Online Appendix to:

Defaults and Decisions: Choice Architecture and Consumer Opt-Out

Kwabena Donkor

Stanford Graduate School of Business

May 14, 2026

Contents

A Survey Instruments

The survey was administered via Prolific and approved by the Stanford Institutional Review Board. Complete instruments are reproduced below.

Pre-screening and Round 1 (Injunctive Norms)

A pre-screening page confirmed informed consent, prior experience with NYC Yellow taxis, and borough of residence. Participants who passed screening proceeded immediately to the elicitation task.

Each participant was presented with five fare scenarios randomly drawn from a set of eleven values ($5, $7.50, $10, $15, $20, $25, $30, $40, $50, $75, $100). For each scenario, the prompt read: "Suppose you take a ride in an NYC Yellow taxicab, what do you think is the appropriate amount to tip the driver if the taxi fare is **$X**?" Participants responded by adjusting an interactive slider that displayed the fare amount, selected tip percentage, and corresponding dollar tip simultaneously. The slider ranged from 0% to 100% with no pre-selected default position. Participants were paid $0.50 for completing this round.

Pre-screening Tipping Survey

DESCRIPTION: We invite you to participate in a research study to help understand tipping behavior in taxicabs.

TIME INVOLVEMENT: Your participation will take approximately 2 minutes.

PAYMENTS: You will be paid $0.50 for completing this survey.

PRIVACY AND CONFIDENTIALITY: Your privacy will be maintained during the research and in all published and written data resulting from the study.

CONTACT INFORMATION:
Questions: If you have any questions, concerns, or complaints about this research, its procedures, risks, and benefits, contact the Protocol Director, Kwabena Donkor, at 650-497-7452 and donkor@stanford.edu.

Independent Contact: If you are not satisfied with how this study is being conducted, or if you have any concerns, complaints, or general questions about the research or your rights as a participant, please contact the Stanford Institutional Review Board (IRB) to speak to someone independent of the research team at (650)-723-2480 or toll-free at 1-866-680-2906, or email at irbnonmed@stanford.edu. You can also write to the Stanford IRB, Stanford University, 1705 El Camino Real, Palo Alto, CA 94306.

Print or save a copy of this page for your records.

If you agree to participate in this research, please click **Yes** to continue.

- ○ Yes
- ○ No

Next

Have you ever taken a trip in an NYC Yellow taxicab?

- ○ Yes
- ○ No

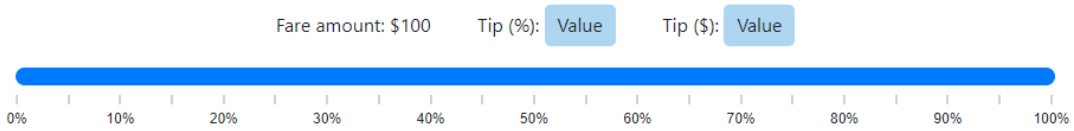
Where do you live in New York: choose one of the NYC boroughs or "Outside NYC"?

- ○ Manhattan
- ○ Brooklyn
- ○ Queens
- ○ Bronx
- ○ Staten Island
- ○ Outside NYC

Next

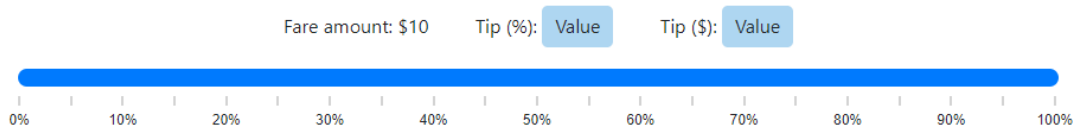
Suppose you take a ride in an NYC Yellow taxicab, what do you think is the appropriate amount to tip the driver if **the taxi fare is $100?**

Please click or drag the slider below to the tip (%) of your choice.



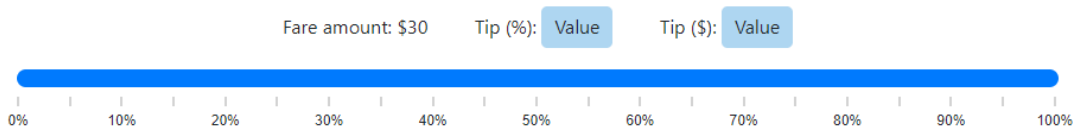
Suppose you take a ride in an NYC Yellow taxicab, what do you think is the appropriate amount to tip the driver if **the taxi fare is $10?**

Please click or drag the slider below to the tip (%) of your choice.



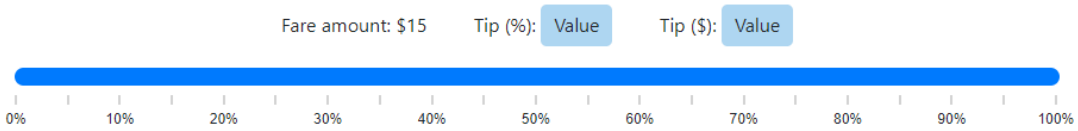
Suppose you take a ride in an NYC Yellow taxicab, what do you think is the appropriate amount to tip the driver if **the taxi fare is $30?**

Please click or drag the slider below to the tip (%) of your choice.



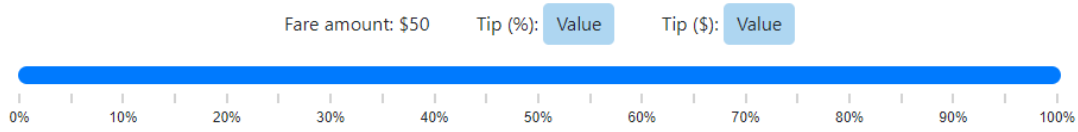
Suppose you take a ride in an NYC Yellow taxicab, what do you think is the appropriate amount to tip the driver if **the taxi fare is $15?**

Please click or drag the slider below to the tip (%) of your choice.



Suppose you take a ride in an NYC Yellow taxicab, what do you think is the appropriate amount to tip the driver if **the taxi fare is $50?**

Please click or drag the slider below to the tip (%) of your choice.



Thank you for taking the time to take this survey. Click finish to be redirected to Prolific.

Finish

Round 2 (Descriptive Norms)

Two weeks after Round 1, returning participants completed a follow-up survey with identical question wording and the same slider interface. The key difference was the incentive structure. Before beginning, participants read the following instructions:

> "Win bonus of up to $3 for next 10 short answer questions. For each of the 10 questions following this page, you will be paid a **$0.30** bonus if your guess is within one percentage point of the average answer across all survey participants. [...] You will be paid **$0.10** if your guess is above one percentage point and within two percentage points of the average answer. There is no payment if your answer exceeds two percentage points of the average."

This tiered bonus structure makes the task an incentive-compatible coordination game: the payoff-maximizing strategy is to report one's best guess of the population average tip rate. Round 2 presented eleven fare scenarios ($5, $7.50, $10, $15, $20, $25, $30, $40, $50, $75, $100). Base payment was $1, with up to $4 in bonuses.

Tipping Survey 2

DESCRIPTION: We invite you to participate in a research study to help understand tipping behavior in taxicabs.

TIME INVOLVEMENT: Your participation will take approximately 3-5 minutes.

PAYMENTS: You will be paid **$1** for completing this survey and an additional bonus of up to **$4** depending on your answers. The bonus will be paid to your Prolific account a week after completing the survey.

PRIVACY AND CONFIDENTIALITY: Your privacy will be maintained during the research and in all published and written data resulting from the study.

CONTACT INFORMATION:
Questions: If you have any questions, concerns, or complaints about this research, its procedures, risks, and benefits, contact the Protocol Director, Kwabena Donkor, at 650-497-7452 and donkor@stanford.edu.

Independent Contact: If you are not satisfied with how this study is being conducted, or if you have any concerns, complaints, or general questions about the research or your rights as a participant, please contact the Stanford Institutional Review Board (IRB) to speak to someone independent of the research team at (650)-723-2480 or toll-free at 1-866-680-2906, or email at irbnonmed@stanford.edu. You can also write to the Stanford IRB, Stanford University, 1705 El Camino Real, Palo Alto, CA 94306.

Print or save a copy of this page for your records.

If you agree to participate in this research, please click **Yes** to continue.

- ○ Yes
- ○ No

Next

Win bonus of up to $3 for next 10 short answer questions.

For each of the 10 questions following this page, you will be paid a **$0.30** bonus if your guess is within one percentage point of the average answer across all survey participants.

E.g., if the average answer is X% and your guess is between X±1%, you get **$0.30.**

You will be paid **$0.10** if your guess is above one percentage point and within two percentage points of the average answer.

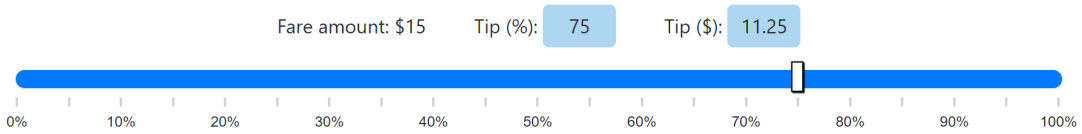
There is no payment if your answer exceeds two percentage points of the average answer.

Next

Q1: Win a bonus of up to $0.30.

Suppose you take a ride in an NYC Yellow taxicab, what is the appropriate amount to tip the driver if the taxi fare is **$15?**

Please click or drag the slider below to the tip (%) of your choice.

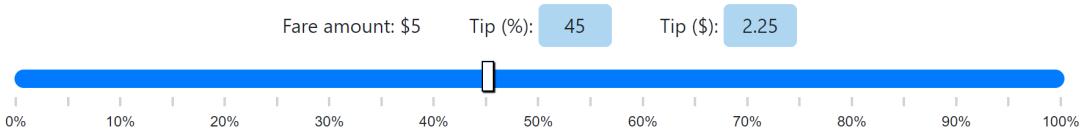


Next

Q2: Win a bonus of up to $0.30.

Suppose you take a ride in an NYC Yellow taxicab, what is the appropriate amount to tip the driver if the taxi fare is **$5?**

Please click or drag the slider below to the tip (%) of your choice.

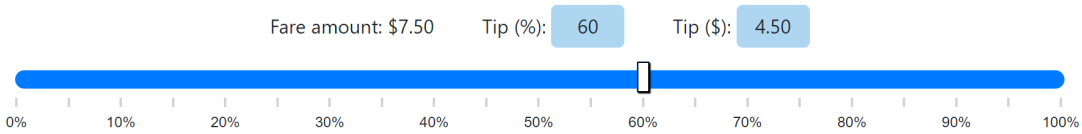


Next

Q3: Win a bonus of up to $0.30.

Suppose you take a ride in an NYC Yellow taxicab, what is the appropriate amount to tip the driver if the taxi fare is **$7.50?**

Please click or drag the slider below to the tip (%) of your choice.

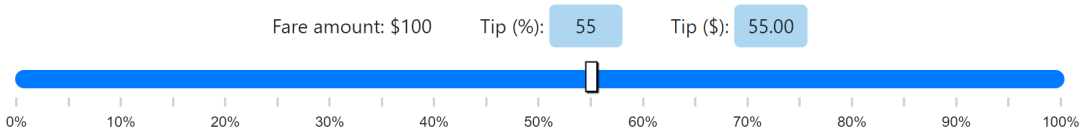


Next

Q4: Win a bonus of up to $0.30.

Suppose you take a ride in an NYC Yellow taxicab, what is the appropriate amount to tip the driver if the taxi fare is **$100?**

Please click or drag the slider below to the tip (%) of your choice.

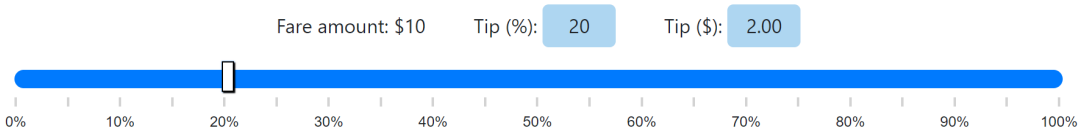


Next

Q5: Win a bonus of up to $0.30.

Suppose you take a ride in an NYC Yellow taxicab, what is the appropriate amount to tip the driver if the taxi fare is **$10?**

Please click or drag the slider below to the tip (%) of your choice.

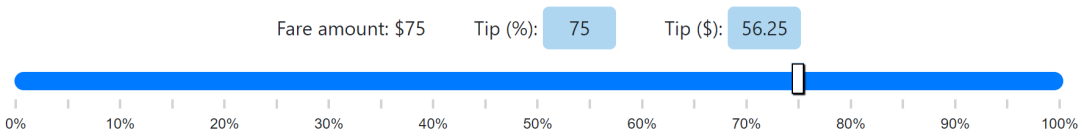


Next

Q6: Win a bonus of up to $0.30.

Suppose you take a ride in an NYC Yellow taxicab, what is the appropriate amount to tip the driver if the taxi fare is **$75?**

Please click or drag the slider below to the tip (%) of your choice.

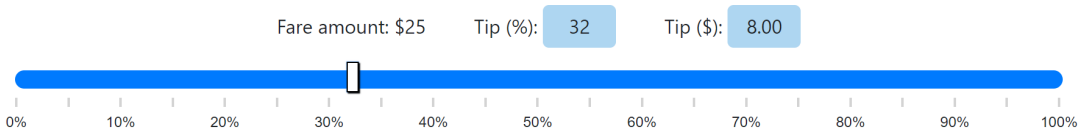


Next

Q7: Win a bonus of up to $0.30.

Suppose you take a ride in an NYC Yellow taxicab, what is the appropriate amount to tip the driver if the taxi fare is **$25?**

Please click or drag the slider below to the tip (%) of your choice.

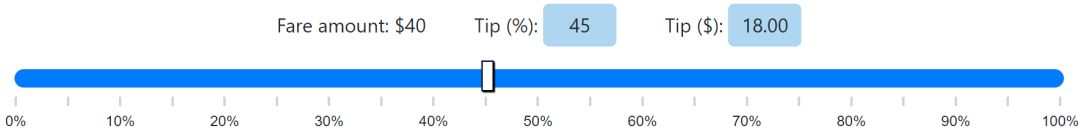


Next

Q8: Win a bonus of up to $0.30.

Suppose you take a ride in an NYC Yellow taxicab, what is the appropriate amount to tip the driver if the taxi fare is **$40?**

Please click or drag the slider below to the tip (%) of your choice.

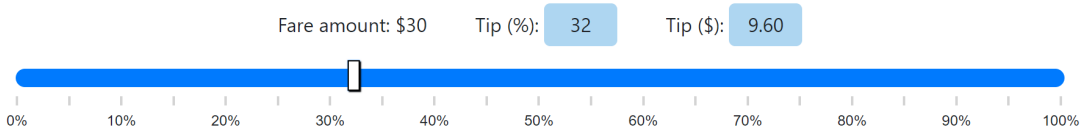


Next

Q9: Win a bonus of up to $0.30.

Suppose you take a ride in an NYC Yellow taxicab, what is the appropriate amount to tip the driver if the taxi fare is **$30?**

Please click or drag the slider below to the tip (%) of your choice.

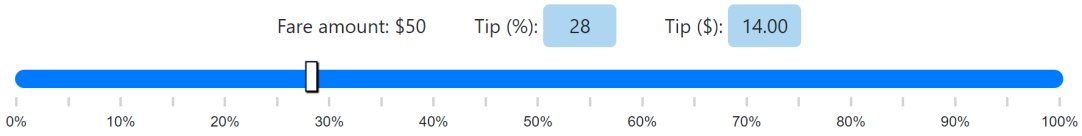


Next

Q10: Win a bonus of up to $0.30.

Suppose you take a ride in an NYC Yellow taxicab, what is the appropriate amount to tip the driver if the taxi fare is **$50?**

Please click or drag the slider below to the tip (%) of your choice.

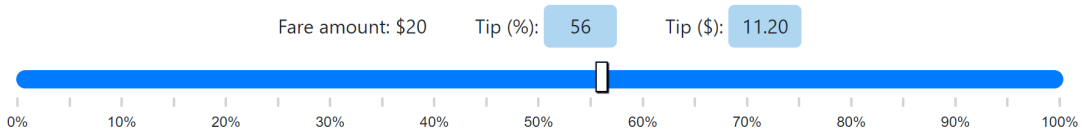


Next

Q11: Win a bonus of up to $0.30.

Suppose you take a ride in an NYC Yellow taxicab, what is the appropriate amount to tip the driver if the taxi fare is **$20?**

Please click or drag the slider below to the tip (%) of your choice.



Next

Win bonus of up to $1

You will be paid a **$0.20** bonus if your answer for each condition listed below is within one percentage point of the average answer across all survey participants.
E.g., if the average answer is X% for condition Y and your answer is between X±1%, you get **$0.20**.
You will be paid **$0.10** if your guess is above one percentage point and within two percentage points of the average.
There is no payment if your answer exceeds two percentage points of the average answer.

Q12: For each condition listed below, what do you think is the appropriate tip (% of taxi fare) for the average NYC Yellow taxi ride?

Input a tip percent between 0% and 100%.

Condition	**Tip (% of taxi fare)**
Sunny	
Rain	
Snow	
Rush Hour	
Travelling with a coworker	

Q13: How much do you think an average NYC Yellow taxi fare cost?

Thank you for taking the time to take this survey. Click finish to be redirected to Prolific.

Finish

Survey Summary Statistics

Table A1 reports respondent demographics and elicited-rate summary statistics for both rounds.

Table A1: Survey Summary Statistics: Elicited Ideal Tip Rates

	Injunctive Norms (Round 1)	Descriptive Norms (Round 2)
N Participants	607	402
N Responses	3035	4422
Mean Tip Rate (%)	17.6	18.4
Median Tip Rate (%)	18.0	20.0
Std. Dev. (%)	9.9	8.9

Notes: This table reports summary statistics from a two-round stated-preference survey administered via Prolific to participants prescreened for New York residency and familiarity with NYC taxi travel. Round 1 (Injunctive Norms) elicited personal beliefs about appropriate tip rates via the prompt "What tip rate is appropriate for a \$X fare?" across five randomly assigned fare scenarios (\$5–\$100) per participant. Round 2 (Descriptive Norms), administered two weeks later to a subset of Round 1 participants, elicited beliefs about typical peer behavior via an incentive-compatible coordination game in which participants were rewarded for matching the group average response across eleven fare scenarios. The difference between mean injunctive (17.8%) and descriptive (18.4%) norms indicates that participants believe others tip slightly more than what they personally consider appropriate. Neither distribution exhibits the pronounced bunching at default rates (20%, 25%, 30%) observed in transaction data (Table 1), consistent with survey-elicited ideals reflecting unconstrained preferences rather than menu-influenced choices.

B Model Derivations

Best Default Cutoffs

For a passenger selecting default d_j, utility is $u_{ij}(T_i) = -F_i d_j - (\theta/2)[T_i - d_j]_+^2$. To determine which default is best, compare adjacent options d_1 versus d_2.

Case 1: $T_i \leq d_1$. Both deviation costs are zero, so $u_{i1} - u_{i2} = F_i(d_2 - d_1) > 0$. The cheaper default d_1 dominates.

Case 2: $d_1 < T_i \leq d_2$. The deviation cost for d_1 is $(\theta/2)(T_i - d_1)^2$; for d_2 it is zero. Default d_1 is preferred if the monetary saving exceeds the deviation cost:

$$F_i(d_2 - d_1) \geq \frac{\theta}{2}(T_i - d_1)^2 \quad \Leftrightarrow \quad T_i \leq d_1 + \sqrt{\frac{2F_i(d_2 - d_1)}{\theta}}.$$

Case 3: $T_i > d_2$. Both deviation costs are active. Default d_1 is preferred if

$$T_i \leq \frac{d_1 + d_2}{2} + \frac{F_i}{\theta}.$$

For the parameter values in this application, the Case 2 cutoff is the binding constraint: it falls near or just above d_2, while the Case 3 cutoff $(d_1 + d_2)/2 + F_i/\theta$ is strictly larger for all relevant parameter values. The effective cutoff between d_j and d_{j+1} is therefore:

$$\gamma_j = d_j + \sqrt{\frac{2F_i(d_{j+1} - d_j)}{\theta}}.$$

These cutoffs partition the ideal-tip-rate space into three regions: $A_1 = [0, \gamma_1)$, $A_2 = [\gamma_1, \gamma_2)$, and $A_3 = [\gamma_2, \infty)$, where d_j is the best default for $T_i \in A_j$.

Opt-Out Value and Gain

Let $V_i(T_i)$ denote the value of the optimal custom tip before subtracting the opt-out cost:

$$V_i(T_i) = \begin{cases} -\dfrac{\theta}{2}T_i^2, & \text{if } T_i < F_i/\theta \quad \text{(corner: } t_i^* = 0\text{)}, \\ -F_i T_i + \dfrac{F_i^2}{2\theta}, & \text{if } T_i \geq F_i/\theta \quad \text{(interior: } t_i^* = T_i - F_i/\theta\text{)}. \end{cases}$$

Let $U_i^D(T_i) = \max_j u_{ij}(T_i)$ denote utility from the best default. The gain from opting out is $\Delta(T_i; F_i, \theta) = [V_i(T_i) - U_i^D(T_i)]_+$, and the passenger opts out if and only if $c_i < \Delta_i$. Under the exponential specification $c_i \sim \text{Exp}(\lambda)$, this yields the opt-out probability $\Pr(\text{opt out} \mid T_i, F_i) = 1 - e^{-\lambda \Delta_i(T_i)}$.

Counterfactual Welfare Expressions

Expected revenue under menu D is:

$$R(D) = \mathbb{E}\left[\sum_j d_j \cdot F_i \cdot \mathbf{1}\{T_i \in A_j, c_i \geq \Delta_i\} + t_i^* \cdot F_i \cdot \mathbf{1}\{c_i < \Delta_i\}\right].$$

Expected consumer welfare is:

$$C(D) = \mathbb{E}\left[\max_j u_{ij}(T_i) \cdot \mathbf{1}\{c_i \geq \Delta_i\} + (V_i(T_i) - c_i) \cdot \mathbf{1}\{c_i < \Delta_i\}\right].$$

Social welfare is $W(D) = R(D) + C(D)$, treating tips as a transfer from consumers to drivers rather than deadweight loss.

C Additional Tables and Figures

Table A2: Within-Person Variation in Ideal Tip Rates by Fare Level

	Injunctive Norms (1)	Descriptive Norms (2)
Fare ($)	-0.0371^{***}	-0.0218^{***}
	(4.5×10^{-3})	(3.4×10^{-3})
Person Fixed Effects	Yes	Yes
Observations	3,035	4,422
Participants	607	402
Fare Scenarios per Person	5	11
Fare Range	$5–$100	$5–$100

Notes: This table reports within-person regressions of the stated ideal tip rate (in percentage points) on the fare amount shown in the survey scenario. Both columns include person fixed effects. Column (1) uses responses from the injunctive norm elicitation (Round 1). Column (2) uses responses from the descriptive norm elicitation (Round 2). Standard errors clustered at the participant level are in parentheses. *** $p < 0.01$.

Table A3: Predicted vs. Observed Choice Shares

	Descriptive Norms	Injunctive Norms
	Predicted [Observed]	
20% default	50.7 [48.5]	47.5 [48.5]
25% default	4.4 [11.2]	3.6 [11.2]
30% default	5.1 [4.7]	6.7 [4.7]
Opt-out	39.8 [35.6]	42.2 [35.6]

Notes: Predicted choice shares (with observed shares in brackets) from the structural model under each norm specification. The opt-out category includes zero tips and positive custom tips. Predictions use the full 2014 sample (8,569,381 trips) and fare-conditional survey-elicited distributions of ideal tip rates.

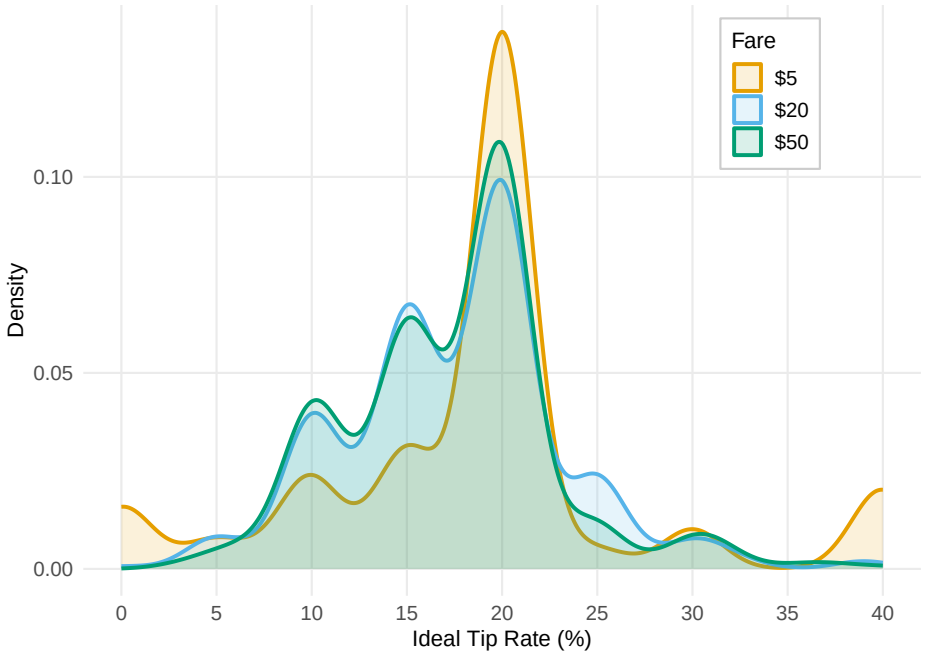
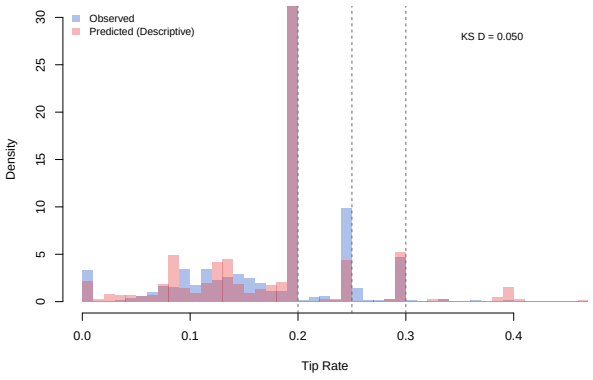


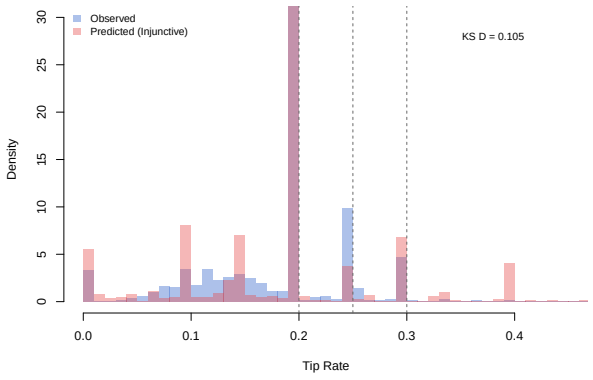
Figure A1: Distribution of Ideal Tip Rates at Representative Fare Levels

Notes: Kernel density estimates of elicited ideal tip rates from the descriptive round at three fare levels: \$5 (orange), \$20 (blue), and \$50 (green). The distribution shifts leftward and compresses as fare increases, illustrating the within-person variation documented in Figure 3 and Table A2. Bandwidth selected using Sheather-Jones plug-in estimator with adjustment factor of 1.5. X-axis truncated at 40% for visualization; 6.7%, 0.7%, and 1.2% of responses at the \$5, \$20, and \$50 fare levels exceed this threshold.

Figure A2: Model Fit: Observed vs. Predicted Tip Rate Distributions



(a) Descriptive Norms ($\hat{\theta} = 749$, $\hat{\lambda} = 1.462$; KS D = 0.050)



(b) Injunctive Norms ($\hat{\theta} = 43{,}665$, $\hat{\lambda} = 1.716$; KS D = 0.105)

Notes: Comparison of observed (blue) and model-predicted (coral) tip rate distributions under both survey specifications. Dashed vertical lines mark default rates (20%, 25%, 30%). Both specifications reproduce the spikes at defaults. The descriptive specification (Panel A) fits substantially better: KS D = 0.050 versus 0.105. X-axis truncated at 45% for visualization; 0.4% of trips have tip rates above this threshold.

Table A4: Summary Statistics: Cross-Vendor Comparison (2010, Fares ≥$15)

	CMT (1)	VeriFone (2)
Trip Characteristics		
Fare amount ($)	19.62	19.65
	(4.86)	(4.90)
Trip duration (min)	24.60	25.04
	(8.24)	(8.06)
Trip distance (mi)	6.70	6.67
	(2.48)	(2.55)
Tipping Behavior		
Default menu (%)	15-20-25	20-25-30
Tip amount ($)	3.21	3.29
	(1.63)	(1.82)
Tip rate (%)	15.69	16.33
	(6.53)	(7.71)
Opt-out rate	0.38	0.50
Share at 15%	0.29	0.03
Share at 20%	0.27	0.37
Share at 25%	0.08	0.10
Share at 30%	0.00	0.03
Share at 0%	0.03	0.06
Observations	2,780,727	2,926,197

Notes: Summary statistics for 2010 NYC Yellow taxi trips with fares ≥$15, when CMT displayed {15%, 20%, 25%} defaults and VeriFone displayed {20%, 25%, 30%}. Sample restrictions identical to Table 1. This cross-vendor comparison provides a contemporaneous test of default effects, holding time and market conditions fixed.

Table A5: Cross-Vendor Anchoring Test (2010, Fare ≥ $15)

	Choice Shares (%)		Custom Tips		Mass Within 1pp of (%)			
Condition	Default	Custom	Mean	Median	15%	20%	25%	30%
Cross-Vendor (2010, Fare ≥ $15)								
CMT (15-20-25%)	64.0	32.8	11.98	10.75	3.6	1.4	0.5	0.8
Verifone (20-25-30%)	51.0	42.8	12.33	11.63	12.0	1.3	0.3	0.3
Difference	-13.0	+10.0	+0.35	+0.88	+8.4	-0.0	-0.2	-0.5

Notes: Statistics computed on 2010 trips with fares ≥$15, comparing CMT (defaults 15-20-25%, N=2,780,727) and VeriFone (defaults 20-25-30%, N=2,926,197). See Table 3 for variable definitions. The cross-vendor pattern is consistent with the within-vendor result: mass at 15% is higher among Verifone custom tippers, where 15% is not a default button, consistent with passengers typing an unchanged ideal rate.

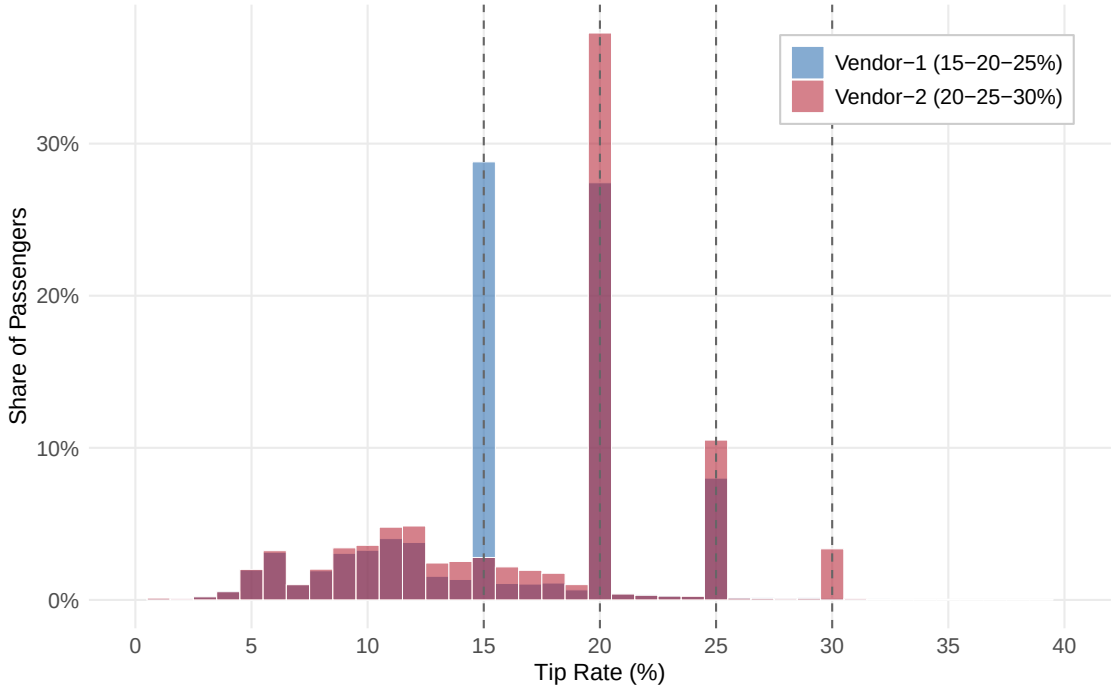


Figure A3: Cross-Vendor Tip Rate Comparison (2010)

Notes: Histogram of tip-rate distributions across payment vendors during 2010, when CMT and VeriFone simultaneously displayed different default menus. CMT displayed {15%, 20%, 25%} defaults (blue, N = 2.9 million trips); VeriFone displayed {20%, 25%, 30%} (red, N = 2.8 million trips). Sample restricted to fares $\geq$ \$15, standard-rate, non-airport, credit-card trips. Vertical dashed lines mark all four default rates. Bin width is 1 percentage point, matching Figure 4. X-axis truncated at 40% for visualization; 0.3% of CMT and 0.3% of VeriFone trips exceed this threshold.

D Mechanism Tests

Estimation Procedure

Each menu condition is estimated independently using the likelihood function from Section 4.1, yielding its own (θ, λ) pair. Stability across conditions supports the maintained assumption that defaults operate through frictions rather than by reshaping preferences.

Within-Vendor Menu Change (CMT)

On February 9, 2011, CMT payment terminals changed their default tip options from {15%, 20%, 25%} to {20%, 25%, 30%}. I estimate the model independently on pre- and post-change CMT transactions. Table 4 in the main text reports the results.

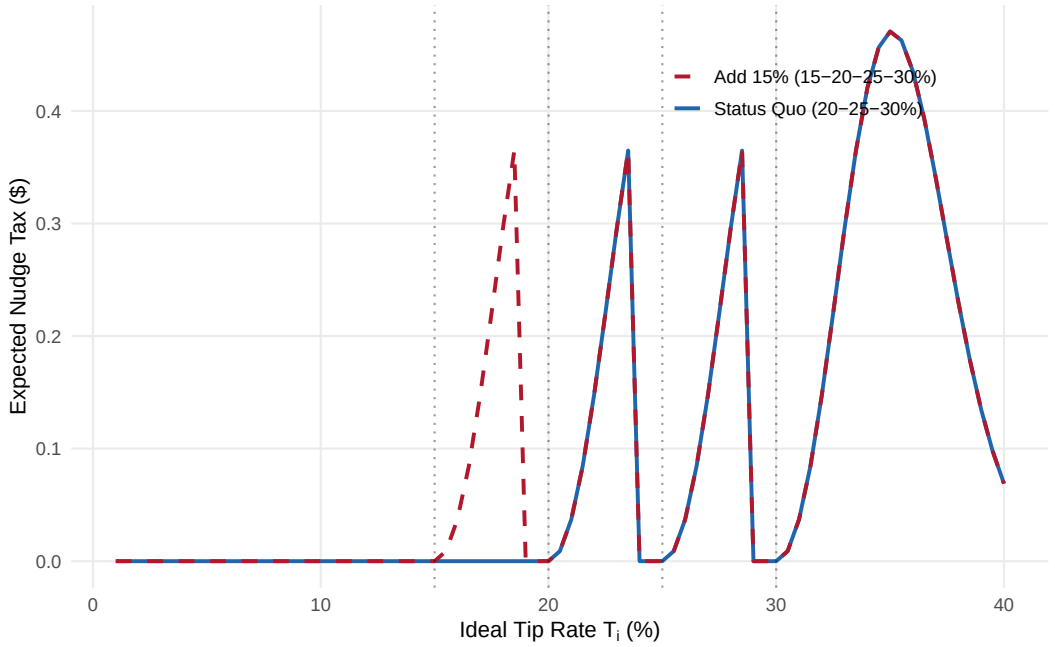


Figure A4: Nudge Tax by Ideal Tip Rate

Notes: Expected nudge tax (the deviation cost $(\theta/2)(T_i - d_j)_+^2$ weighted by the probability of staying at default d_j) as a function of the consumer's ideal tip rate T_i, evaluated at the population-weighted mean fare using baseline estimates ($\hat{\theta} = 749$, $\hat{\lambda} = 1.46$). Solid blue: status quo menu {20%, 25%, 30%}. Dashed red: menu with 15% added {15%, 20%, 25%, 30%}. Vertical dotted lines mark default rates. Adding a 15% option reduces the nudge tax for consumers with T_i between 15% and 24%.

E Inference and Robustness

Two-Step Variance Decomposition

The likelihood in Section 4.1 is a two-step M-estimator: the fare-conditional preference distribution $\widehat{\omega}(T \mid F)$ is estimated in a first stage from a survey of 402 respondents, and the structural parameters (θ, λ) are estimated in a second stage from 8.6 million transactions, treating $\widehat{\omega}$ as a plug-in input. Treating $\widehat{\omega}$ as known data understates standard errors. I follow the standard variance decomposition for two-step M-estimators (Newey and McFadden, 1994; Murphy and Topel, 1985):

$$\operatorname{Var}(\widehat{\boldsymbol{\eta}}) \;=\; \underbrace{\operatorname{Var}(\widehat{\boldsymbol{\eta}} \,|\, \widehat{\omega} \text{ fixed})}_{\text{second-stage (sandwich)}} \;+\; \underbrace{\operatorname{Var}(\widehat{\boldsymbol{\eta}}(\omega))\,\big|_{\omega \sim G}}_{\text{first-stage}}, \tag{1}$$

where $\widehat{\boldsymbol{\eta}} = (\log \widehat{\theta}, \log \widehat{\lambda})$ and G is the sampling distribution of the survey. The decomposition is valid because the transaction sample (TLC administrative records) and the survey sample (Prolific respondents) are operationally independent draws: who appears in one had no influence on who ap-

pears in the other. Both samples target the same NYC tipping population; what the decomposition captures is the finite-sample uncertainty in each, propagated through the estimator. I estimate the two components separately and combine them in quadrature.

Second-Stage Component: Cluster-Robust Sandwich Variance

The second-stage component is the sandwich variance for M-estimators, holding $\widehat{\omega}$ fixed:

$$\hat{V}_{\text{2nd}} = \hat{H}^{-1}\,\hat{B}\,\hat{H}^{-1},$$

where $\hat{H} = \sum_{i=1}^{N} \nabla^2 \ell_i(\widehat{\boldsymbol{\eta}})$ is the Hessian of the total negative log-likelihood evaluated at the MLE, and $\hat{B} = \frac{G}{G-1}\sum_{g=1}^{G} S_g S_g'$ is the clustered outer-product-of-scores ("meat") matrix with $S_g = \sum_{i \in g} \nabla \ell_i(\widehat{\boldsymbol{\eta}})$ summed within cluster g. The $G/(G-1)$ factor is the finite-cluster correction analogous to HC1 in linear models. Standard errors are computed in log-parameter space and converted to the original scale via the delta method: $\text{SE}(\theta) = \widehat{\theta} \cdot \text{SE}(\log\theta)$. Observations are clustered at the date level (365 clusters in the 2014 baseline; 404 pre-change and 326 post-change in the menu-change analysis). Score vectors are computed by central differences on the per-observation log-likelihood, with step size 10^{-5} in log-parameter space.

First-Stage Component: Survey Bootstrap

The first-stage component is estimated by a nonparametric bootstrap of the survey. For each of $B = 500$ replications, I (i) resample 402 descriptive-survey respondents with replacement and rebuild $\widehat{\omega}^{(b)}(T \mid F)$ at each of the eleven fare levels, and (ii) re-estimate (θ, λ) on the full transaction data using $\widehat{\omega}^{(b)}$. The standard deviation of the bootstrap distribution of $\widehat{\boldsymbol{\eta}}$ is the first-stage standard error. Each replication uses the same resampled $\widehat{\omega}^{(b)}$ across all three samples (2014 baseline, pre-change CMT, post-change CMT), so cross-sample contrasts inherit paired bootstrap standard errors. To improve robustness, BFGS is initialized at the original MLE plus a small Gaussian jitter ($\sigma = 0.02$ in log-parameter space) and uses tightened convergence tolerance (`factr` $= 10^4$). Replications in which the optimum hits the parameter boundary or the optimizer fails are excluded from the SE calculation; valid-replication counts are 459 of 500 (baseline), 429 of 500 (pre-change), and 457 of 500 (post-change).

Decomposition of Reported Standard Errors

Table A6 reports the per-source decomposition for each parameter and sample. The first-stage component dominates by orders of magnitude for the deviation-cost parameter θ. With 8.6 million transactions, the second-stage sandwich variance is small in absolute terms; the first-stage bootstrap is dominant because the survey is a small sample (402 respondents) and the structural parameters are responsive to the shape of $\widehat{\omega}$.

Table A6: Decomposition of Standard Errors: Sandwich and First-Stage Bootstrap

Sample / Parameter	Point Estimate	Sandwich SE	First-Stage Bootstrap SE	Total SE	First-Stage Share (%)
Baseline (2014, Descriptive)					
θ	748.5651	0.14	135.00	135.00	100.0%
λ	1.4621	4.4×10^{-3}	0.13	0.13	99.9%
Pre-Change CMT (2010–2011, before Feb 9)					
θ	1,008.4088	0.03	403.00	403.00	100.0%
λ	3.6895	0.02	0.44	0.44	99.8%
Post-Change CMT (2010–2011, on/after Feb 9)					
θ	935.1635	4.7×10^{-3}	224.00	224.00	100.0%
λ	4.1367	0.02	0.42	0.42	99.9%

Notes: Decomposition of total standard errors into the two components of the Newey–McFadden two-step variance formula (Newey and McFadden, 1994; Murphy and Topel, 1985): $\mathrm{Var}(\hat{\theta}) = \mathrm{Var}(\hat{\theta} \mid \hat{\omega} \text{ fixed}) + \mathrm{Var}(\hat{\theta}(\omega))$. The first term ("Sandwich SE") is the cluster-robust sandwich SE with date-level clustering, holding the survey-derived $\hat{\omega}(T \mid F)$ fixed. The second term ("First-Stage Bootstrap SE") is the standard deviation of the bootstrap distribution of $\hat{\theta}$ obtained by resampling the 402 descriptive-survey respondents with replacement (B = 500) and re-estimating (θ, λ) on the full transaction data with each resampled $\hat{\omega}^{(b)}$. The two components are combined in quadrature: $\mathrm{SE}_{\mathrm{total}} = \sqrt{\mathrm{SE}^2_{\mathrm{sandwich}} + \mathrm{SE}^2_{\mathrm{boot}}}$. The final column reports the share of total variance attributable to the first-stage component, which dominates by orders of magnitude for the deviation-cost parameter θ. Bootstrap reps where the optimum hit the parameter boundary or where L-BFGS-B aborted with low iteration count were excluded; valid-rep counts were 459 of 500 (baseline), 429 of 500 (pre-change), and 457 of 500 (post-change).

Robustness to Clustering Level

Within the second-stage component, the sandwich variance can be computed at alternative clustering levels. To assess sensitivity to the choice, the baseline sandwich standard errors are recomputed with observations clustered at the week level (52 clusters). Week-level clustering is more conservative, allowing for correlation across days within the same week.

Week-level standard errors are approximately twice as large as date-level standard errors (Table A7), consistent with modest positive correlation across days within a week. Both clustering

Table A7: Sandwich Standard Errors: Date vs. Week Clustering (Descriptive)

	Date (365 clusters)	Week (52 clusters)
SE(θ)	0.14	0.32
SE(λ)	4.4×10^{-3}	8.3×10^{-3}

Notes: Sandwich standard errors for the descriptive specification ($\widehat{\theta} = 748.6$, $\widehat{\lambda} = 1.4621$), holding $\widehat{\omega}$ fixed. Both clustering levels apply the $G/(G-1)$ finite-cluster correction. These values are the second-stage component of the variance decomposition only; the total standard error is dominated by the first-stage bootstrap component reported in Table A6.

choices imply that the second-stage component is small in absolute terms; the dominant contribution to total SE comes from the first-stage bootstrap.

Computational Approach

With 8.6 million observations, numerical integration over the empirical T_i distribution for each transaction is computationally intensive. A sufficient-statistic structure resolves this: for passengers who select a default or tip zero, the choice probability depends only on fare, so per-observation log-likelihoods can be aggregated to the fare level. Only the approximately 37% of transactions involving a custom interior tip require individual evaluation. This fare-aggregation reduces the effective dimension of the summation by several orders of magnitude, enabling maximum likelihood estimation on the full 8.6 million-transaction sample without subsampling.

Sample Size Robustness

To verify that the full-sample estimates are stable, the baseline model is re-estimated on stratified random subsamples of 100,000, 500,000, and 1,000,000 transactions drawn from the 2014 sample. The opt-out cost rate λ is stable across all subsample sizes, varying by less than 2% from the full-sample estimate. The deviation cost parameter θ, however, is systematically approximately 10% higher on subsamples ($\hat{\theta} \approx 822$–837) than on the full sample ($\hat{\theta} = 749$), reflecting sensitivity to the empirical fare distribution weights that enter the aggregated likelihood. This pattern motivates the full-sample estimation approach: the fare-aggregation technique described above eliminates the need for subsampling and avoids the attendant bias.

References

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. In *Handbook of Econometrics*, volume 4, pages 2111–2245. Elsevier.

Murphy, K. M. and Topel, R. H. (1985). Estimation and inference in two-step econometric models. *Journal of Business & Economic Statistics*, 3(4):370–379.