Gender Bias in Auto Repair Pricing

Sophia Ayele*
sophia.ayele@berkeley.edu

Vaibhav Beohar*
vbeohar@berkeley.edu

Ying Chen*
chentim@berkeley.edu

Chris Donaton*
chrisd@ischool.berkeley.edu

05/09/2021

Abstract

This study attempts to replicate previous evidence of auto repair price discrimination against women. It also explores hypotheses about the underlying mechanisms, specifically that women customers are less respected and perceived to be less knowledgeable about auto repair. We use an audit study to compare price quotes received over the phone by female and male callers for the same repair on the same make and model of car. We also examine the interaction between a caller's gender and accent. Contrary to previous studies, we do not find evidence of price discrimination against women. In fact, we find that, on average, women and people that speak English as a second language (ESL) are quoted lower prices. We also find evidence of a gender and ESL affinity effect, with male ESL callers receiving lower quotes, on average, from male ESL shop representatives. Lastly, we observe that women are questioned about their knowledge ten percent more often than men, but this does not appear to translate into higher price quotes.

Introduction

Industries without fixed pricing, such as the auto repair industry, provide an ideal testing ground for studies on price discrimination. In these industries, sellers can dynamically adjust pricing on a transaction-by-transaction basis. While pricing decisions are impacted by market dynamics, sellers in these industries also have the opportunity to adjust pricing based on other factors, such as consumer characteristics. This opportunity results from the information asymmetry between sellers and consumers. Since prices are not typically posted, the consumer must rely on the seller to quote them a fair market price.

There has been considerable research into price discrimination with several studies finding evidence of discrimination based on consumer characteristics including as gender, race, age, sexual orientation, and disability status. However, the evidence is mixed and few studies have attempted to examine the underlying mechanisms. A 1995 field experiment by Ayres and Siegelman sent buyers to car dealerships to negotiate to buy a new car using a script.[1] After over 300 paired audits, they found that white men received lower starting prices and were able to negotiate lower prices than black or female buyers. In a 2012 study, Gneezy, List, and Price used a series of field experiments, encompassing over 3,000 transactions in several different markets, to examine price discrimination based on gender, age, sexual orientation, race, and disability.[2] They found evidence of discimination in each market.

In an observational study, Chen, Hu, Szulga, and Zhou examined price differentials from over 145,000 transactions in the Chinese automobile market between 2007 to 2009. They found that men paid lower prices on average than women.[3] Chen et al. also found that local men received an additional discount that was not present for local women. However, another study of auto repair pricing in North America failed to find price differences for local versus out of town customers. In a 2012 field experiment, Schneider used undercover visits to auto repair shops to test whether mechanics were more likely to provide a thorough and accurate diagnosis to local customers versus customers that

*These authors contributed equally to this work.

signaled that they were moving out of town.[4] Schneider did not find a significant difference, however, he did conclude that under and over diagnosis are common.

In a 1996 observational study, Goldberg used transaction price data from the US Consumer Expenditure Survey to analyze variation in car dealer discounts.[5] Goldberg found that the variation was explained by the car model, market-specific factors, and purchase transaction characteristics (first-time purchase, trade-in, and financing through the dealer) versus consumer characteristics. In a 2013 field experiment involving the taxi market in Lima, Peru, Castillo, Petrie, Torero, and Vesterlund obtained data from over 1,000 negotiations between male taxi drivers and 12 trained passengers. They found that men received worse bargaining outcomes than women.[6] In a follow-up study, where customers signalled their willingness to pay before bargaining, the gender differences disappeared.

The study that most closely relates to ours is a 2013 field experiment by Busse, Israeli, and Zettelmeyer that examined whether perceptions about women's knowledge of market prices drives gender discrimination.[7] Busse et al. leveraged AutoMD's call center to have undercover callers obtain over 4,600 price quotes from auto repair shops. The researchers randomly assigned shops to a female or male caller and into three different experimental conditions: (1) the uninformed caller that signalled unfamiliarity with market rates, (2) the uninformed caller that mentioned a price above the market rate, and (3) the informed caller that mentioned a price at the market rate. They found that when customers signalled unfamiliarity with market rates, women were quoted $23 more, on average, than men. However, the gender difference disappeared in the other two experimental conditions. Interestingly, they also found that shops were more likely to provide a discount when asked by a female versus male caller.

In this study, we expand the existing knowledge around gender-based price descriminiation by attempting to replicate the causal effect observed by Busse et al. Due to the project timeline and resource constraints, our experiment is limited to the uninformed caller condition where Busse et al. observed a statistically significant difference in price. It is well known that a gender wage gap exists in the United States.[8] If female consumers are, indeed, charged more in markets without fixed pricing, this would further compound the economic disparities that women face.

Research Question

This study investigates the causal effect of gender on auto repair pricing. Specifically, our research question is: Does a caller's gender influence auto repair price quotes? We also examine the interaction between caller gender and English as a second language (ESL) status.

Hypotheses

We attempt to test several hypotheses, primarily drawn from the existing literature. The first three relate to price discrimination and the second two explore biases as mediators (also known as mechanisms) underlying the relationship between gender and pricing. In this section we outline these hypotheses and the outcome measures used to evaluate them.

Price Discrimination Hypotheses

Three primary hypothesese about price discrimination were made for the experiment:

- **Gender-Based**: Female callers receive higher auto repair quotes.
- **Gender-ESL Interaction**: Female and ESL callers receive even higher auto repair quotes.

- **Gender-ESL Affinity Interaction**: Callers receive lower auto repair quotes when their gender and/or ESL status is the same as the person providing the quote.

Our hypotheses on price discrimination are measured by one primary metric:

- **Price Quoted**: The price quoted in US dollars by the auto repair shop. If this was provided as a price range, the median point of the range was used for analysis and comparison.

Mediation Hypotheses

Two secondary hypotheses were made in addition to price discrimination:

- **Knowledge**: Female and/or ESL callers are believed to be less knowledgeable about auto repair and are subject to higher pricing.
- **Respect**: Female and/or ESL callers are less respected by auto repair shops.

Our mediation hypotheses are measured by three metrics:

- **Script Knowledge**: An indicator for whether the caller was asked questions about the vehicle symptoms or history. We assessed that questioning the caller's diagnosis of the problem was an indication that the shop representative doubted the caller's knowledge.
- **Quote Breakdown Provided**: An indicator for whether the caller was provided with an unsolicited price breakdown. We assessed this as a measure of professionalism and respect for the customer.
- **Attrition**: An indicator for whether a shop refused to provide a quote or failed to call back with a quote. We also assessed this as a measure of professionalism and respect for the customer.

Two other secondary indicators were originally considered, but ruled out due to the low number of positive cases. The first was an indicator for whether a shop representative refused to provide a quote (12 positive cases, also included in attrition). This could be store policy, but could also indicate that the representative wants the customer to come to the shop so that they can find, or invent, other things that are wrong with the car. The other was an indicator for whether the shop representative suggested additional repairs or asked about the caller's budget (18 positive cases). This could indicate that the representative is looking for an opportunity to charge the customer more.

Experiment Design

Experiment Overview

This audit study used a post treatment design, measuring outcomes only after successfully exposing a subject to either treatment or control.

Treatment Group	R X O
Control Group	R - O

Table 1: Experiment Design

The quote to replace a radiator was the primary outcome measurement of interest.

Y_0 = Quote received by a Male and/or non-ESL caller
Y_1 = Quote received by a Female and/or ESL caller

Response to treatment was measured using three between-subjects comparisons to measure the average treatment effect. These comparisons are represented here as they appear in the linear regression models in the Results section of this paper. Each comparison built on the previous in order to control for additional factors that may have affected model estimates. The first comparison was based solely on the treatment assignment hypothesis that female callers are likely to receive a higher quote than male callers.

$$H_0: \frac{outcome|caller_{female}}{N|caller_{female}} = \frac{outcome|caller_{male}}{N|caller_{male}}$$
$$H_1: \frac{outcome|caller_{female}}{N|caller_{female}} \neq \frac{outcome|caller_{male}}{N|caller_{male}}$$

The next comparison controlled for the ESL status of the caller as an interaction term to account for interactions between treatment groups.

$$H_0: \frac{outcome|caller_{female,ESL}}{N|caller_{female,ESL}} = \frac{outcome|caller_{male,!ESL}}{N|caller_{male!ESL}}$$
$$H_1: \frac{outcome|caller_{female,ESL}}{N|caller_{female,ESL}} \neq \frac{outcome|caller_{male!ESL}}{N|caller_{male!ESL}}$$

Finally, the last comparison controlled for the gender and ESL status of the auto shop representative as it compared to the caller's to capture any potential heterogeneous treatment effects.

$$H_0: \frac{outcome|(caller_{female,ESL}, representative_{female,ESL})}{N|caller_{female,ESL}, representative_{female,ESL}} = \frac{outcome|caller_{male,!ESL}}{N|caller_{male,!ESL}}$$
$$H_1: \frac{outcome|(caller_{female,ESL}, representative_{female,ESL})}{N|caller_{female,ESL}, representative_{female,ESL}} \neq \frac{outcome|caller_{male,!ESL}}{N|caller_{male,!ESL}}$$

The affinity indicators were created from covariates collected during the experiment and interacted with treatment.

Data Scraping

To compile a dataset of auto repair shop locations and contact information, data was scraped from Yelp on February 27, 2021. The team created a python script to search for auto repair shops in the 25 most populous cities in Washington State (population greater than 50,000) and to scrape data from the first four search result pages for each city This scraping procedure returned 1,000 auto repair shops. After removing duplicate records, the dataset consisted of 731 unique auto repair shops.

The dataset was manually scrubbed to identify whether a shop was a chain and to remove invalid records. A total of 76 shops that either did not provide radiator services, did not service Toyotas, were mobile auto repair services, or had gone out of business were removed. An additional 13 shops with incomplete information were also removed. Lastly, 6 shops had inadvertently been scrapped from over the border in Oregon which were also removed.
After deduplication, data cleaning, and pilot data allocation, there were 615 auto repair shop records remaining in the dataset. Each record was enriched with county location and an indicator for whether the shop was a chain, defined as having more than one business location.

The compiled data set included data from 11 of the 39 counties in Washington State, comprising 83 chain shops and 532 shops with only one location. The map below shows the distribution of shops around the state, with chain shops displayed as red dots and shops with only one location displayed as green dots.

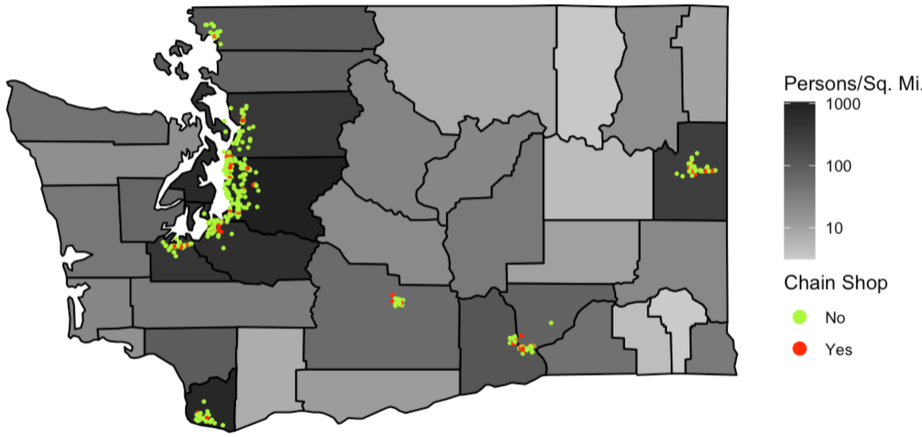

Figure 2: County Population Density Overlaid With Auto Shop Locations Across the Washington State

Pilot Study & Power Analysis

In early March 2021, a small pilot study was conducted using a random sample of 30 shops from the full dataset. In addition to the power analysis, the pilot study helped assess the viability of the scraped data, identify additional parameters for removing shops from the list of potential subjects, expand the call script, and inform the data collection rubric.

The pilot study helped identify several areas of improvement for the call script and data collection rubric. Following the pilot, the call script was expanded to include commonly-requested vehicle information, such as four door, gas operated, and automatic transmission, and reorganized to present the most-often requested information first. The data collection rubric was also amended to include additional reasons why a shop could not provide a quote, the amount of additional information that a shop requested, whether they asked about the customer's budget or tried to add additional services, and how many call attempts were made to each shop.

To estimate the sample size needed to detect a statistically significant effect, we conducted a power analysis using pilot data. As mentioned previously, our callers were volunteers and had limited time to devote to calling. Pilot calls took approximately 2 minutes each, therefore, we determined that it would only be feasible for callers to make between 25 and 50 calls each, a total sample size of between 200 and 400 calls. As shown by Figure 3, the pilot data indicated that, for a power of level of 0.8 and significance level of 0.05, we would need an effective size of $77.04 for a sample of 200 calls and an effect size of $54.34 for 400 calls.

We observed a high level of attrition during the pilot study, with only 60% of calls successfully obtaining a quote. The high attrition was mainly due to shops not picking up the phone, failing to call back with a quote, or refusing to provide a quote over the phone because they required the caller to bring in the vehicle. Since we expected a similarly high level of attrition in our actual experiment, we set the goal of collecting 200 quotes (100 each for treatment and control) by contacting 400 shops. As mentioned above, our power analysis indicated that this sample size would require an effect size of $77.04. However, note that as a result of the unexpected high attrition rate, our power analysis is performed based on an insufficient sample size (i.e. smaller than 30). The results presented in this section should be interpreted with caution.

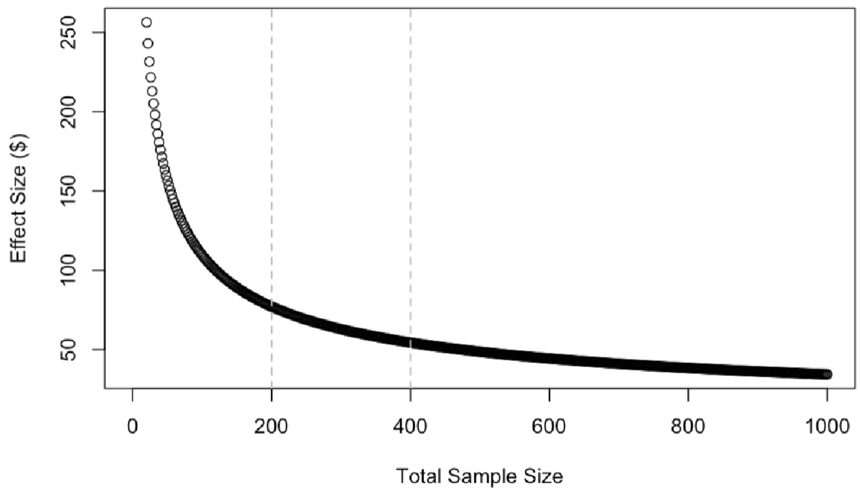

Figure 3: Power Analysis - Sample Size vs. Effect Size

Block Randomization

We hypothesized that chain shops would have more standardized pricing and charge a premium over smaller shops. We also hypothesized that different counties would have different factors impacting their pricing, such as the county-specific tax rate, minimum wage rate, and cost of living. In order to account for these fixed effects, randomization was blocked on chain and county to ensure equal representation of each level in the study.

Although the goal of the experiment was to enroll at least 100 subjects in each treatment and control, all 615 shops were randomized and assigned to one of the eight callers on the experimental team. The experimental team would then proceed down the list attempting to accumulate the target number of subjects in each group while accounting for attrition and non-compliance. This method posed the risk of low sampling-variance within the blocks, but the logistical constraints of the experiment forced the acceptance of this risk.

Treatment Assignment

Once randomized blocks were created based on county and chain, they were then randomly assigned into treatment and control groups. The makeup of the experimental group was balanced with respect to gender, but included more male than female ESL speakers.

	Treatment (Female Caller)	**Control (Male Caller)**
ESL Caller	Female, ESL	Male, ESL
Native English Caller	Female, Native English	Male, Native English

Table 4: Gender/ESL Treatment Matrix

This imbalance in the group provided a natural oversampling of the specific subset of the population of interest for the second hypothesis related to the effect of ESL status on potential outcomes.

Experiment Protocol

The team researched the most popular name for each gender in the Washington state area, then selected a female and male name that didn't necessarily convey an ethnic or racial identity. For consistency, callers identified themselves with this same name and used a script to play the role of the uninformed caller (Appendix A). In the original Busse et al. study, auto repair shops were asked to provide a quote for a 6-cylinder 2003 Toyota Camry LE. This make and model was selected because the AutoMD database showed customers frequently requested repair information for this car. They chose a radiator repair because it can easily be diagnosed by a customer and chose the year 2003 because radiator leaks are common among 10-year old vehicles. To replicate this study, the same make, model, and service was used but the year of the vehicle was adjusted to account for the age of the study. This experiment requested a quote to replace a 2012 Toyota Camry LE.

Each caller started with the first shop in their respective list of potential subjects. Callers used *67 to block their caller ID in order to prevent any potential bias associated with an out-of-town phone number. If there was no answer, it was logged and only one additional attempt to contact would be made, if necessary; No voicemails were left by any caller.

If contact was made with a shop representative, the shop was enrolled in the experiment, the interaction was logged in a data collection form (Appendix B), and all callers' communication was dictated by the experimental script. If a shop requested to call back with a quote, a callback number was provided. The same callback number was used by all callers. Although no additional attempts were made to contact this shop, any quote left by the shop via voicemail was collected into the study. The following information categories were collected:

- **Call Timing**: day and time that the call occurred
- **Shop Demographics**: gender and ESL status of the shop representative
- **Price Quoted**: whether a quote was provided, the price quoted and whether it was provided as a fixed price or range, and whether it included a cost breakdown
- **Call Characteristics**: whether the shop representative asked about the caller's budget, tried to offer additional repairs, or questioned the caller's ability to diagnose the problem

Results

Data Flow

The experiment was conducted over a one week period in March of 2021. In total, contact was attempted with 416 of the 615 randomized shops and 207 price quotes were obtained (100 treatment and 107 control). The response rate, defined as obtaining a quote from a shop, was 50%, with a slightly higher response rate among female callers (53%) and slightly lower response rate among male callers (47%).

	Overall, N = 416	Control	Treatment
Quote Obtained			
No	209 (50%)	120 (53%)	89 (47%)
Yes	207 (50%)	107 (47%)	100 (53%)

Table 5: Obtained Quotes Distribution

Figure 6, shows the entire data flow process. Of the 209 shops that we did not obtain quotes from, 158 were excluded from the analysis because they were never contacted, had gone out of business, or did not provide the service. Another 51 were considered attrition. These shops were contacted but either refused to provide a quote over the phone, claimed that they were unable to provide a quote, or failed to call back with a quote. See Appendix C for a detailed breakdown of exclusion and attrition categories. Only three shops claimed that they were unable to provide a quote, two because the mechanic was unavailable and one because their computer was down.

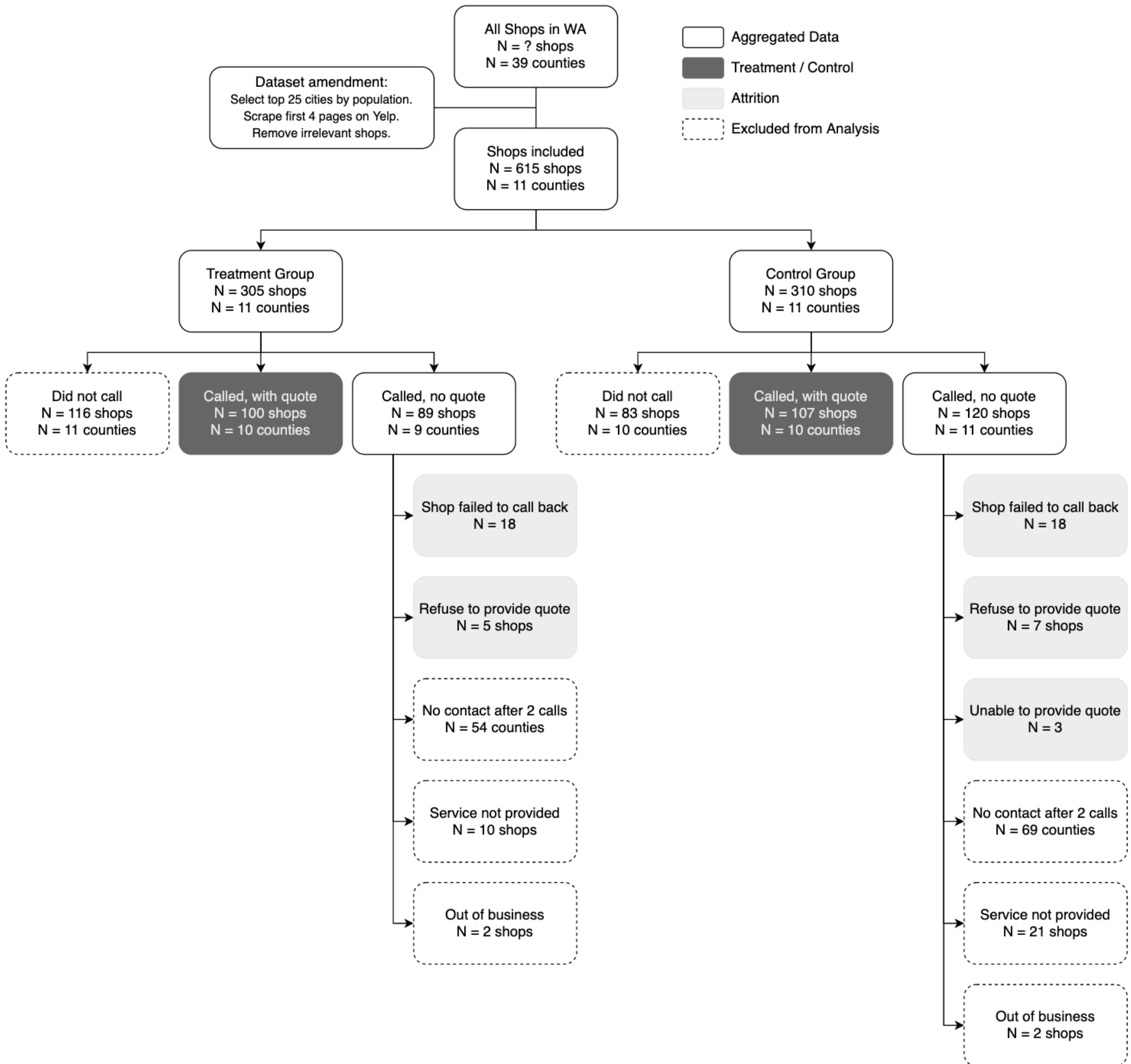

Figure 6: Data Flow Chart

Randomization Check

Table 7 shows the distribution of potential subjects post block randomization - that is, all shops scraped from Yelp. You can see the small sample size in many of the blocks, which will be discussed further in the Regression Models section. After the block randomization, each caller was assigned a list of roughly 80 shops to contact. As mentioned previously, we elected to scrape far more shops than our objective of 200 subjects (100 each in treatment and control). This offered the experiment more robustness to withstand loss of potential subjects due to shop closures or the inability to contact a shop for other reasons. Due to a minor administrative issue one caller's block assignments were switched. While this did not impact the overall treatment and control randomization, it does mean that the blocks are slightly less balanced than originally intended.

Characteristic	Quote Obtained	Attrition
chain	35 (17%)	7 (14%)
county		
benton	11 (5.3%)	5 (9.8%)
clark	10 (4.8%)	6 (12%)
franklin	6 (2.9%)	0 (0%)
island	0 (0%)	1 (2.0%)
king	79 (38%)	16 (31%)
pierce	19 (9.2%)	1 (2.0%)
snohomish	17 (8.2%)	7 (14%)
spokane	25 (12%)	5 (9.8%)
thurston	17 (8.2%)	2 (3.9%)
whatcom	14 (6.8%)	2 (3.9%)
yakima	9 (4.3%)	6 (12%)

Table 7: Pretreat Distribution by County and Chain

Figure 8 shows a covariate balance check using the absolute standardized mean differences between treatment and control. The first section of covariates, highlighted grey, are covariates that we had control over. Of note, our ESL caller group was unbalanced due to the makeup of our experimental team. Ultimately, we elected to maintain this natural oversampling given our hypothesis about the effect of ESL status on potential outcomes.

The second section, highlighted mid-grey, are covariates that we could have controlled. However, due to logistical constraints such as geographic separation and availability of our callers, we were unable to reliably control for day and time factors. See (Appendix E) for more information about variation in day and time of calls made by each caller. Finally, the last section, highlighted light-grey are covariates that we did not have control over. While we did not have control over these covariates, shop representative gender and ESL status were well balanced between the groups. Although the affinity covariates, which indicated like-gendered and like-ESL status between the caller and the shop representative, are unbalanced, this was expected given both the breakdown within our group (i.e. more ESL than non-ESL) and the fact that male shop representatives outnumbered female representatives nearly 4 to 1 across subjects.

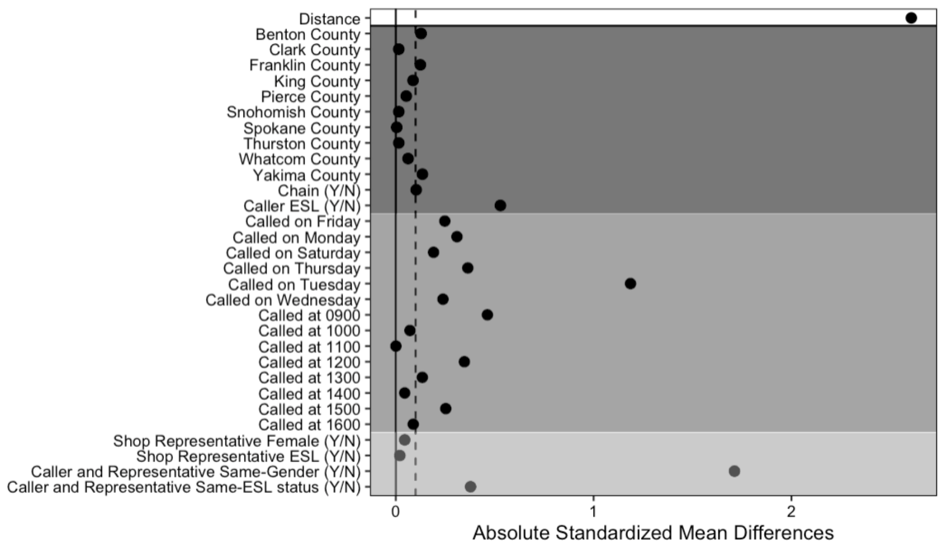

Figure 8: Covariate Balance Check

Attrition Analysis

As mentioned previously, we excluded shops that we were unable to contact, that had gone out of business, or that did not provide the service. We were unable to contact some shops because they never answered the phone or because they did not accept unidentified numbers. We attempted to verify all claims that a shop did not provide the service. The majority of these shops were tire shops or auto body shops that we had overlooked during our data cleaning process. After removing these shops, 51 (20%) of the remaining shops were considered attrition. These were shops that we had contacted, but that did not provide a quote for one of three reasons: they refused to provide a quote over the phone, they did not call back with a quote, or they claimed that they were unable to provide a quote.

To assess the impact of attrition on our results we conducted an attrition balance check (Appendix D). The attrition balance check reveals a slightly lower percentage of chain shops among the attrition group than shops that we obtained a quote from (14% versus 17%, respectively). Among counties, we see slightly higher attrition in Benton, Clark, Snohomish, and Yakima counties, but it should be noted that there are very small group sizes across most counties for both the attrition and quote obtained categories. In addition, a two-sample t-test fails to reject the null hypothesis of no statistically significant difference in means between attrition in the treatment and control groups (Appendix D).

As a final precaution, we used extreme value bounds to test the reliability of our results under different potential outcome scenarios (Appendix D). To evaluate the effects of Extreme Value Bounds, we imputed attrited outcomes using a bracketing technique which included the minimum and maximum extreme values, as well as the 25% and 75% quantiles. With the exception of one coefficient in the maximum extreme range in Model 3, there was no sign change on treatment effect. This bolsters confidence in our concluded direction of treatment effect, although leaves to question the magnitude of the effect.

Outcome Variables

Primary Outcome

Price Quote: The primary outcome measure for this analysis is the price quote obtained from each shop. As mentioned previously, 34 (16%) shops provided a price range rather than a fixed price and these outcomes were encoded as the median of the range for the purpose of analysis and comparison. Figure 9 below shows the raw distribution of quotes by treatment and control groups. Overall, quotes ranged from a low of $330 to a high of $1,018. Surprisingly, the control group mean is $40 higher than the treatment group mean ($688 versus $648).

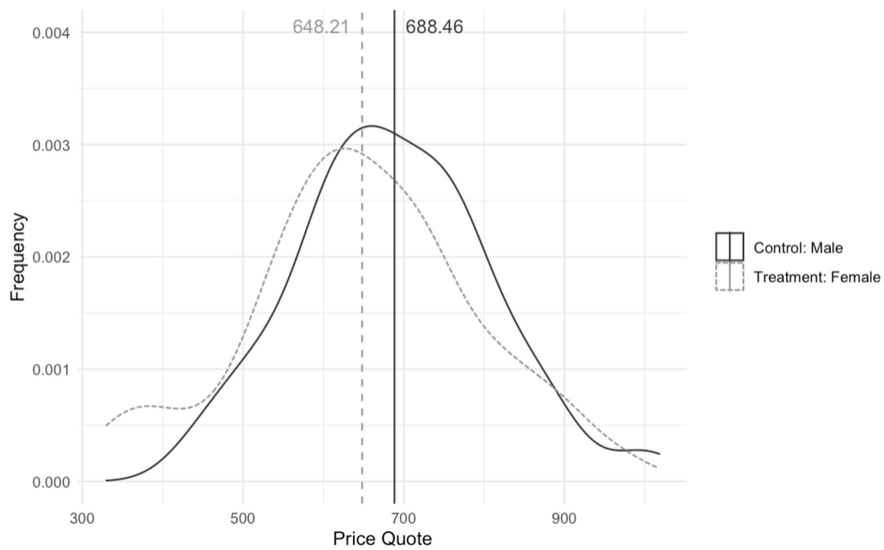

Figure 9: Outcome Distribution by Gender

Secondary Outcomes

Knowledge: As an indicator of the perceived level of knowledge of a caller, we measured whether a caller was asked to provide more information about how they knew the radiator needed to be replaced. This was measured by whether or not a caller used sections D and/or E from the caller script (Appendix A). Figure 10 depicts the measured values. Female callers (treatment) were asked for this information in 29% of calls versus 19% for male callers (control).

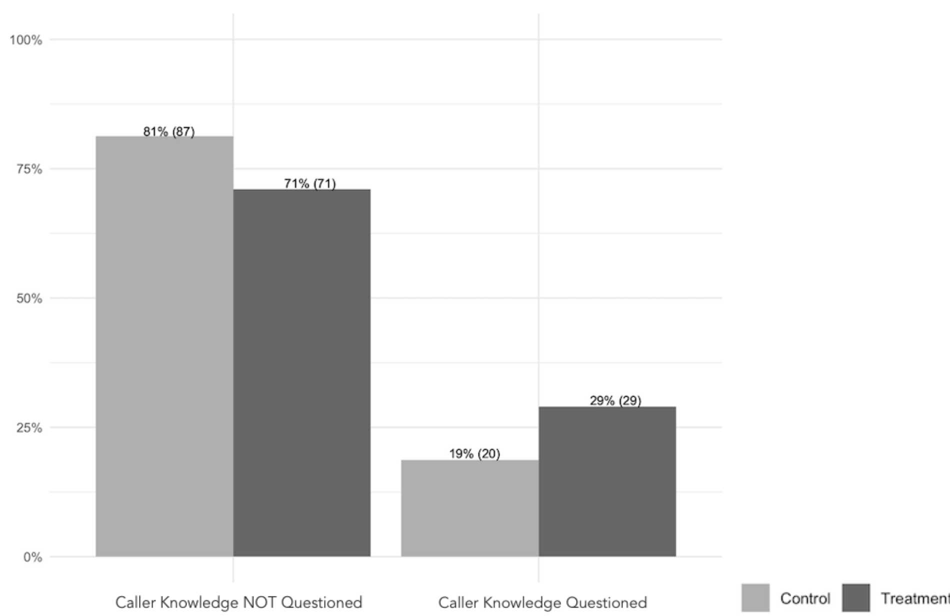

Figure 10: Distribution of Knowledge Measurement by Gender. “Yes” means the

Respect: As indicators of the level of professionalism and respect provided to the customer, we measured attrition, defined as a contacted shop failing to provide a quote or follow-up with a quote by voicemail, and whether a quote came with an unsolicited cost breakdown. Figure 11 depicts the breakdown between control and treatment.

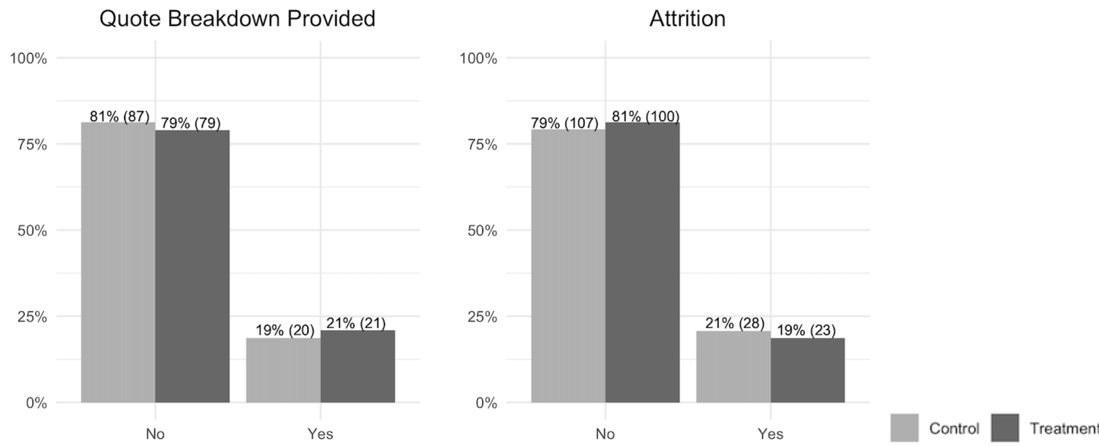

Figure 11: Distribution of Respect Measurements by Gender

We observed a slightly higher level of attrition for male callers (control 21%) than for female callers (treatment 19%). In terms of unsolicited quote breakdowns, female callers (treatment) received them 21% of the time versus 19% for male callers (control). These are surprising results as we had anticipated that auto repair shops would show less professionalism and respect towards female versus male callers.

Regression Models

To prevent multiple comparisons (i.e. fishing for results), we determine the model specifications **before** running the actual regressions with real data in R. Our model analysis is separated into three parts: Primary Outcome (price quote), Secondary Outcomes (knowledge and respect), and Fun Models (which we build just for showing interesting findings unrelated to the research question. Fun Models are included in Appendix G.

Primary Outcome

Our primary results measured by price quotes are summarized in Figure 12. Note that R^2 is small across all four models, suggesting that most of the variance in our outcome variables are explained by the residuals and consumed by the constant term. This is strong evidence that our two randomized treatments, caller gender and ESL status, only explain a small portion of the variance of the quotes we received from shops in the experiment. The issue of gender bias might be irrelevant if the focus is on why prices are different instead of gender inequality.

Without controlling for the fixed effects (i.e. blocked on “county” and “shop chain”), Model 1 shows an ATE of -$40.254 between the treatment and control groups. Model 2 controls for the fixed effects and the coefficient of the treatment variable stays relatively the same with a slightly lower standard error. Both models suggest that female callers in our experiment receive a slight discount of roughly $40. However, in Model 3, the treatment effects for both gender and ESL become insignificant after controlling for the caller’s ESL status and the interaction term. This is evidence that at least some of the variance explained by gender in Model 1 can in fact be explained by the caller’s ESL status.

The last model, Model 4, includes our two randomized variables, gender/ESL affinities, and the interaction terms between gender (our primary treatment) and the three other independent variables. Table 13 summarizes the various outcomes from each of the subgroups formed by the model. The highest effect we observe is when a male native English speaker calls a female ESL shop representative. The lowest effect, on the other hand, is when a male ESL caller calls another male ESL shop representative; note that this is a case with significant effect when both "gender affinity" and "ESL affinity" are positive. Across the possible combinations that are statistically significant in Model 4, we find that being a female or ESL caller generally leads to deeper discounts compared to being a male, non-ESL caller. The affinity factors, contrary to our expectation, do not result in further discounts for female callers. Male callers, on the other hand, do seem to receive some discounts as a result of affinity. The two observed effects, however, can be explained by a behavior unrelated to our research question, as detailed in the next paragraph.

Overall across the four models, we conclude that prices received by female callers are different from those received by male callers. **Female callers generally receive a discount** but the prices increase when the quote is given by a female shop representative. We suspect that the price increase is a result of female representatives being more conservative on estimates in general. The behavior is evidenced that male callers receive quotes that are on average \$43.95 higher when the shop representatives are female. Lastly, **ESL callers also tend to receive a discount**, but the effect is independent of the caller gender with no significant interaction.

	Dependent variable:			
	price_quote			
	(1)	(2)	(3)	(4)
Female Caller	-40.254**	-39.390**	-44.500	-179.343**
	(18.241)	(17.410)	(35.290)	(76.730)
ESL Caller			-21.608	-101.207**
			(28.234)	(48.098)
Gender Affinity				-43.945*
				(24.414)
ESL Affinity				-96.060**
				(43.620)
Female Caller:ESL Caller			-0.875	85.754
			(41.900)	(73.756)
Female Caller:ESL Affinity				97.366
				(72.697)
Female Caller:Gender Affinity				74.811*
				(41.022)
Intercept	688.463***	627.681***	644.981***	769.040***
	(11.563)	(38.204)	(40.805)	(59.410)
Fixed effects?	No	Yes	Yes	Yes
Observations	207	207	207	207
R2	0.023	0.147	0.152	0.186
Adjusted R2	0.018	0.099	0.095	0.113
Residual Std. Error	131.056 (df = 205)	125.568 (df = 195)	125.823 (df = 193)	124.580 (df = 189)
F Statistic	4.877** (df = 1; 205)	3.057*** (df = 11; 195)	2.669*** (df = 13; 193)	2.545*** (df = 17; 189)
Note:				*p<0.1; **p<0.05; ***p<0.01

Figure 12: Primary Outcome Models

	Male Non-ESL Caller	Female Non-ESL Caller	Male ESL Caller	Female ESL Caller
Female & ESL Shop Rep.	769.04* (Intercept)	-148.48	-197.27	-162.63
Male & ESL Shop Rep.	-43.95*	-179.34	-241.22*	-193.49
Female Non-ESL Shop Rep.	-96.06*	-147.17*	-101.21	-163.94
Male Non-ESL Shop Rep.	-140.01*	-92.28*	-145.16*	-194.8

Table 13: Model 4 Treatment Effects Summary

Note that Table 13 summarizes the results from Model 4. The first cell on the top left hand corner is the intercept from Model 4; all other numbers in the table are the differences from the intercept. Numbers marked by the star sign are the results calculated from equations with only significant coefficients from Model 4.

Secondary Outcomes

In addition to our primary outcome, we tested our hypotheses on the possible meditations that could cause the price discrimination. As discussed in the Hypotheses section, the mediations measured are: (1) whether the caller was questioned about the vehicle symptoms as a measure of perception about the caller's knowledge of auto repair, and (2) whether the shop provided a quote to the caller and whether the quote included a price breakdown as measures of the level of respect for the caller. All three secondary outcome variables are binary.

Of the three models, we observe a statistically significant effect for use of the script knowledge section. This indicates that, overall, female callers were questioned 10% more frequently about their knowledge about the radiator or car in general compared to male callers. The finding is rather unexpected because we expect the price to increase when the caller is stereotyped as "less knowledgeable," but our primary models show that female callers in the experiment actually receive a discount. One possible explanation is that the auto shops we call in the experiment are honest and refuse to take advantage of a less knowledgeable customer.

	Dependent variable:		
	script_knowledge	attrition	quote_breakdown
	(1)	(2)	(3)
Felmale Caller	0.100*	-0.026	0.012
	(0.058)	(0.048)	(0.053)
Intercept	0.233	0.337***	0.370**
	(0.142)	(0.121)	(0.150)
Fixed effects?	Yes	Yes	Yes
Observations	207	258	207
R2	0.062	0.081	0.057
Adjusted R2	0.009	0.036	0.003
Residual Std. Error	0.424 (df = 195)	0.392 (df = 245)	0.399 (df = 195)
F Statistic	1.168 (df = 11; 195)	1.795** (df = 12; 245)	1.062 (df = 11; 195)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 14: Secondary Outcome Models

Limitations and Future Enhancements

Generalizability

There were eight callers used in this experiment and several accents were represented in the ESL caller group, however, the caller group is not representative of the entire female and male population or all foreign and regional accents in the United States. Nevertheless, having a multi-caller group does increase the generalizability of this study. As this study is restricted to auto repair shops in Washington State, the generalizability of the results to other geographic contexts is also limited. In terms of the generalizability within Washington State, due to the way that shops were selected, the subject pool is skewed towards shops located in more populous and urban areas.

As mentioned previously, we excluded shops from our analysis that we were unable to contact because they did not answer the phone or did not accept unidentified numbers. It is possible that these shops have different potential outcomes to the shops that we were able to contact, therefore our results only generalize to shops that answer their phone regularly and that accept calls from unidentified numbers. Lastly, this study relies on price quotes provided over the phone. It is possible that auto repair price quotes have a different relationship to gender when a quote is given in person versus over the phone. To increase the generalizability of this study, it could be conducted nationally with callers from a more diverse variety of backgrounds. However, that type of study was not feasible given available resources and project timeline.

Attrition

The high level of attrition in this study may have impacted the reliability of our results. However, there is no statistically significant difference between the percentage of shops that did not provide quotes in the treatment and control groups and our attrition balance check does not not reveal any worrying differences in county or chain characteristics. In addition, our extreme value bound analysis confirms that, even at the extremes, attrition does not impact the overall direction of the treatment effect or alter the overall conclusions of this study. Future studies may be able to reduce the percentage of shops that attrit by altering the script to better handle refusals. They may also consider following up with shops that fail to call back with a quote. However, this creates an additional challenge because these shops will have received a different treatment to shops that were only contacted once.

Measurement

Covariates and Outcomes: While most shops provided a fixed price quote, 34 (16%) shops provided a price range. We used the median price in our analysis when a range was provided, but the actual price that a shop would have charged may have been higher or lower. Since we were unable to determine the actual gender of the auto repair shop representative, we chose to record their assumed gender. This is a subjective measurement and could be incorrectly determined. Also, because we are treating gender as binary for the purposes of this experiment, we may have incorrectly classified any gender non-binary auto repair shop representatives. Both of these challenges are difficult to overcome because shops cannot be required to provide a fixed price quote and we could not ask a shop representative to disclose their gender without raising suspicion. The study might have also benefited from further categorization of chain shops. For example, large versus small chains, but, as this coding must be done manually, it is a very time consuming process.

Statistical Models: Our pilot data power analysis indicated that, given the experiment sample size, a simple model regressing treatment on price quote, would require an effect size of at least $77.04 for 0.8 statistical power and a 0.05 significance level. Although we see a statistically significant treatment effect with our base model, the effect size is only $40.25, approximately $37 less than the effect size indicated by our power analysis. We do see a slight reduction in standard error after adding chain and county fixed effects, but, given the limited sample size, these results should still be interpreted with caution. As we explore heterogeneous treatment effects in the more complex models, we necessarily add more covariates. Given the small sample size, further reducing group sizes increases the unreliability of our estimates, as evidenced by considerable increases in standard error. Therefore, effect sizes in these models should be interpreted with even more caution.

Additionally, there were only a small number of positive cases in each of the secondary outcomes that we used to assess mediation. These analyses may have benefitted from synthetic oversampling to address class imbalance. Finally, we did not adjust p-values for multiple model comparisons, therefore, all statistically significant effects should be confirmed with a follow-up study.

Administration Consistency: While callers used a standardized script, it is impossible to ensure perfectly consistent delivery in every call. Differences in the caller's tone of voice, style of delivery, background noise (or lack thereof), and even how fast they talk may have all influenced price quotes. Future studies could consider more rigorous training for callers. Also, due to logistical constraints such as geographic separation and the schedule availability of our callers, we were not able to randomize when calls were made or to make all calls during the same time or day. If the day and time when calls are made impacts price quotes, then caller specific variation in calling day and time, could have impacted our results. Future studies with dedicated callers could consider standardizing or randomizing call day and time.

Lastly, shop representatives that requested to call back with a quote were given an out of state number where they could leave a message. While the same call back number was used for all callers, shops that called back rather than provided a quote directly to the caller were given more information about the caller. Specifically, that the caller had an out of state number. This additional knowledge could have impacted their price quotes. Future studies could explore other ways of handling call-backs to ensure that each shop receives the same information about the caller.

Conclusion

Although this study was not able to recreate the results of the Busse et al experiment, it did conclude a statistically significant difference in the way uninformed males and females, as well as native English and ESL speakers are quoted for service. Surprisingly, our research concluded that female callers in general do receive discounts, a direct contrary to the other researches referenced in the paper. Perhaps more interestingly, males with no apparent accent, presenting as native English-speakers, are most disadvantaged when offered quotes. This may have been a byproduct of the scoped selection of shops in Washington state given the restraints of the experiment (time, team size, etc), and warrants additional research. Naturally, additional research with greater focus would expand upon the limited understanding this study provides. Recommended future studies should include an ROXO design where the same shop receives both control and treatment in a manner that extracts any shift in response with greater fidelity.

Regardless of outcomes, the study stands on merit and furthered the existing research. Perhaps the shift in outcome to treatment is a result of the changing times, or perhaps a more regional-focused culture shift. Regardless, the study provides couples and household units the opportunity to be more informed about how, when able, to best achieve the most optimal service quote for future automotive repairs.

References

[1] Ayres, I., & Siegelman, P. (1995). Race and gender discrimination in bargaining for a new car. The American Economic Review, 304-321.
https://opencommons.uconn.edu/cgi/viewcontent.cgi?article=1454&context=law_papers

[2] Gneezy, U., List, J., & Price, M. K. (2012). Toward an understanding of why people discriminate: Evidence from a series of natural field experiments (No. w17855). National Bureau of Economic Research.
https://www.nber.org/system/files/working_p apers/w17855/w17855.pdf

[3] Chen, L. Z., Hu, W. M., Szulga, R., & Zhou, X. (2018). Demographics, gender and local knowledge—Price discrimination in China's car market. Economics Letters, 163, 172-174.

[4] Schneider, H. S. (2012). Agency problems and reputation in expert services: Evidence from auto repair. The Journal of Industrial Economics, 60(3), 406-433.
https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-6451.2012.00485.x

[5] Goldberg, P. K. (1996). Dealer price discrimination in new car purchases: Evidence from the consumer expenditure survey. Journal of Political Economy, 104(3), 622-654.
https://www.journals.uchicago.edu/doi/abs/10.1086/262035

[6] Castillo, M., Petrie, R., Torero, M., & Vesterlund, L. (2013). Gender differences in bargaining outcomes: A field experiment on discrimination. Journal of Public Economics, 99, 35-48.
https://www.nber.org/system/files/working_papers/w18093/w18093.pdf

[7] Busse, M. R., Israeli, A., & Zettelmeyer, F. (2013). Repairing the damage: The effect of price knowledge and gender on
auto repair price quotes. Journal of Marketing Research, 54(1), 75-95. https://www.nber.org/papers/w19154

[8] Foster, T. B., Murray-Close, M., Landivar, L. C., & deWolf, M. (2020). An Evaluation of the Gender Wage Gap Using Linked Survey and Administrative Data (No. 20-34).
https://www.census.gov/library/working-papers/2020/adrm/CES-WP-20-34.html

Appendix

Appendix A: Caller Script

Greeting
[Time-appropriate Greeting], How's it going?
I am hoping you can help me out. I am trying to get an idea of how much it would cost to replace the radiator on a vehicle.

Basic Information
2012 Toyota Camry LE
6-cylinder
2.5 liter engine size
Just over 100,000 miles
4 door
Automatic transmission
Gas

Contingencies
A: Shop Provides Quote
- Thanks very much. I'll make some calls and get back to you.

B: Shop Asks For Additional repairs/parts
- Let's just start with replacing the radiator for now

C: Shop Asks About Budget
- I'm not sure how much it should cost. You are the first shop that I have called.

D: Shop Asks "How do you know it's the radiator?"
- I am looking to purchase this car; It appears to be in good shape but the owner says the radiator needs to be replaced
- I work with the owner so I have no reason to doubt his/her assessment

E: Shop Asks About History of the Car
- Yes, it's leaking radiator fluid
- I don't know much else about the car other than it looks to be in good condition and the owner told me about this one problem
- I am purchasing it for $6000

F: Shop Will Not Provide Quote
- I'd like to know how much it will cost before bringing the car in.

G: Still refuses
- Well, thanks anyway for your time.

H: Provides quote
- Thanks very much. Let me look this over and I'll get back to you. Have a great day.

Appendix B: Data Collection Rubric

Table 15 shows the template callers used to collect information. The template is converted into Google spreadsheets, one sheet per caller.

Column Name	Instruction	Response Options
id	*information from randomized dataset*	
name		
address_city		
address		
phone_number		
county		
chain		
treat		
caller_id		
call_attempts	*Select the number of attempts made to each shop*	1 to 5
day	*select day of week when quote was provided/refused*	Monday Tuesday Wednesday Thursday Friday Saturday Sunday
time	*select hour when quote was provided/refused (24hr Pacific Time)*	8 to 20
female_rep	*select yes if person that provided quote/refused is female, no for male*	Yes No Unknown
esl_rep	*select yes if person that provided quote/refused is a non-native English speaker, no for native English speaker*	Yes No Unknown
script_budget_upcharge	*Did you have to use script language from section B or C?*	Yes No
script_knowledge	*Did you have to use script language from section D or E?*	Yes No
quote_point	*enter dollar amount, if fixed quote given, leave blank if no quote OR range given* ***If a breakdown was provided, provide the TOTAL price***	*numeric*
quote_range_low	*if range given, provide dollar amount of bottom of range (leave blank if fixed quote OR no quote)*	*numeric*
quote_range_high	*if range given, provide dollar amount of top of range (leave blank if fixed quote OR no quote)*	*numeric*
no_quote_reason	*select reason for no quote, leave blank if quote given*	No contact (after 2 tries) Does not provide this service Refused to provide quote Other
quote_breakdown	*select yes if unsolicited breakdown of costs was provided, no if not (leave blank if no quote given)*	Yes No
notes	*provide other important context, as needed*	*text*

Table 15: Data Collection Rubric

Appendix C: No Quote Reason Breakdown

Table 16 summarizes the distributions for shops that attrit from the experiment. Table and 17 summarizes the distribution for the shops that are excluded from the experiment. The main difference between attrition and exclusion is that shops that attrit are the ones that receive the treatment (i.e. shops answer the calls our callers make) but whose results cannot be measured (e.g. unable to provide quotes). Exclusions, on the other hand, are the shops that never receive the treatment because our callers are unable to reach them.

No Quote Attrition	Overall, N = 51	Control	Treatment
Reason			
Failed callback	36 (71%)	18 (64%)	18 (78%)
Refused to provide quote	12 (24%)	7 (25%)	5 (22%)
Unable to provide quote	3 (5.9%)	3 (11%)	0 (0%)

Table 16: Attrition Reason Distribution

No Quote Exclude	Overall, N = 158	Control	Treatment
Reason			
Does not provide this service	31 (20%)	21 (23%)	10 (15%)
No contact (after 2 tries)	123 (78%)	69 (75%)	54 (82%)
Out of business	4 (2.5%)	2 (2.2%)	2 (3.0%)

Table 17: Exclusion Reason Distribution

Appendix D: Attrition Balance Check

Differential Attrition T-Test

Table 18 includes a t-test with the null hypothesis that the attrition ratio between treatment and control are equal. In addition, Table 19 summarizes the attrition ratios between the different counties and between the two different types of shops (chain vs. no chain) included in the experiment. Table 20 to 23 are the results of the extreme bounds analysis in the presence of attritions conducted for the four models in Table 12.

```
##  Welch Two Sample t-test
##
## data:  attrition by treat
## t = 0.41054, df = 255.23, p-value = 0.6818
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.07751446  0.11834554
## sample estimates:
## mean in group 0 mean in group 1
##       0.2074074       0.1869919
```

Table 18: T-Test for Attritions Between Treatment and Control

	Overall, N = 615	Control	Treatment
chain	83 (13%)	46 (15%)	37 (12%)
county			
benton	34 (5.5%)	15 (4.8%)	19 (6.2%)
clark	31 (5.0%)	14 (4.5%)	17 (5.6%)
franklin	14 (2.3%)	9 (2.9%)	5 (1.6%)
island	2 (0.3%)	1 (0.3%)	1 (0.3%)
king	240 (39%)	119 (38%)	121 (40%)
pierce	53 (8.6%)	30 (9.7%)	23 (7.5%)
snohomish	62 (10%)	27 (8.7%)	35 (11%)
spokane	59 (9.6%)	30 (9.7%)	29 (9.5%)
thurston	50 (8.1%)	27 (8.7%)	23 (7.5%)
whatcom	36 (5.9%)	20 (6.5%)	16 (5.2%)
yakima	34 (5.5%)	18 (5.8%)	16 (5.2%)

Table 19: Attrition Balance Check

	Dependent variable:				
	outcome (1)	outcome_min (2)	outcome_25 (3)	outcome_75 (4)	outcome_max (5)
treat	-40.254**	-25.409	-30.809**	-32.248**	-39.455*
	(18.420)	(22.472)	(15.281)	(14.833)	(22.913)
Constant	688.463***	614.115***	668.981***	683.597***	756.811***
	(11.672)	(15.623)	(9.806)	(9.270)	(14.815)
SE Flavor	Robust	Robust	Robust	Robust	Robust
County-Chain fixed effects	No	No	No	No	No
Observations	207	258	258	258	258
R2	0.023	0.005	0.016	0.019	0.012
Adjusted R2	0.018	0.001	0.012	0.015	0.008
Residual Std. Error	131.056 (df = 205)	179.693 (df = 256)	121.365 (df = 256)	117.569 (df = 256)	182.089 (df = 256)
F Statistic	4.877** (df = 1; 205)	1.287 (df = 1; 256)	4.148** (df = 1; 256)	4.842** (df = 1; 256)	3.022* (df = 1; 256)
Note:					*p<0.1; **p<0.05; ***p<0.01

Table 20: Extreme Value Bounds Analysis for Model 1 in Table 12

	Dependent variable:				
	outcome (1)	outcome_min (2)	outcome_25 (3)	outcome_75 (4)	outcome_max (5)
treat	-39.390**	-21.268	-28.143	-29.974	-39.149
	(18.495)				
Constant	627.681***	529.615	618.719	642.456	761.360
	(42.390)				
SE Flavor	Robust	Robust	Robust	Robust	Robust
County-Chain fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	207	258	258	258	258
R2	0.147	0.097	0.117	0.119	0.097
Adjusted R2	0.099	0.053	0.073	0.076	0.052
Residual Std. Error	125.568 (df = 195)	174.939 (df = 245)	117.542 (df = 245)	113.877 (df = 245)	177.955 (df = 245)
F Statistic	3.057*** (df = 11; 195)	2.205** (df = 12; 245)	2.695*** (df = 12; 245)	2.752*** (df = 12; 245)	2.183** (df = 12; 245)
Note:					*p<0.1; **p<0.05; ***p<0.01

Table 21: Extreme Value Bounds Analysis for Model 2 in Table 12

	Dependent variable:				
	outcome (1)	outcome_min (2)	outcome_25 (3)	outcome_75 (4)	outcome_max (5)
treat	-44.500**	-62.789	-38.267	-31.734	0.989
	(18.495)				
esl_caller	-21.608	-56.004	-26.277	-18.358	21.311
treat:esl_caller	-0.875	47.173	4.603	-6.738	-63.546
Constant	644.981***	569.046	637.475	655.705	747.020
	(42.390)				
SE Flavor	Robust	Robust	Robust	Robust	Robust
County-Chain fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	207	258	258	258	258
R2	0.152	0.105	0.124	0.126	0.104
Adjusted R2	0.095	0.053	0.073	0.075	0.052
Residual Std. Error	125.823 (df = 193)	174.946 (df = 243)	117.560 (df = 243)	113.900 (df = 243)	177.994 (df = 243)
F Statistic	2.669*** (df = 13; 193)	2.031** (df = 14; 243)	2.447*** (df = 14; 243)	2.494*** (df = 14; 243)	2.005** (df = 14; 243)
Note:					*p<0.1; **p<0.05; ***p<0.01

Table 22: Extreme Value Bounds Analysis for Model 3 in Table 12

	Dependent variable:				
	outcome (1)	outcome_min (2)	outcome_25 (3)	outcome_75 (4)	outcome_max (5)
treat	-56.186	-94.362	-139.457	-151.471	-211.648
	(40.497)				
esl_caller	-65.268	-104.750	-96.445	-94.233	-83.151
	(50.739)				
gender_affinity	-11.142	14.144	-19.020	-27.854	-72.109
	(23.783)				
esl_affinity	-50.043	-64.118	-83.279	-88.384	-113.953
	(44.575)				
treat:esl_caller	2.809	92.044	81.179	78.285	63.787
	(46.501)				
treat:esl_affinity		41.147	86.387	98.438	158.807
treat:gender_affinity		-23.633	33.991	49.342	126.238
Constant	695.476***	608.673	728.968	761.014	921.541
	(60.490)				
SE Flavor	Robust	Robust	Robust	Robust	Robust
County-Chain fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	207	249	249	249	249
R2	0.164	0.099	0.142	0.159	0.164
Adjusted R2	0.098	0.028	0.075	0.093	0.099
Residual Std. Error	125.628 (df = 191)	172.762 (df = 230)	119.009 (df = 230)	114.827 (df = 230)	168.962 (df = 230)
F Statistic	2.494*** (df = 15; 191)	1.402 (df = 18; 230)	2.122*** (df = 18; 230)	2.413*** (df = 18; 230)	2.512*** (df = 18; 230)
Note:					*p<0.1; **p<0.05; ***p<0.01

Figure 23: Extreme Value Bounds Analysis for Model 4 in Table 12

Appendix E: Callers by Week Day and Day Time

Figure 24 and 25 summarize the distributions of the calls made between the callers, days of the week, and times of the day.

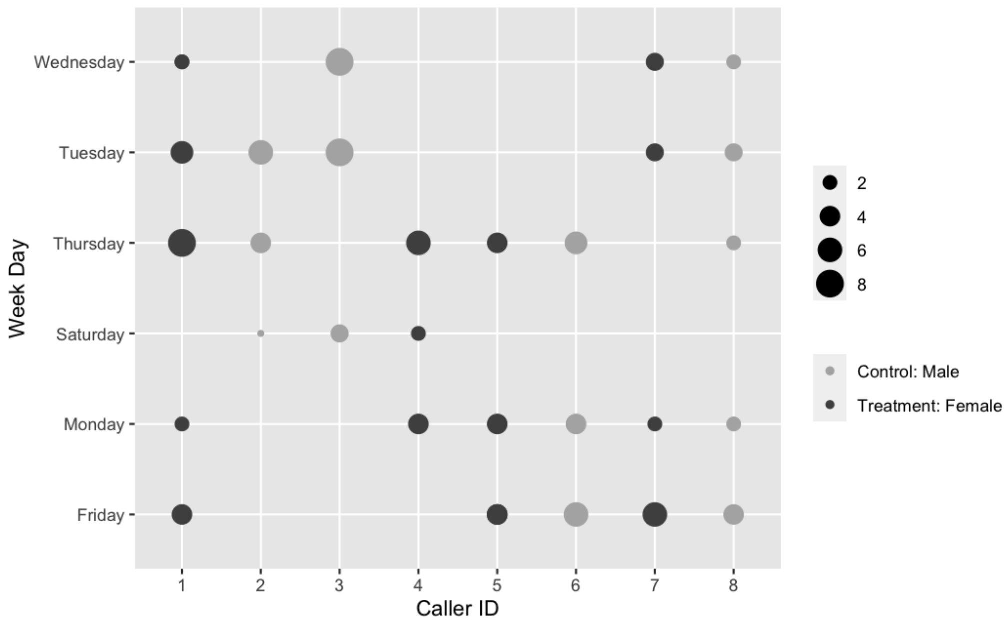

Figure 24: Caller ID by Day Breakout

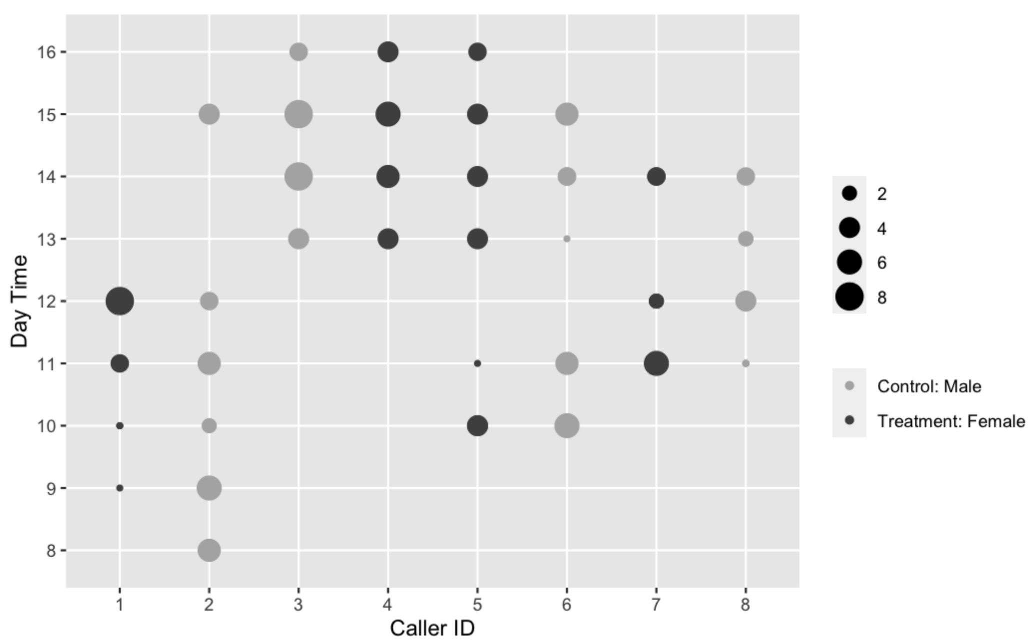

Figure 25: Caller ID by Day Time

Appendix F: Gender/ESL Affinity Breakdown

Affinity is the match of the characteristics between the callers and shop representatives who provide the quotes. Three affinities are analyzed in the experiment: gender, ESL (English as a second language), and gender + ESL.

	Overall, N = 207	Control	Treatment
Gender Affinity			
No	103 (50%)	20 (19%)	83 (83%)
Yes	104 (50%)	87 (81%)	17 (17%)
ESL Affinity			
No	136 (66%)	80 (75%)	56 (56%)
Yes	71 (34%)	27 (25%)	44 (44%)
Gender/ESL Affinity			
No	175 (85%)	85 (79%)	90 (90%)
Yes	32 (15%)	22 (21%)	10 (10%)

Table 26: Affinity Breakdown

Appendix G: Fun Models

In addition to the models discussed in Regression Models, we built three more models just for fun. The first model tests if calling on a particular day of the week results in higher quotes, on average. The base case of Model 1 is Friday and Sunday is missing from the model because no call was attempted on Sunday. The second model tests if any of the eight callers received higher quotes, on average, than the other callers. Lastly, Model 3 includes only the fixed effects to examine the county and chain effects in isolation.

Across the three models, we observe a significant discount of $45.50, on average, for calls made on Mondays, but no significant effect for any of the eight callers. We also find a price premium of $85.11, on average, in King County. This makes sense because King County is a relatively wealthy county where multiple leading high-tech firms have their headquarters, including Amazon, Microsoft, and Expedia. More surprisingly, Thurston County has an even higher premium of $116.54, on average. Thurston County includes the state capital of Olympia and has the fifth highest per capita income of all counties in Washington State which may partly explain this premium. Finally, we find that chain shops charge $50.18 more, on average, than shops with only a single business location.

	Dependent variable:		
	price_quote		
	(1)	(2)	(3)
Monday	-45.497*		
	(23.705)		
Tuesday	32.390		
	(24.385)		
Wednesday	50.834		
	(34.053)		
Thursday	-0.613		
	(23.574)		
Saturday	-60.221		
	(39.187)		
Caller 2		18.307	
		(36.847)	
Caller 3		-13.726	
		(36.434)	
Caller 4		-64.132	
		(40.662)	
Caller 5		-27.005	
		(32.353)	
Caller 6		-1.209	
		(30.930)	
Caller 7		-69.137	
		(43.996)	
Caller 8		21.994	
		(38.105)	
Clark County	56.744	60.637	67.595
	(56.411)	(59.197)	(58.890)
Franklin County	29.510	31.831	45.661
	(68.674)	(64.172)	(64.778)
King County	77.531*	81.595**	85.110**
	(41.907)	(39.310)	(43.101)
Pierce County	20.759	14.616	17.387
	(47.301)	(45.908)	(48.807)
Snohomish County	54.431	52.341	62.399
	(54.776)	(50.283)	(54.915)
Spokane County	-42.617	-35.128	-39.181
	(45.439)	(43.931)	(46.492)
Thurston County	118.806**	111.248**	116.540**
	(46.704)	(47.096)	(47.403)
Whatcom County	75.131	74.090*	80.989*
	(46.757)	(43.054)	(47.403)
Yakima County	50.325	82.685	65.852
	(51.657)	(54.877)	(52.860)
Chain	57.554**	50.678**	50.178**
	(24.172)	(24.723)	(24.668)
Intercept	603.485***	621.365***	603.562***
	(43.384)	(48.381)	(41.040)
Observations	207	207	207
R2	0.176	0.174	0.125
Adjusted R2	0.112	0.100	0.081
Residual Std. Error	124.671 (df = 191)	125.528 (df = 189)	126.840 (df = 196)
F Statistic	2.729*** (df = 15; 191)	2.339*** (df = 17; 189)	2.806*** (df = 10; 196)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 27: Fun Models