

Eliciting the Endowment Effect under Assigned Ownership

W241-3: Experiments and Causality

Patrick Barranger, Rohit Nair, Rob Mulla, Shane Conner

8/19/2017

Abstract

The endowment effect is the tendency for people who own a good to value it more than people who do not. Its economic impact is consequential. It creates market inefficiencies and irregularities in valuation such as differences between buyers and sellers, reluctance to trade, and mere ownership effects. This study (n=495) presents evidence that endowment effect can be elicited merely by assigned ownership. Employing survey responses we were able to generate an endowment effect size of 15-20 times (at $p < 0.05$) increase between willingness to pay (WTP) and willingness to accept (WTA).

I. Introduction

Many economic theories are developed under a rational agent model, where consumers are expected to treat all information unbiasedly in their decision making process. In the 1960s economists started to realize that consumers did not always act in this anticipated rational manner. Research lead to the development of behavioral economics, the study of mental processes such as “attention, language use, memory, perception, problem solving, creativity, and thinking”. One of the hypothesis developed was the endowment effect, the hypothesis that consumer ascribe more value to things merely because they own them. The endowment effect captures the observation of the valuation paradigm, where people will tend to pay more to retain something they own than to obtain something they do not own –even when there is no cause for attachment, or even if the item was only obtained minutes ago. Studies of the effect typically focus on the difference in Willingness to Pay (WTP) in comparison to the Willingness to Accept (WTA). In one famous example participants were given a mug and then offered the chance to sell or trade for an equally valued alternative, pens. The researchers found that once participants had the mug their WTA was twice as high as their WTP. In another similar study, researchers found that participants selling price of NCAA final four tickets (WTA) was 14 times higher than their WTP. With magnitudes ranging from two to 14 times in research, it is clear that the endowment effect can cause people to assign drastically different values due to simple possession. What if this effect extended beyond physical ownership to hypothetical possession? In this paper, we attempt to answer this question by means of a questionnaire that sets up this hypothetical ownership. This paper is organized as follows. Section II give an overview of current theories on endowment effect. Section III describes our experiment design while in Section IV we share data analysis. Results and findings are discussed in section V. Section VI concludes this paper.

II. The Endowment Effect

Endowment effect is typically demonstrated using two experimental paradigms, exchange paradigm and valuation paradigm[4]. In exchange paradigm, participants are endowed with one of two goods and are tested for their willingness to exchange it for the other good. What was found was that they were more reluctant to exchange than would be found by chance. In valuation paradigm, half of the participants are endowed with a good (sellers) and the other half (buyers) is asked what they are willing to pay to acquire said good. In this scenario what was found was that the *maximum* amount the buyers were willing to pay to acquire the good was lower than the *minimum* amount of money the sellers of the goods were willing to accept to relinquish

it. This violates the foundational assumption of law governing allocation and distribution of entitlements, the Coase theorem; while providing evidence towards Prospect theory. It assumes loss aversion, diminishing marginal utility, reference-dependence and non-linear decision weights influence how people value goods. In addition to above, evolutionary advantage, strategic misrepresentation, biased information processing and psychological advantage theories try to explain the different instantiations of the endowment effect.

Loss Aversion

Endowment effect can be attributed to reference-dependence feature of prospect theory which states that buyers frame goods as gains while sellers frame goods as losses. As people are loss averse, ie. the psychological impact of loss is significantly higher than gain of same amount, there is a difference in value that a seller attributes to good compared to what buyer values it at. In conjunction with this are the findings that people's valuation of a good is dependent on how long they have owned it and whether they intend to own the good in the future. Higher valuation due to these factors result in the endowment effect.

Evolutionary Advantage

Evolutionary studies have suggested that people overvalue goods they own because it bestowed advantage in bargaining. The better they bargained, the more resources they acquired which enabled them to support more offsprings. Although studies have shown that this effect is dependent on other covariates such as education, culture, upbringing.

Strategic Misrepresentation

This theory suggests that endowment effect materializes because the participants understand that they are in a negotiation and they strategically value the good to their advantage. Although studies have shown that this is entirely not true as the participants do not predict the effect leading to the understanding that it's not premeditated. Further the effect were present in cases where there were no negotiations and participants were given one opportunity to buy or sell a good. Another theory is that participants behave in the way they think the experimenter wants them to behave.

Reference Prices

A good can have multiple reference prices depending on who and in which context it is being valued. Reference price theory proposes that when the value of good to a person compares unfavorably, a buyer will reduce their WTP while a seller will increase their WTA so as to hedge against having a bad experience while negotiating. Additionally buyer favor low reference price while sellers favor high reference price. For example, while selling a car, a setter values the car at KBB's *Excellent* price point while a buyer will at most value the car at KBB's *Good* price point.

Biased Information Processing

Cognitive process theory proposes that confirmatory biases are in play while evaluating a good. Buyers invoke memory or information that suggests retaining money is much better than owning the good while for a seller there is increased accessibility to information that suggests keeping the good is better than parting with it. These informations are spontaneously processed and acted on. Studies have shown that asking buyers to consider the positive features of the good increased their WTP while asking sellers to consider the negative aspects of the good decreased their WTA which in effect reduced the endowment effect.

Psychological Ownership

Studies have shown that ownership of a good increases its perceived value. The longer a person owns a good, the more favorably they value the good. One theory explains this by suggesting that ownership creates association with the good that were previously non-existent. These associations causes the seller to value the good higher than what they would have valued it at. When a good becomes part of owner's identity, the evaluation rises higher. Opposite effects is noticeable in the cases where there is a negative association with the good in which case the seller values it less and have lower WTA. Examples of this would be goods from ex-relationships that bring forth bad memories or goods that someone think is jinxed or unlucky.

Our study attempts to add to this body of work by trying to understand if an individual would respond differently between a question where ownership is assigned and where it is not. Our hypothesis is that assigning ownership will cause the participant to price the object more highly.

III. Experimental Design

To test our hypothesis we set up a post-treatment measurement 2x2 test design where we asked subjects to price seating features on a five hour flight. Specifically, we asked them to either price legroom or the recline feature. Half were asked their willingness to accept a payment (WTA) to give up the feature and the other was asked the willingness to pay (WTP) for the feature.

Data Gathering and Subject Interaction

Our team decided to continue to use Google Consumer Surveys to gather our data responses. Google surveys was chosen because of its relative cost compared to other survey options, quick response time, and additional information provided about the respondents like age, gender, and location. Our team make a cautious decision to potentially trade off quality of responses so that we could have more responses.

Our participants were mainly visitors of news websites who wanted to read articles behind the website's paywall. Instead of paying, Google allows customers to complete a surveys such as ours to read the article. Each subject was asked only one of our four questions, and 133 responses were gathered for each question. Participants could be reaching using their computer or mobile device. We note that because most of our respondents were attempting to get through paywalls they in a sense had "self selected" into our study. We also note that our sample population may be similar in other ways due to this type of experiment, specifically that they all are the type of people who desire to read news articles.

Pilot Study

Before embarking on a full fledged study of endowment effect, we wanted to ensure that different aspects of the study are working as expected. To this effect we initiated a pilot study. Our object in running a pilot study were:

- Ensuring Google Customer Survey (GCS) randomization works,
- Survey takers understand our survey questions,
- We are able to deliver treatment as we intended to,
- Verify that responses are valid and meets expectation,
- Distill presence of Endowment effect and
- Get a baseline to calculate statistical power required for full fledged study.

The pilot study ran on Google Customer Survey from 25th June - 27 June, 2017. The survey consisted of two questions.

Question	Description
1	You are traveling on a 5 hour flight. What is the maximum amount you would be willing to pay to recline your seat?
2	You are traveling on a 5 hour flight. What is the minimum amount you would accept to not recline your seat?

Each survey taker is randomly assigned to one of the two questions and they are provided with a textbox to post their response. The textbox doesn't run any validation on the responses and accepts free form text. This was done so as to gauge the kind of responses we get. We set the survey to end when GCS receives at least 50 responses per question. Both the questions assigns the respondent ownership of the chair. Question 1 is testing for their willingness to pay while question 2 tests their willingness to accept. Once the survey ends, GCS provides functionality to export survey responses with additional attributes pertaining to respondents. The data consists of following columns:

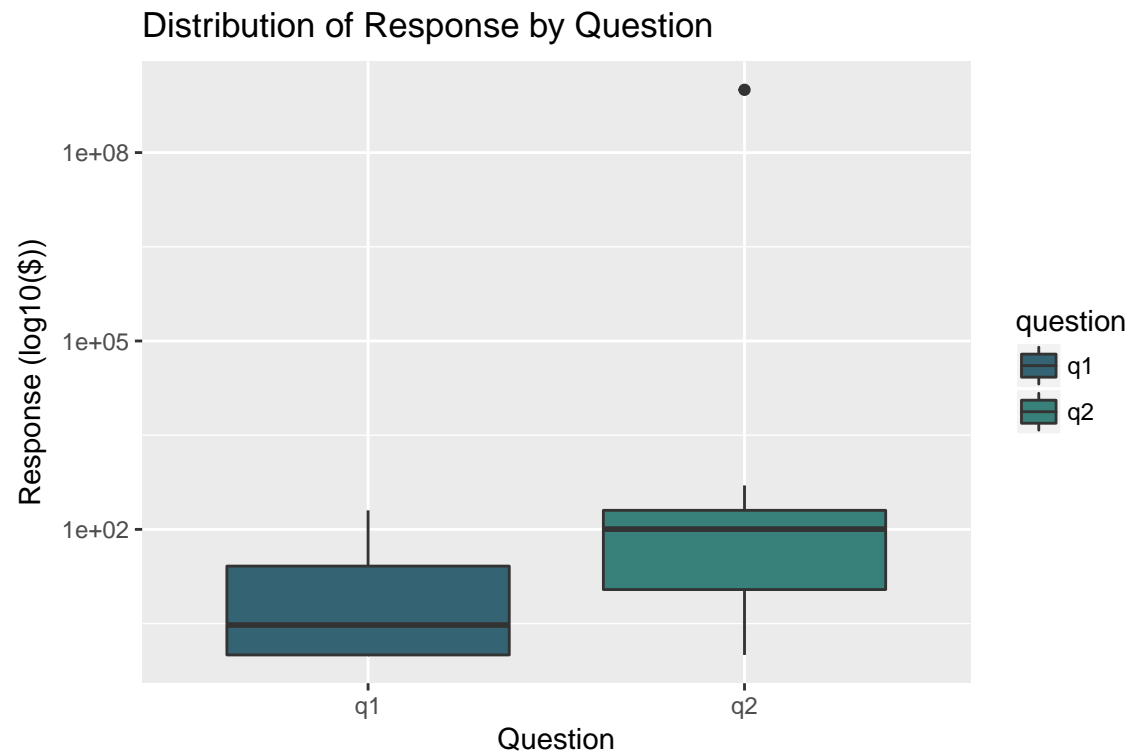
Data Column	Description
User ID	A unique user ID for each survey respondent
Time (UTC)	The time in which the survey was completed
Publisher Category	The type of website the user was trying to view when filling out the survey
Gender	Gender of survey respondent or Unknown
Age	The age bracket of the survey respondent or Unknown
Geography	The Country, region and State of the respondent
Urban Density	Population density of user's location (Urban, suburban, rural)
Income	Income bracket user falls into
Parental Status	Parental status of the user
Question ## Response	Raw response of the respondent
Response Time #1 (ms)	Time taken to respond

Once we exported the data, we loaded it in R and checked the validity of the responses. Of the 101 responses, following are the non-numeric ones. All but three ("Ly", "Cff", "5 hours") of them can be easily converted to numeric format. We will mark these three as NA so as not to affect our calculations.

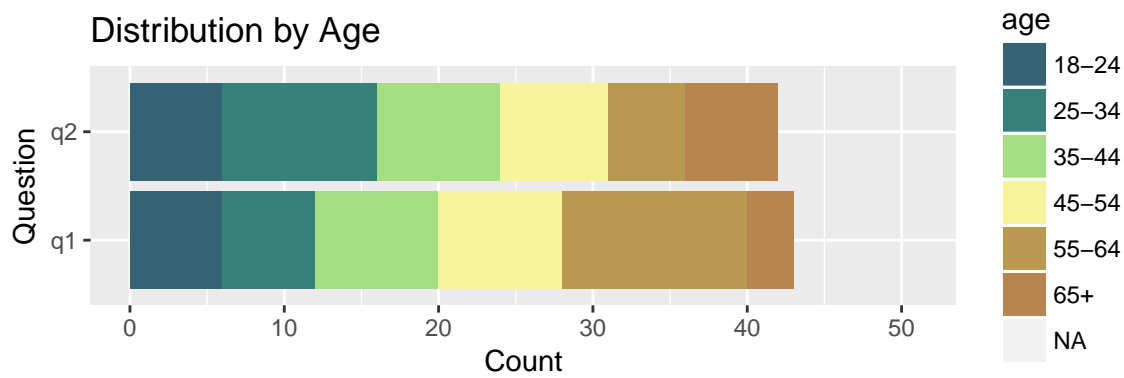
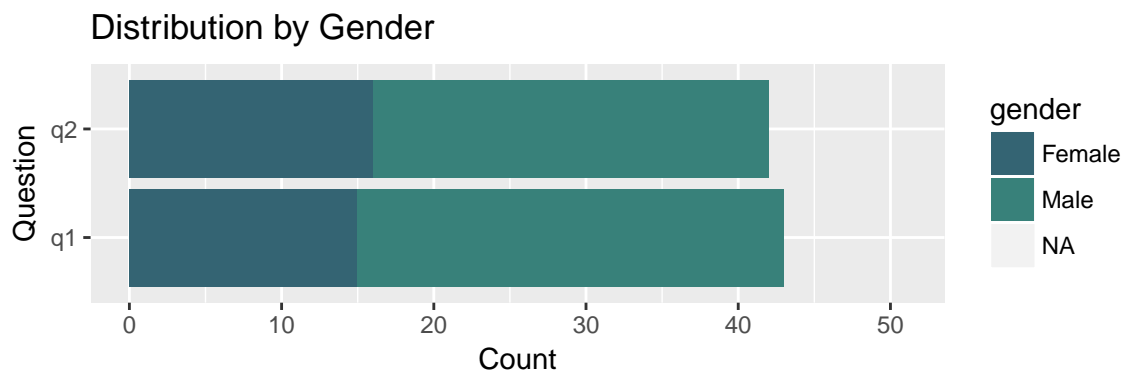
Odd responses

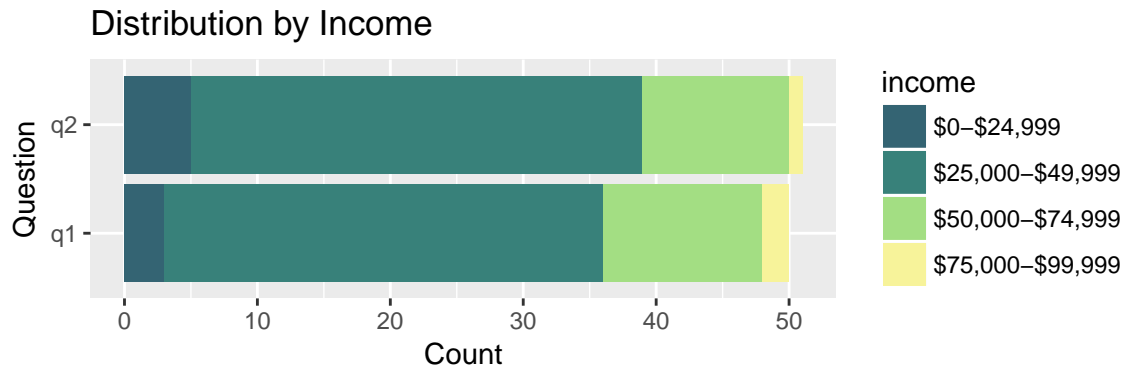
Odd	Responses
Ly	\$2
\$50	\$10
\$0	0...I think it's rude to recline on any flight...ever
\$100	50 dollars
Cff	\$100
5 hours	\$100
\$150	300.00 us
Zero	1,000000000.00

Following this we plot parsed responses on a box plot to see whether the distribution is as expected. We see that the median response for question 1 is around \$10 while that for question 2 is around \$100. The difference between responses to these two questions is the endowment effect which we were able to elicit in our pilot study which is promising.



Next we check whether the responses are distributed uniformly among different demographics. Below we show distribution of responses for each question by Gender/Age/Income.





Next, we ran a test for covariate balance. We create a subset of responses by omitting NAs for the covariates. We create two regression models, one with all the covariates and another with no covariates. Using ANOVA we test whether any of these

```
## Analysis of Deviance Table
##
## Model 1: treatment ~ gender + region + age
## Model 2: treatment ~ 1
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1         75      18.435
## 2         84      21.247 -9   -2.8117   0.2468
```

The result is a non statistically significant p-value of 0.2468 which suggests the randomization design was not violated. The data from the pilot study shows that we were able to get right responses for our questions and that GCS did a good job of randomizing respondents.

Actual Study

In our actual study we expanded our questions to four, gathering 133 responses per question. These questions were:

Question Number	Question Text
q1	On a 5 hour flight what is the maximum amount you would be willing to pay to recline the seat?
q2	On a 5 hour flight what is the minimum amount you would accept to not recline your seat?
q3	On a 5 hour flight what is the maximum amount you would be willing to pay to stop the passenger in front of you from reclining into the space in front of you?
q4	On a 5 hour flight what is the minimum amount you would accept to allow the passenger in front of you to recline into your space?

IV. Data Analysis

Our raw data is provided to us by google surveys in the form of a xls file. While our main concern is the answers provided by the survey respondents, we are also provided additional information about our sample population. We can run some statistical tests to determine if these are the same for each group.

Data Column	Description
User ID	A unique user ID for each survey respondent
Time (UTC)	The time in which the survey was completed
Publisher Category	The type of website the user was trying to view when filling out the survey
Gender	Gender of survey respondent or Unknown
Age	The age bracket of the survey respondent or Unknown
Geography	The Country, region, State and City of the respondent
Question Raw Response	Raw response of the respondent
Response Time #1 (ms)	Time taken to respond

Response filtering and Noncompliance

Our survey allowed the participant to type any text in the response box. The text box, before the respondents start entering text stated “Enter your answer in US \$”. Because of this the responses were not all easily converted into numerical values. Some manual interpretation was required. Many of the responses could be converted, for instance ‘2 dollars’ could be interpreted as ‘2’. Other responses such as those in units of measurement (3 inches), units of time (3 hours) or nonsense responses (wa) were all converted to NA values and ignored for our analysis. These participants are considered ‘noncompliers’ in our study.

Below are the responses that were manually converted to numeric values.

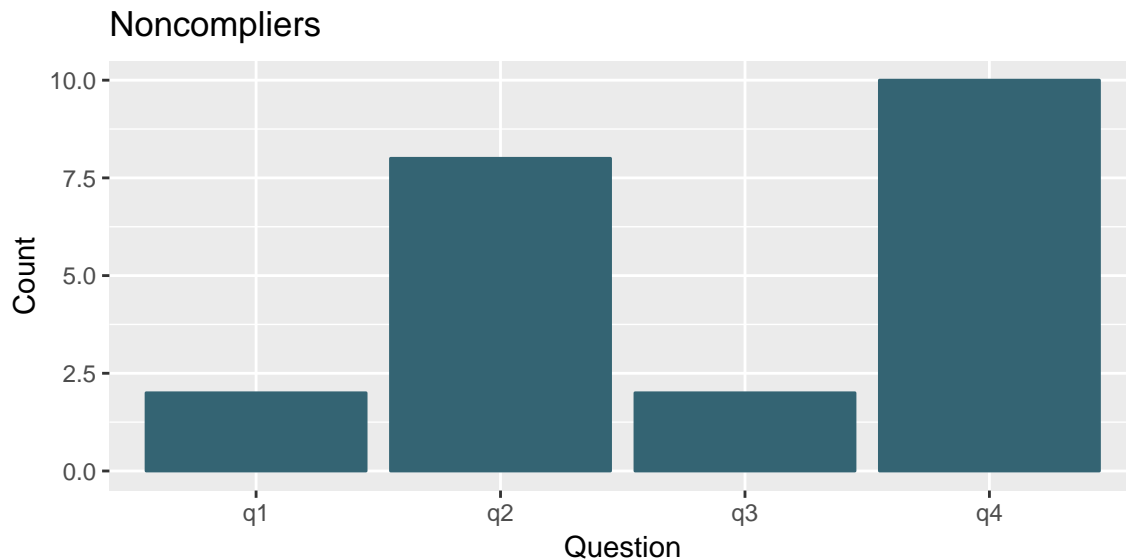
```
# Cases which can be turned into numerical values
d[d$response == "none it should be free"]$response <- 0
d[d$response == "1 000 000 000.69"]$response <- 1000000000.69
d[d$response == "two"]$response <- 2
d[d$response == "zero"]$response <- 0
d[d$response == "nothing"]$response <- 0
d[d$response == "none"]$response <- 0
d[d$response == "195 500 812.50"]$response <- 195500812.50
d[d$response == "100 000 000.836215"]$response <- 100000000.836215
d[d$response == "non"]$response <- 0
d[d$response == "10 000"]$response <- 10000
d[d$response == "0.00 as long as the seat reclines it is their right to recline it"]$response <- 0
d[d$response == "25 dollars"]$response <- 25
d[d$response == "zedo"]$response <- 0
d[d$response == "1 000 000.00"]$response <- 1000000.00
d$response <- as.numeric(d$response)
```

The following were responses that were converted to NA.

```
## [1] "i don't know"          "i don't fly commercial"
## [3] "2 hours"               "reject offer"
## [5] "250 free flight"       "2 hours"
## [7] "2 hours"               "2 min"
## [9] "do not follow question" "na"
## [11] "yes"                   "wa"
## [13] "2.5 hour"              "not at all"
## [15] "yes"                   "3 hours"
```

```
## [17] "5 hours"          "15 minutes"
## [19] "5 hours"          "1 hour"
## [21] "not sure"         "3 inches"
```

We can plot a count by question to inspect noncompliance by question. We note that visually it appears that the distribution of NA responses is not even between questions. The ‘Willingness to Pay’ questions (Question #1 and Question #3) have fewer NA responses compared to our ‘Willingness to Accept’ questions (Question #2 and Question #4). This may be attributed to the fact that the willingness to accept questions were fundamentally harder to understand, which may have impacted our randomization and must be accounted for in our final analysis.



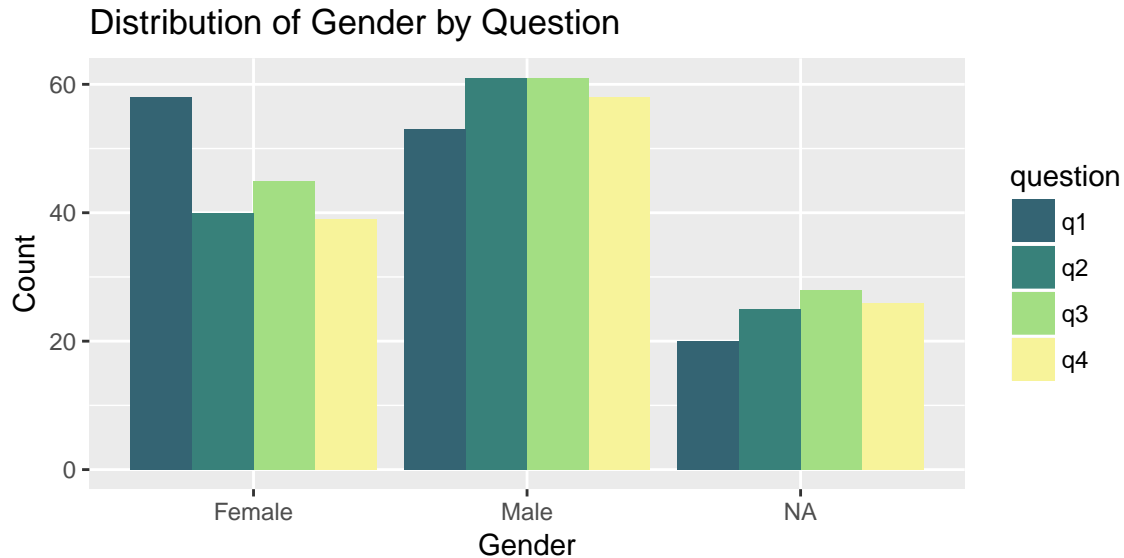
Additionally we conduct a `t.test` and `cohen.d` test to compare the compliance rate of our treatment and control groups. The results of the t-test show a statistically significant difference between the compliance rate while the Cohen’s D shows a small effect size between the two groups. In this study we are mainly concerned with the complier average causal effect (CACE). Simply looking at the intention to treat (ITT) would give these nonsense responses some weight, while the CACE give us the effect for those who actually understood the question and answered the question in compliance with our survey request of “Enter your answer in US \$”.

```
##
## Welch Two Sample t-test
##
## data: d$compliance[d$treatment == 1] and d$compliance[d$treatment == 0]
## t = -3.0802, df = 382.93, p-value = 0.002218
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.08608726 -0.01900442
## sample estimates:
## mean of x mean of y
## 0.9325843 0.9851301
##
## Cohen's d
##
## d estimate: -0.2667107 (small)
## 95 percent confidence interval:
##      inf      sup
## -0.43748608 -0.09593523
```

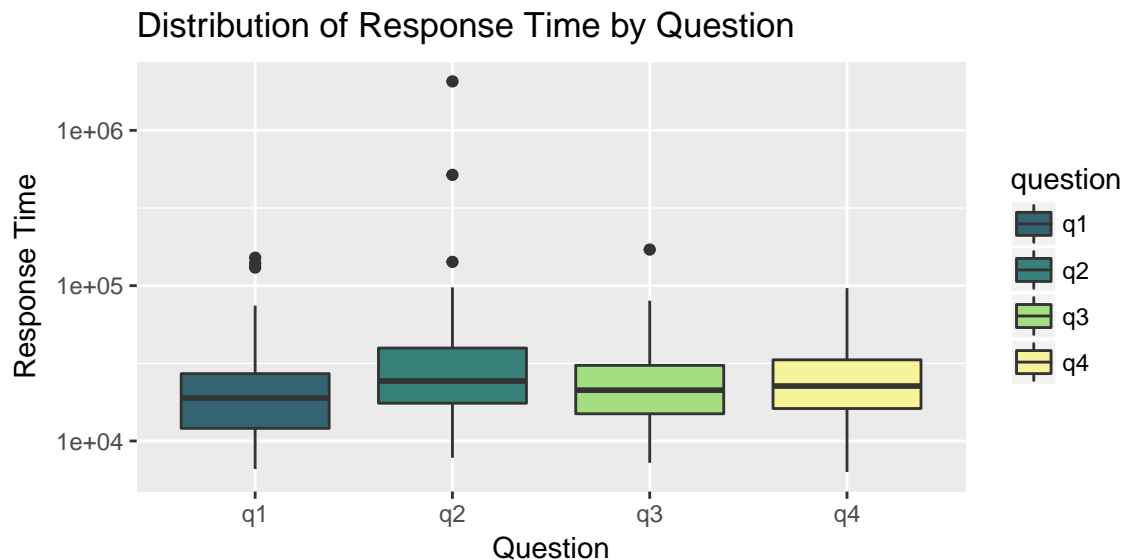

Check for Randomization

To check for randomization we will compare the responses between questions. If randomization was done correctly there should be no statistically significant difference between the response rates for each question. First we will look at the results visually.

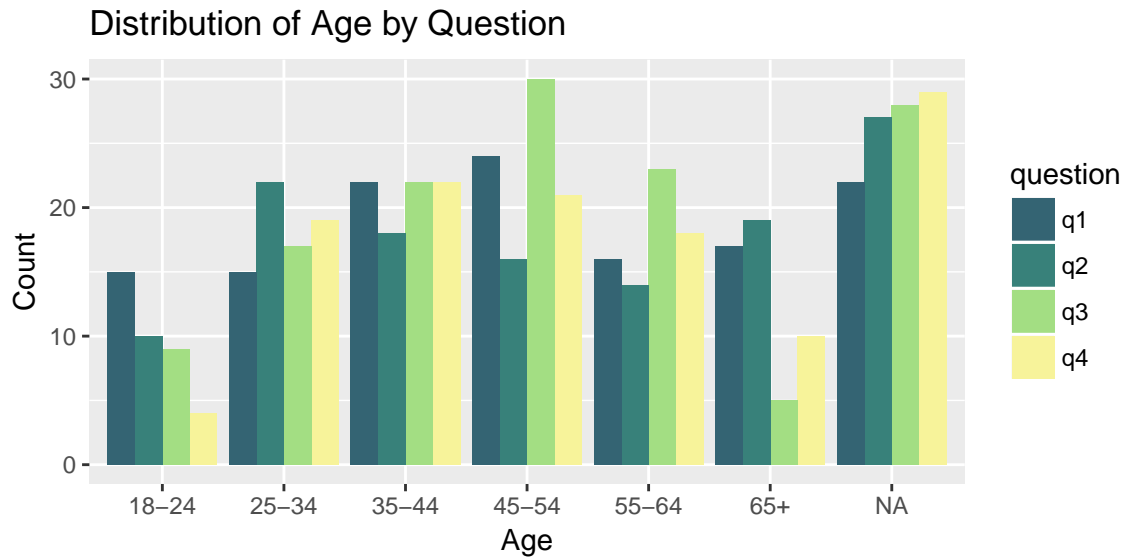
The distribution of gender per question shows a slightly less number of known female respondents for all questions except question #1.



We can compare how long it took respondents to complete the question by using a boxplot. Note the y-axis is on a logarithmic scale. Question 2 appears to be the only question with a slightly longer average response time.



Exploring the age distribution of respondents we see no obvious bias. The number of respondents in the 18-24 age bracket appears to have the largest difference between questions. With the small number of respondents we cannot say this is statistically significant.



ANOVA test

Next, we create a subset from our dataset and remove any observations that do not include the categorical data we are leveraging to check covariate balance.

We isolate the first question set's observations and check if any categorical factors had an effect on whether treatment was assigned.

```
## Analysis of Deviance Table
##
## Model 1: treatment ~ gender + region + age
## Model 2: treatment ~ 1
##   Resid. Df Resid. Dev  Df Deviance Pr(>Chi)
## 1      199      278.43
## 2      209      290.44 -10  -12.006  0.2847
```

The result is a non statistically significant p-value of 0.3341 which suggests the randomization design was not violated.

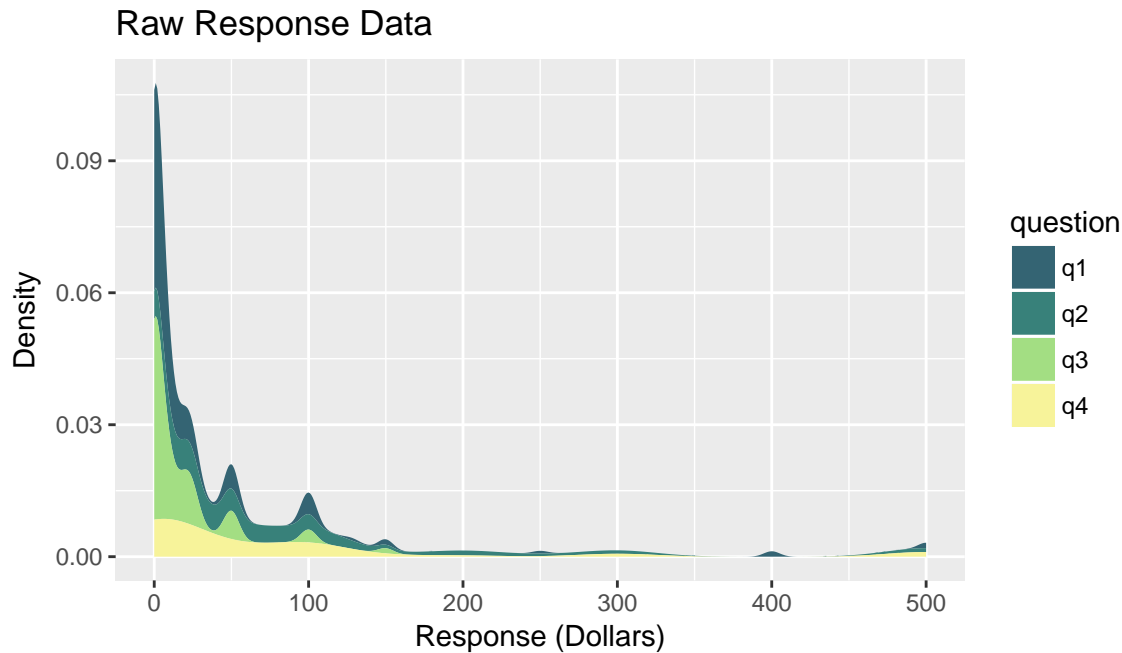
We repeat this on the second test:

```
## Analysis of Deviance Table
##
## Model 1: treatment ~ gender + region + age
## Model 2: treatment ~ 1
##   Resid. Df Resid. Dev  Df Deviance Pr(>Chi)
## 1      190      268.68
## 2      200      278.04 -10  -9.3636  0.498
```

Which also results in a non statistically significant p-value of 0.4674.

Transformations

Due to the continuous user input there are a handful of extreme outliers, particularly in the WTA pool (e.g., willing to accept \$10,000+).



The responses of the raw data resulted in an average user WTP as \$9126012 (\$5 median) and WTA \$794316 (\$75 median) for the first question set (recline). WTP as \$15.27 (\$1 median) and WTA \$8259.81 (\$20 median) for the second question set (legroom).

At first, we considered removing these observations altogether since they are disruptive to the response mean and likely an unconsidered response. And perhaps they are. However, after discussing amongst ourselves we decided we ultimately will treat these responses as valid data despite our reservations. For the purposes of our analysis, we are concluding these users are unwilling to negotiate an accept value which is a legitimate response. Although this decision compromises the usefulness of our data as is. In light of these considerations, our data required transformation.

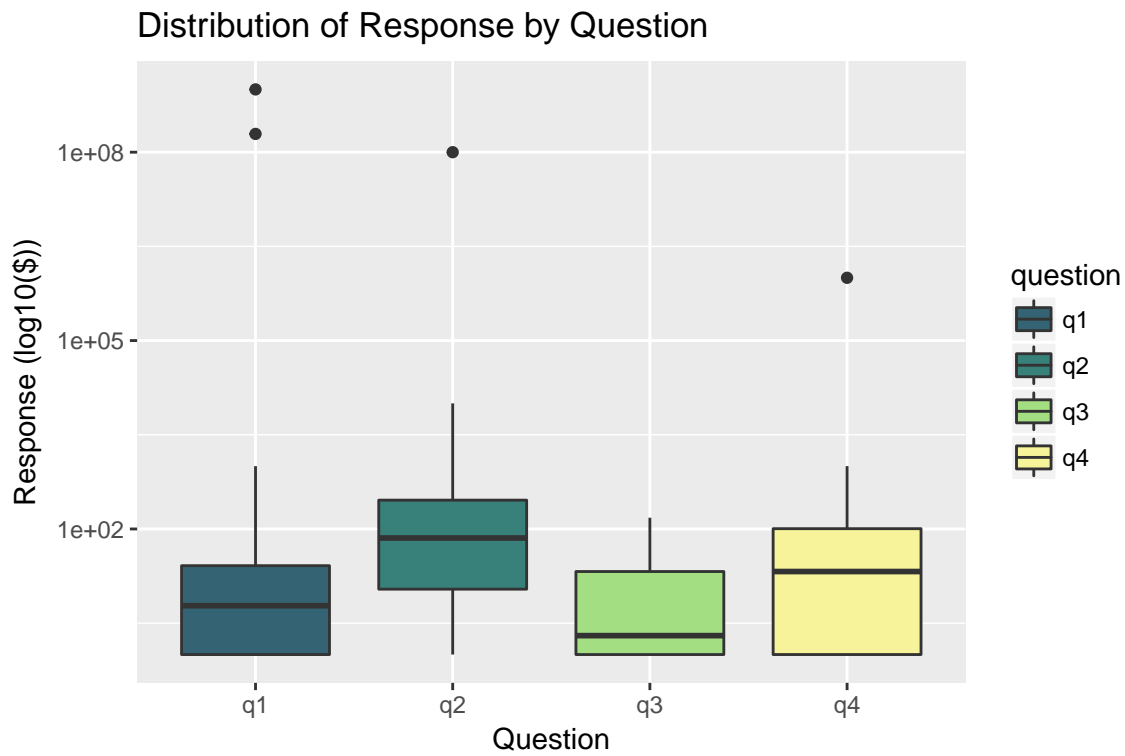
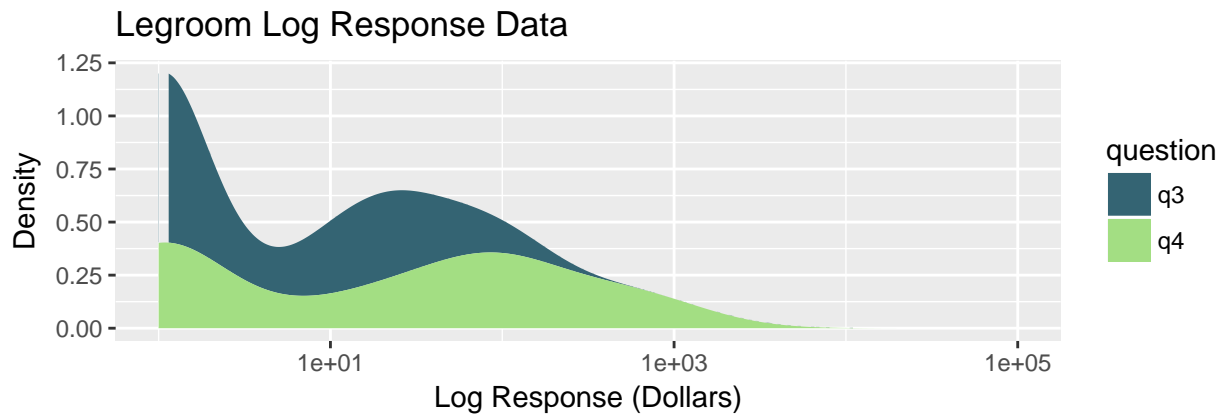
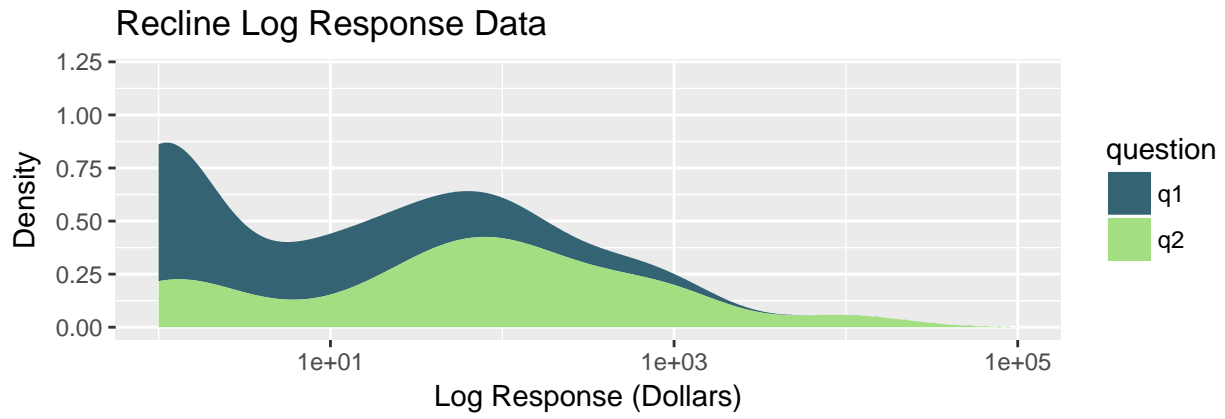
To get a better sense of the response data relative to itself we explored the data using nonparametric statistics.

Resultantly, a bimodal distribution was unveiled. Collectively, the largest density of responses were subjects who answered \$0 which resulted in one of the two humps. The second is due to the subjects in the treatment (WTA) consistently responding in a higher max value than their control (WTP) counterparts. The first question set identified the average WTP subject in the 0.381 percentile vs the 0.632 percentile for WTA subjects. 0.427 percentile for WTP and 0.595 percentile for WTA in the second question set

In addition to better understanding the distribution, both question sets were highly statistically significant using the nonparametric Wilcoxon Rank-sum test model. Despite the promising results, our analysis still left us without an answer for the average a subject is willing to pay in comparison to what they're willing to accept.

To better answer that question, we used a log transformation on the data.

```
# Stacked plot
grid.arrange(log_recline, log_legroom, ncol=1, nrow=2)
```



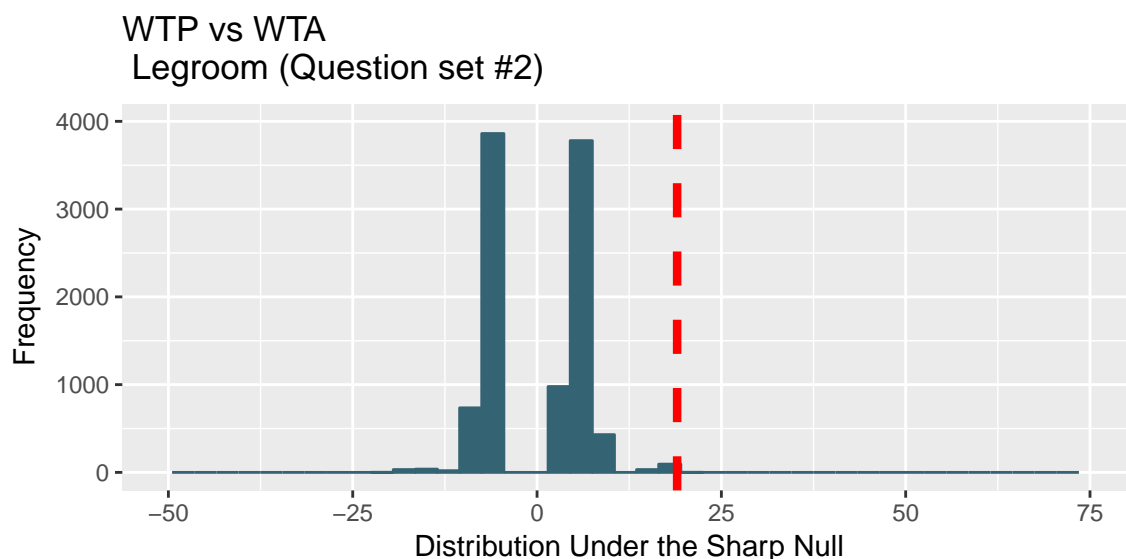
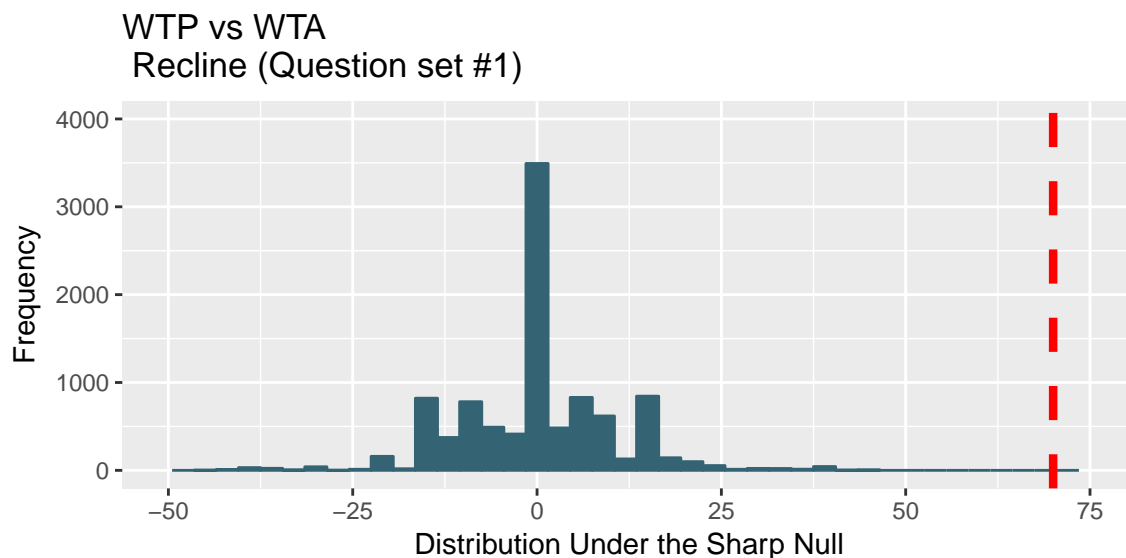
V. Results & Discussion

Given the non-parametric distribution of our data we first analyzed our data using a Wilcoxon rank sum test.

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: d$response[(d$recline == 1) & (d$treatment == 1)] and d$response[(d$recline == 1) & (d$treatm
## W = 12232, p-value = 1.247e-11
## alternative hypothesis: true location shift is not equal to 0
##
## Wilcoxon rank sum test with continuity correction
##
## data: d$response[(d$recline == 0) & (d$treatment == 1)] and d$response[(d$recline == 0) & (d$treatm
## W = 10858, p-value = 5.68e-06
## alternative hypothesis: true location shift is not equal to 0
```

Our results showed that for both our recline and legroom question sets we reject our null hypothesis that there is no shift in rank. These results gave us evidence to reject the null hypothesis that assigned ownership does not have an impact on the value of an object, but failed to show the magnitude of the effect.

We wanted to quantify the amount of difference observed between subjects beyond simply stating their rank was different. Given the skewed distribution of our observations and the large difference between the mean and median in each respondent group we decided to analyze our results using randomization inference on the median response.



Subjects who answered our WTPxRecline question (q1) had a median response of (\$5) and our WTAxRecline (q2) subjects had a median response of (\$75). The difference in median of \$70 (15:1 WTA:WTP) is statistically significant with 0 of 10000 draws showing a larger difference (p value of 0).

Similar to our recline question set, subjects who answered our WTPxLegroom question (q3) had a median response of \$1 and our WTAxLegroom (q4) subjects had a median response of (\$20). The difference in median of \$19 (20:1 WTA:WTP) is statistically significant with 46 of 10000 draws showing a larger difference (p value of 0.0046).

Given the statistically significant results of both the recline and legroom question sets we can reject our null hypothesis that assigned ownership does not have an impact on the value of an object. We observed that the ratio of WTA:WTP ranged from 15:1 for reclining and 20:1 for legroom.

Our 2x2 design allows us to set up hypothetical pairs of real world ‘negotiations’ between a recliner and the person behind them. We used the same Randomization Inference methodology for the medians of q1 vs q4 (WTPxRecline vs WTAxLegroom) and q2 vs q3 (WTAxRecline vs WTPxLegroom). As the endowment effect would predict, we observe that the WTA individuals have a statistically significant higher median minimum amount required than the median WTP. WTAxlegroom individuals wanted a median minimum payment of 20 while the WTPxrecline individuals wanted to spend a median maximum payment of 5. This

difference is statistically significant with a p-value of 0.0233 at an observed 4:1 WTA:WTP. WTAxrecline individuals wanted a median minimum payment of 75 while the WTPxlegroom individuals wanted to spend a median maximum payment of 1. This difference is statistically significant with a p-value of 0 at an observed 75:1 WTA:WTP. This has interesting real-world implications in that the object, legroom/reclining in this study, has different negotiated values depending on who is given the ownership - even though rational-agent economic theory suggests individuals should value the object identically!

VI. Conclusion

Our results show that we can reject our null hypothesis that assigned ownership does not have an impact on the value of an object. While this is exciting, we need to be cautious about how far to extend the implications of these results. There does not exist a standing framework with airlines to allow passengers to negotiate seat privileges. We would anticipate real world tests to have slightly different outcomes than our hypothetical negotiation tests due to additional behavioral economic interactions, such as price setting (from overheard negotiations). That said, the intent for this study was to evaluate how assigned ownership in a question causes an individual to price an object differently, not how to help airline best set seat pricing.

The statistically significant results we observed have implications outside of the airline industry. For example, voters are often asked through ballot measures to vote on approval of public work projects or government services. For many voters this decision includes determining in their mind what the price of the good or service is and if the proposal is an acceptable deal. Our research would indicate that the way the question on the ballot is posed could have a material impact on the outcome. Consistent with our results, we would anticipate voters to have a higher assumed value to objects or services they perceive to “own”, such as a park or river. Just like the reclining seat in our study, these objects or services on ballots rarely have a known market price. Moreover just like our subjects, voters are simply given a question and rarely involved in the negotiation to get to the ballot-stated value. While our results lack a real world mechanism for seat negotiations they have strong implications for similar real-world situations where individuals are asked to determine the value of a good or service.

Our work gives strong evidence that assigned ownership changes how an individual perceives the worth of an object. While we can’t extend these results too far, future research into how an individual comes to believe ownership would be interesting. Additionally, given our positive result it would be interesting to see how assigned ownership interacts with political beliefs and if this effect extends into how individuals vote.

References

1. Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler. “Experimental tests of the endowment effect and the Coase theorem.” *Journal of political Economy* 98.6 (1990): 1325-1348.
2. Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler. “Anomalies: The endowment effect, loss aversion, and status quo bias.” *The journal of economic perspectives* 5.1 (1991): 193-206.
3. Ericson, Keith M. Marzilli, and Andreas Fuster. “The endowment effect.” (2014).
4. Morewedge, Carey K., and Colleen E. Giblein. “Explanations of the endowment effect: an integrative review.” *Trends in cognitive sciences* 19.6 (2015): 339-348.
5. Horowitz, John K., and Kenneth E. McConnell. “A review of WTA/WTP studies.” *Journal of environmental economics and Management* 44.3 (2002): 426-447.
6. Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. “Salience in experimental tests of the endowment effect.” *The American Economic Review* 102.3 (2012): 47-52.
7. Price, Curtis R., and Roman M. Sheremeta. “Endowment origin, demographic effects, and individual preferences in contests.” *Journal of Economics & Management Strategy* 24.3 (2015): 597-619.