Rider Characteristics	X
Observations	46,521,513
R^2	0.043
Adjusted R^2	0.043
Residual Std. Error	0.353 (df = 46508332)
Note:	*p<0.1; **p<0.05; ***p<0.01

Appendix Table 2: Regression estimates for the effect of various predictors discussed in the text on the likelihood a trip is tipped. Controlling for time, location, and other trip, rider, and driver covariates. For covariates marked normalized, we subtracted the mean and divided by the standard deviation before including it in the regression.

	Dependent variable:
	Tip Amount (conditional on tipping)
Rider Rating (normalized)	0.012***
	(0.0001)
Rider Trips Taken (normalized)	−0.016***
	(0.0001)
Driver Rating (normalized)	0.014***
	(0.0001)
Driver Trips Taken (normalized)	−0.002***
	(0.0001)
log(Fare + 1)	0.006***
	(0.0002)
Delay in Pickup (normalized)	−0.002***
	(0.0001)
Delay in Dropoff (normalized)	0.003***
	(0.0001)
Any Hard Accelerations	−0.001***
	(0.0001)
Any Hard Brakes	−0.001***
	(0.0001)
Was Speeding	−0.005***
	(0.0001)
Model Year Before 2009	−0.003***
	(0.0001)
Driver Language Not English	−0.019***
	(0.0002)
Date	X
Hour of Week	X
Pick-up Geo	X
Drop-off Geo	X
Trip Characteristics	X
Driver Characteristics	X
Rider Characteristics	X
Observations	7,169,095
R^2	0.197
Adjusted R^2	0.196
Residual Std. Error	1.932 (df = 7155946)
Note:	*p<0.1; **p<0.05; ***p<0.01

Appendix Table 3: Regression estimates for the effect of various predictors discussed in the text on the average tip, including only trips that are tipped. Controlling for time, location, and other trip, rider, and driver covariates. For covariates marked normalized, we subtracted the mean and divided by the standard deviation before including it in the regression.

	Dependent variable:
	Tip Amount
Rider Rating (normalized)	0.039***
	(0.0002)
Rider Trips Taken (normalized)	−0.052***
	(0.0003)
Driver Rating (normalized)	0.046***
	(0.0002)
Driver Trips Taken (normalized)	−0.005***
	(0.0003)
log(Fare + 1)	0.137***
	(0.001)
Delay in Pickup (normalized)	−0.007***
	(0.0002)
Delay in Dropoff (normalized)	0.036***
	(0.0002)
Any Hard Accelerations	−0.007***
	(0.0005)
Any Hard Brakes	−0.013***
	(0.0005)
Was Speeding	−0.029***
	(0.001)
Model Year Before 2009	−0.013***
	(0.001)
Driver Language Not English	−0.078***
	(0.001)
Date	X
Hour of Week	X
Pick-up Geo	X
Drop-off Geo	X
Trip Characteristics	X
Driver Characteristics	X
Rider Characteristics	X
Observations	46,521,513
R^2	0.053
Adjusted R^2	0.053
Residual Std. Error	1.368 (df = 46508332)
Note:	*p<0.1; **p<0.05; ***p<0.01

Appendix Table 4: Regression estimates for the effect of various predictors discussed in the text on the average tip. Controlling for time, location, and other trip, rider, and driver covariates. For covariates marked normalized, we subtracted the mean and divided by the standard deviation before including it in the regression.

	Dependent variable:		
	Tip Amount		
	(1)	(2)	(3)
Female Driver	0.045***	0.037***	0.037***
	(0.001)	(0.001)	(0.001)
Date	X	X	X
Hour of Week	X	X	X
Pick-up Geo	X	X	X
Drop-off Geo	X	X	X
Trip Characteristics	X	X	X
Driver Characteristics		X	X
Rider Characteristics			X
Observations	23,146,167	23,146,167	23,146,167
R^2	0.041	0.042	0.055
Adjusted R^2	0.040	0.042	0.054
Residual Std. Error	1.360 (df = 23133126)	1.359 (df = 23133095)	1.350 (df = 23133052)

Note: *p<0.1; **p<0.05; ***p<0.01

Appendix Table 5: Regression output for tip differences between male and female drivers. Controlling for time, location, and trip, rider, and driver covariates.

a. No controls added

	Male Rider	Female Rider	Unmatched Rider
Male Driver	0	-0.089	-0.205
Female Driver	0.068	-0.041	-0.171

b. Location and time controls added

	Male Rider	Female Rider	Unmatched Rider
Male Driver	0	-0.057	-0.167
Female Driver	0.058	-0.018	-0.135

c. Full set of controls added.

	Male Rider	Female Rider	Unmatched Rider
Male Driver	0	-0.054	-0.145
Female Driver	0.046	-0.026	-0.123

Appendix Table 6: Fitted values for interactions between driver and rider genders. Estimates are relative to male drivers matched to male riders. In table a no controls are added. Table b includes controls for the time and location of the trip. Table c includes controls for time, location, and other trip, rider, driver controls used in estimating Equation 1.

	Dependent variable:	
	Tip Amount	
	(1)	(2)
Second Interaction	0.093***	0.070***
	(0.011)	(0.011)
Constant	0.274***	
	(0.010)	
Trip Controls		✓
Observations	34,678	34,678
R^2	0.001	0.026
Adjusted R^2	0.001	0.025
Residual Std. Error	1.365 (df = 34676)	1.348 (df = 34641)
Note:		*p<0.1; **p<0.05; ***p<0.01

Appendix Table 7: Regression results for tip levels when a rider matches with the same driver twice, including only instances where the driver uses a default app language other than English. The constant gives the expected tip amount for the first interaction between rider r and driver $d(r)$. The coefficient on Second Interaction shows the change in tip amount on the second interaction. The increase in tip levels on the second interaction is very similar to the effect size in Appendix Figure 7. If conversation is less likely when the driver is not a native English speaker, then conversation is not the dominant mechanism through which repeated interaction leads to higher tips.

	Effect on Percent of Trips Tipped	Effect on Mean Tip in USD (conditional on tipping)	Effect on Mean Tip in USD
	(1)	(2)	(3)
First Option \$2 Instead of \$1	−0.0107***	0.2530***	0.0068***
	(0.0003)	(0.0034)	(0.0011)
Second Option \$4 Instead of \$3	−0.0016***	0.0344***	0.0006
	(0.0003)	(0.0035)	(0.0011)
Third Option \$6 Instead of \$5	−0.0003	0.0386***	0.0050***
	(0.0003)	(0.0035)	(0.0011)
Constant	0.1613***	2.8728***	0.4637***
	(0.0003)	(0.0034)	(0.0011)
Observations	21,401,912	3,317,893	21,401,912
R^2	0.0002	0.0038	0.00001
Adjusted R^2	0.0002	0.0038	0.00001
Residual Std. Error	0.3619 (df = 21401908)	2.0854 (df = 3317889)	1.3713 (df = 21401908)
Note:			*p<0.1; **p<0.05; ***p<0.01

Appendix Table 8: Marginal effect of changes in preset options for experiment 1. In our presets experiment, riders were randomized into having \$1 or \$2 as the first preset digit, \$3 or \$4 as the second preset digit, and \$5 or \$6 as the third preset digit. Estimates in the table above are clustered by rider.

3. Imputing Rider Gender

The Social Security Administration maintains an extensive record of names given at the time of birth for both males and females for each year from 1880 to the present. All names that occur at least 5 times nationally for a year-gender pair are included in the data for that year. We collect all data from 1916 through 2016 and aggregate across years to construct a data set with each name and the number of times a baby was given that name at birth for each gender. Because women are more likely to have very uncommon names than men and the most uncommon names are excluded, there are 4.4% more men in the SSA data than women. Let n_{female} and n_{male} be the total number of females and males in the SSA data, respectively. Let $n_{female,name}$ and $n_{male,name}$ be the number of occurrences of a given name for females and males in the data. To estimate the probability a name corresponds to a female we compute:

$$P(female|name) = \frac{(n_{female,name}) / (n_{female})}{(n_{female,name}) / (n_{female}) + (n_{male,name}) / (n_{male})}$$

Not every Uber rider name matches with the SSA names data set. Some modifications we make to rider names to improve the match rate are removing case and keeping only the first word in names that are multiple words or use hyphens. After these modifications the remaining unmatched names tend to be foreign names that are likely given infrequently in the US, abbreviations of more common names, or fictitious names the rider provided instead of their real name. 93.3% of trips have a matched rider name, but 77.6% of unique rider names are unmatched. A list of the 40 most common names that are unmatched is in Appendix Table 9.

First Name	Number of Riders
deleted	16,800
j	5,604
d	2,819
m	2,492
a	2,301
c	2,164
k	1,995
t	1,950
b	1,628
s	1,538
rmsad	1,506
r	1,374
l	1,289
g	1,163
e	1,088
cat	1,076
h	823
p	739
v	655
n	633
dr	590
andr	508
f	366
w	348
o	326
pankaj	320
z	318
i	316
venkata	305
q	297
stef	286
em	272
madhu	270
raghu	239
bala	220
ms	216
fei	214
vik	214
ant	211
y	205

Appendix Table 9: The 40 most common unmatched first names from our rider gender imputation procedure.

4. Variance Decomposition - Robustness

4.1 Results When Including the Tails of the Distribution

In Table 9 we remove estimated effects that are below the 2nd and above the 98th percentile to ensure results are not driven by outliers. In Appendix Table 10 we do not remove the tails of the effect distributions. Client effects explain an even larger share of variance.

Effect	Asheville, NC	Bloomington, IN	Boston	Chicago	Salt Lake City	San Francisco
Driver	0.278	0.140	0.161	0.151	0.290	0.167
Client	0.931	0.417	0.618	0.595	0.945	0.654
Time x Location	0.354	0.089	0.086	0.093	0.345	0.113
Residual	0.910	0.542	0.672	0.688	0.835	0.753

Appendix Table 10: Standard deviation of estimated effects for the different sources of tip variation.

We find a similar result for ratings as well, shown in Appendix Table 11. Client effects for ratings are relatively less important than for tipping.

Effect	Asheville, NC	Bloomington, IN	Boston	Chicago	Salt Lake City	San Francisco
Driver	0.222	0.143	0.195	0.178	0.204	0.159
Rider	0.360	0.365	0.410	0.392	0.384	0.372
Time x Location	0.202	0.124	0.094	0.088	0.245	0.096
Residual	0.339	0.445	0.448	0.457	0.363	0.425

Appendix Table 11: Standard deviation of estimated effects for the different sources of rider to driver rating variation.

4.2 Accounting for Different Number of Trips

While we only kept drivers, riders, and (time cross location) pairs with at least 10 trips overall between August 18 and September 14, 2017, in the resulting data set there are fewer observations per effect. As an example, though a rider may have taken ten or more trips between August 18 and September 15, 2017, any of those trips that occurred with a driver who took fewer than ten trips would get dropped. Appendix Table 12 shows summary statistics of the number of trips per source of variation in the resulting data set for Chicago.

Effect	Mean	Std. Dev.	Min	25th Perc.	Median	75th Perc.	Max
Driver	39.337	34.756	1	13	29	54	278
Rider	12.943	9.510	1	7	11	16	157
Time x Location	136.641	415.809	1	11	22	58	5,794

Appendix Table 12: Summary statistics for the number of trips each effect type is estimated over.

Higher variance in rider effects could result from them taking fewer trips. Appendix Table 13 below shows the standard deviation of effects across trips for each source of variation when only considering effects built on between 10 and 20 observations. Appendix Table 13 excludes the tails of the effect distributions.

Effect	Asheville, NC	Bloomington, IN	Boston	Chicago	Salt Lake City	San Francisco
Driver	0.227	0.094	0.140	0.137	0.250	0.157
Rider	0.639	0.172	0.271	0.281	0.558	0.320
Time x Location	0.259	0.077	0.083	0.100	0.237	0.115

Appendix Table 13: Standard deviation of estimated effects for the different sources of tip variation. We only include effects for riders, drivers, and (time x location pairs) estimated with between 10 and 20 observations. Effects below the 2nd percentile and above the 98th percentile for a given effect type are excluded to ensure estimates are not driven by outliers.

When making trip counts more similar, rider effects remain about three times more important than driver effects in cities with high tip levels. They are about twice as important in cities with lower tip levels.

Finally, it is still possible that driver effects are deflated because more of their trips are matched with riders that have few trips. Most of the variation on these trips could get picked up by the rider effects. In Appendix Table 14 we first remove all riders with fewer than 5 trips in the data set and then recompute the fixed effects. We make no other restrictions on drivers or (time cross location) pairs. Results are very similar to before.

Effect	Asheville, NC	Bloomington, IN	Boston	Chicago	Salt Lake City	San Francisco
Driver	0.192	0.066	0.107	0.108	0.212	0.114
Rider	0.627	0.181	0.296	0.301	0.612	0.335
Time x Location	0.232	0.057	0.047	0.054	0.229	0.066
Residual	0.568	0.191	0.290	0.307	0.505	0.330

Appendix Table 14: Standard deviation of estimated effects for the different sources of tip variation. Before estimating the fixed effects we remove all riders with fewer than 5 trips in the sample.

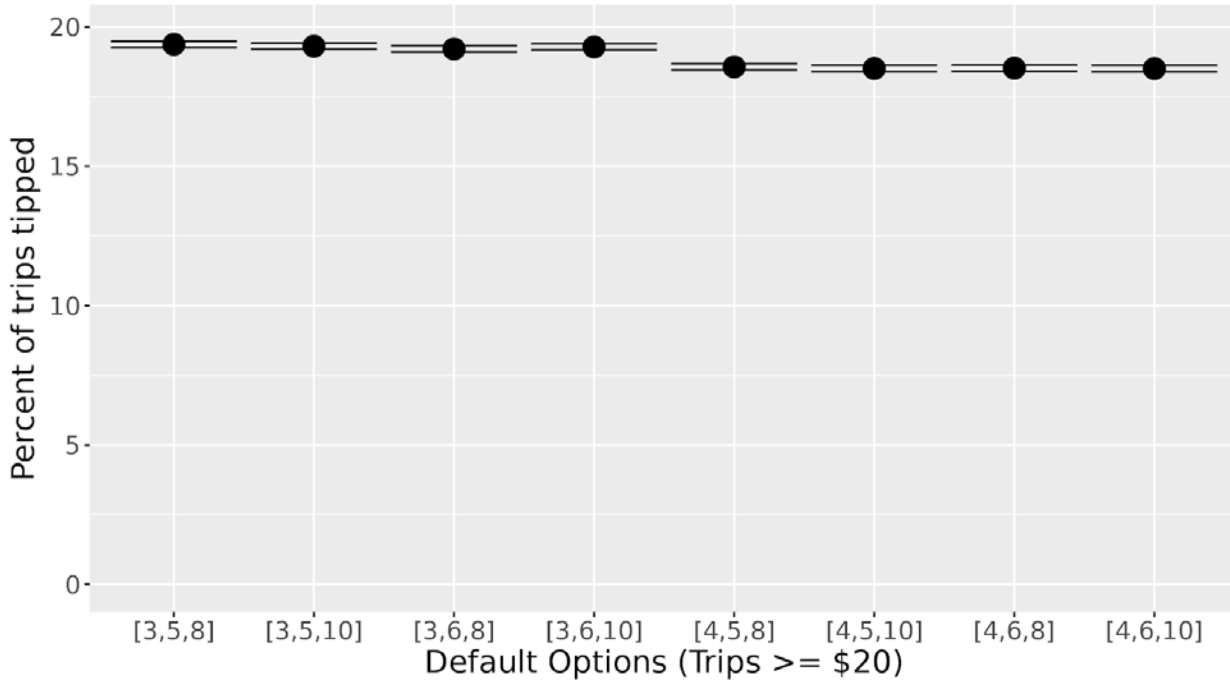
5. Results from Experiment 2 (Variable Preset Group)

We consider participants who received a different preset for trips under $20 and over $20. The preset options were randomized and so the group was placed into 64 different groups (eight options for trips under $20 times eight options for trips $20 and over). For ease of analysis we split the data into trips eligible for the lower presets and trips eligible for the upper presets.

5.1 Percent of Trips Tipped

For trips under $20, riders in experiment 2 were shown one of the presets from experiment 1. Results for the effect of presets on these trips largely mimic those seen in experiment 1 and therefore are not reported. We turn our focus to trips that cost $20 or more and the new presets shown to riders in this experiment.

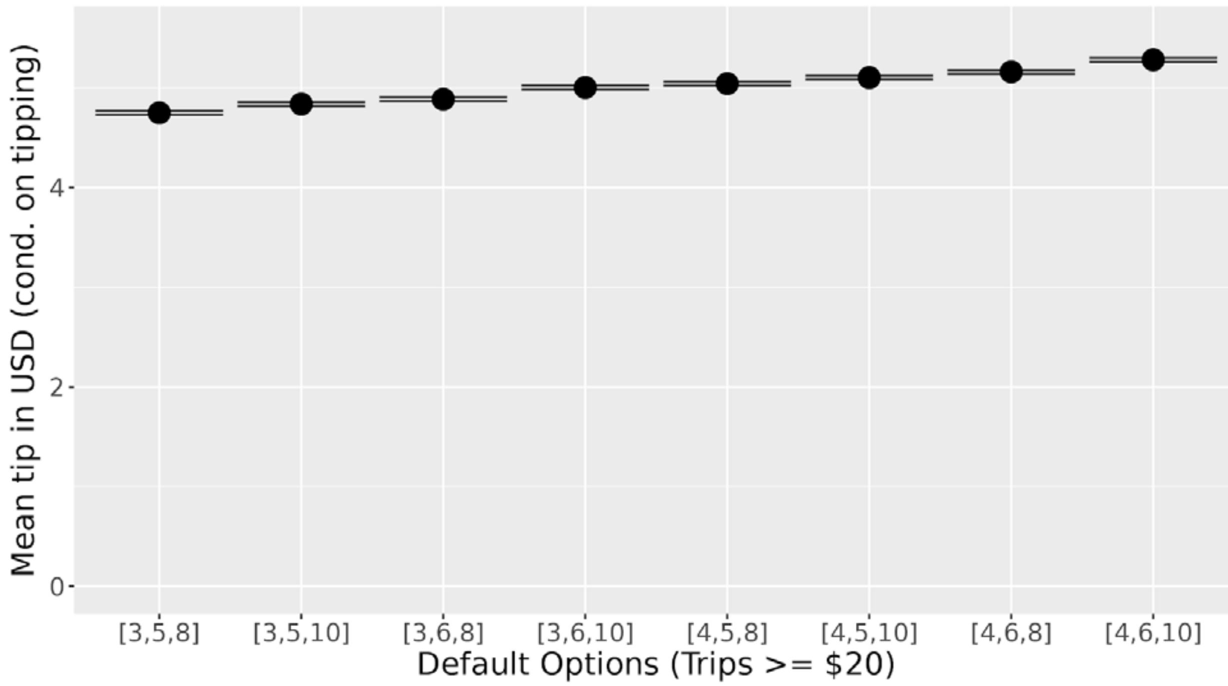
In Appendix Figure 8 we see that presets that begin with a $4 option instead of a $3 option decrease the probability that a trip is tipped. This result mimics that of shifting from presets starting with $2 instead of $1 in experiment 1. The highest probability of tipping occurs with the [$3, $5, $8] preset at 19.4% while the lowest probability is associated with the [$4, $6, $10] preset at 18.5% of trips tipped. For reference, similarly priced trips in experiment 1 were tipped 20.5% of the time when preset [$1, $3, $5] was shown, and least likely, 19.7% of the time, with the preset [$2, $4, $6].



Appendix Figure 8: Probability of being tipped as a function of presets for experiment 2.

5.2 Mean tip conditional on being tipped

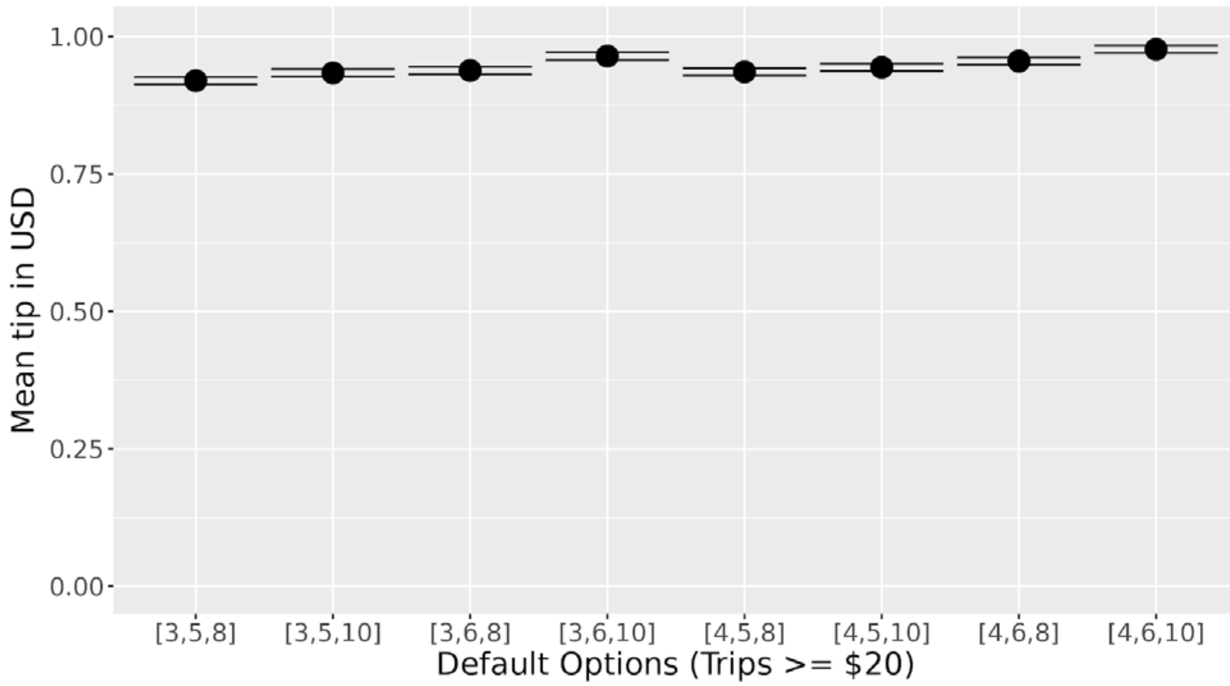
Similar to experiment 1, we see that different presets lead to different amounts tipped conditional on a trip being tipped. Results are depicted in Appendix Figure 9. For this experiment, [$4, $6, $10] yields a $5.28 average tip, while [$3, $5, $8] only yields $4.75 on average, a difference of $0.53. In the previous experiment, the difference was smaller for similarly priced trips, where the highest mean tip amount, $4.31, occurred with the [$2, $4, $6] while the lowest was $4.03 for the [$1, $3, $5] preset, a difference of only $0.28.



Appendix Figure 9: Mean amount tipped conditional on tipping as a function of presets for experiment 2.

5.3 Mean tip

Again, similar to experiment 1, we see that the effect on the probability of tipping and the mean tip conditional on tipping counteract each other and lead to much more muted effects on the average tip on a given trip including $0 when the rider did not tip. Results are depicted in Appendix Figure 10. The highest mean tip of $0.977 is associated with preset [$4, $6, $10] while the lowest mean tip of $0.920 is associated with the preset [$3, $5 $8], a difference of only $0.057. For trips over $20 in experiment 1, the highest mean tip was $0.849, for preset [$2, $4, $6], while the lowest was $0.811, for preset [$2, $3, $5], a difference of $0.038. Although there is little difference within either experiment, the difference across all eight presets from both experiments ends up being $0.166, suggesting presets have some effect for more expensive trips with a wider range of price points for the presets.



Appendix Figure 10: Mean amount tipped as a function of presets for experiment 2.

As in experiment 1, the presets were designed to be able to estimate the marginal impact of changing a single option in the preset. In Appendix Table 15 we see that when the first option is set at $4 instead of $3 the probability of being tipped decreases by 77 basis points (4.0%), while changes in the other two positions did not statistically significantly affect the probability of a trip being tipped. When subsetting to trips that were tipped we see that the first option being $4 instead of $3 increased tips $0.278 (5.8%), the second option being $6 instead of $5 increased tips $0.15 (3.2%), and the third option being $10 instead of $8 increased tips $0.10 (2.0%). Lastly, we see that these effects offset each other such that a higher first option increases tips by 1.4¢ (1.5%), the higher second option increases tips by 2.5¢ (2.8%), and the higher third option increases tips by 1.8¢ (1.9%) on average across all trips that cost $20 and above.

	Effect on Percent of Trips Tipped	Effect on Mean Tip in USD (conditional on tipping)	Effect on Mean Tip in USD
	(1)	(2)	(3)
First Option \$4 Instead of \$3	-0.0077^{***}	0.2776^{***}	0.0139^{***}
	(0.0006)	(0.0086)	(0.0033)
Second Option \$6 Instead of \$5	-0.0006	0.1496^{***}	0.0254^{***}
	(0.0006)	(0.0086)	(0.0033)
Third Option \$10 Instead of \$8	-0.0002	0.0971^{***}	0.0176^{***}
	(0.0006)	(0.0086)	(0.0033)
Constant	0.1933^{***}	4.7456^{***}	0.9179^{***}
	(0.0006)	(0.0085)	(0.0033)
Observations	3,696,356	698,962	3,696,356
R^2	0.0001	0.0031	0.0001
Adjusted R^2	0.0001	0.0031	0.0001
Residual Std. Error	0.3916 (df = 3696352)	2.9407 (df = 698958)	2.3412 (df = 3696352)
Note:			*p<0.1; **p<0.05; ***p<0.01

Appendix Table 15: Marginal effect of changes in preset options for trips $20 and over in experiment 2.