

**1. Experiments on Amazon Mechanical Turk (3 points).** Go to the [Amazon Mechanical Turk website](#). Either sign in as a worker, if you already have a Mechanical Turk account or click "Get Started with Amazon Mechanical Turk" and then click "Create a Worker account" under the "Make Money" icon. This worker account will let you browse the different options currently available on MTurk. Once signed in, select "HITs" at the top of the window. These are "Human Intelligence Tasks." Most of these are jobs in which Amazon workers do some manual classification. A smaller number are surveys. The following questions have to do with experiments.

**NOTE:** I am an international student under an F-1 visa which means that according to U.S. federal law to work off-campus or do any remunerated task off-campus I will have to get permission to do it. This permit can only be granted after the first academic year has passed, and I am currently in my first year, so I cannot get permission to do any remunerated work off-campus.<sup>1</sup>

It is possible to argue that if I apply or just sign in into Amazon Mechanical Turk, a violation of my student status has not occurred and the only scenario in which a violation will exist is if I do a task and actually get paid for it. However, under immigration law, the enforcement agents have a great amount of discretion, which means that in the end just submitting my application could lead to complicated situations. And as we have seen in the last months even when we have an explanation, law enforcement agents use an action in a decontextualized way to apply the "law."

Of course, I am assuming an extreme position, and probably nothing will happen by submitting my application. I am aware that the scenario I am presenting is one with almost zero probability of happening, but I think we can agree that the things that have been happening in the last months were unthinkable in the past. We are living in a time in which family separation is allowed, the mobilization of the U.S. military forces to block the entry of an immigrant caravan looking for asylum is justified, and a discourse under which immigrants are delinquents has been normalized in national television and the federal government.

These are extremely difficult times for all immigrants in the U.S., and I truly believe that special times required special sensitivity. I think that this assignment did not take into consideration the context in which we are today. Let's take what we have learned in this class: the information we enter into web pages leaves a record. This assignment requires us to submit our social security numbers and our immigration status in the U.S. to a private company, asking to be allowed to enter a system in which we will be doing things that violate the rules under which we were accepted in the U.S.

Once again, we could argue that Amazon will never release that data to the government. But history teaches us that sometimes private-companies partner with governments to carry out xenophobic projects. With this I am not saying that something like this will happen, once again I think that all these scenarios have an extremely low probability of happening. My point with this note is to highlight that innocent assignments can create unnecessary risks to vulnerable populations, and that it is easy to forget the risks when we are not the ones facing them.

As you understand after the explanation presented above, I do not feel comfortable applying to Amazon Mechanical Turk because I consider I will be taking an unnecessary risk. Even if I am mistaken about the consequences of signing in under my current immigration status, you can see that this created unnecessary anxieties that could have been easily avoided. I tried to get access to the webpage by using someone else's account, but the persons that agreed to share their information (username and password) were not approved. In light of these complications, I decided to google pictures of MTurk and read the instructive that MTurk

---

<sup>1</sup> <https://www.uscis.gov/working-united-states/students-and-exchange-visitors/students-and-employment>

publishes to answer the questions. I understood the assignment as a way to get familiarized with a tool that is currently used to carry out experiments, so I tried to do it using the tools that I found available online.

(a) Search for an experiment on MTurk that interests you. (Hint: You might want to “Search all HITs” for a job or project that has the word “experiment” in it.)

The screenshot shows the Amazon Mechanical Turk interface. At the top, there's a navigation bar with 'Your Account', 'HITS', and 'Qualifications'. A banner indicates '227,886 HITs available now'. Below this is a search bar with 'Find HITs containing' and a filter for 'that pay at least \$ 0.00'. A yellow box contains an announcement about direct deposit to bank accounts. The main section is titled 'All HITs Available to You' and shows '1-10 of 273 Results'. The results are sorted by 'Reward Amount (most first)'. The first five results are listed in a table:

Requester	HIT Expiration Date	Reward	HITs Available
Mark Littlewood	Mar 2, 2015 (2 days 22 hours)	\$14.00	1
Julia Ann Mossbridge	Feb 28, 2015 (23 hours 11 minutes)	\$5.00	1
Melissa	Mar 5, 2015 (6 days)	\$2.00	1
Melissa	Mar 4, 2015 (4 days 22 hours)	\$2.00	1
Turk Experiment	Feb 27, 2015 (59 minutes)	\$1.20	7

Source: [http://1.bp.blogspot.com/-m9\\_5a9mbEsM/VPcWHoa9yFI/AAAAAAAAABNY/LVE5sqQPjBM/s1600/Mturk.jpg](http://1.bp.blogspot.com/-m9_5a9mbEsM/VPcWHoa9yFI/AAAAAAAAABNY/LVE5sqQPjBM/s1600/Mturk.jpg)

I chose the “Women encouraged! Only new workers...” HIT to do the assignment.

(b) Describe the full payment structure of this experiment. That is, the reward column says an amount, but there is a lot more information available as to what that amount means.

Amazon Mechanical Turk pricing depends not only of the reward paid to each worker.<sup>2</sup> In addition to the reward, Amazon charges a Mechanical Turk fee that is 20% on the reward, and in case there is a bonus amount then the 20% is calculated using the reward + the bonus amount. If the HITs necessary are 10 or more, there is an additional 20% fee on the reward of the workers.

Amazon Mechanical Turk offers options that help to target specific segments of the population. The premium qualifications tool allows requesting specific attributes of the workers, for example, gender, insurance policyholder, language fluency, job function, among others. Each specification has a cost (between \$0.05 and \$0.50), and this fee is charged per approved assignment based on the specific premium qualification chosen.

On the other hand, if a worker who takes the job has demonstrated excellence and she has been awarded the Masters Qualification then an additional fee of 5% of the reward is charged per “master” worker.

(c) Describe any qualifications, eligibility requirements, or restrictions (or lack thereof).

This HIT only accepted persons with android cellphones with flash and a computer. Although it seems like the HIT did not require that the workers were women it encouraged women to take it.

<sup>2</sup> Information taken from this webpage: <https://requester.mturk.com/pricing>

(d) How long does this job take? What is the implied hourly rate (dollars per hour)?

The task takes 12 hours, and the reward is \$5.00 for the entire time, so the hourly rate is approximately \$0.417 U.S. dollars.

(e) When does this job expire?

Because I was looking at an old screenshot, the experiment said it expired on February 28, 2015. Next to the expiration date, it says 23 hours and 11 minutes, which means that the screenshot was taken on February 27, 2015.

(f) What is the most this project would cost the HIT experiment creator if 1 million people participated in the task?

To answer this question, I assume that all the workers that participated (1'000.000) were approved, which means that the creators of the HIT did not reject an assignment. Based on the description of the Mechanical Turk pricing the cost of the HIT would be the following if 1 million people participated in the task:

- a) Rewards: \$5'000.000
- b) 20% Fee: \$1'000.000
- c) Additional 20% fee (No. Workers>10): \$1'000.000

I have no information regarding if the creators used a premium qualification or if the what percentage of the 1'000.000 persons that did the assignment were "masters." So, assuming there were no master qualified workers and no premium qualification was used the total cost of the HIT would have been \$7'000.000 U.S. dollars. If all the workers were masters but no premium qualification then the total cost would have been \$7'250.000 U.S. dollars. If all workers were "masters" and two premium qualification, like Gender-female (\$0.50) and household income \$25.000-\$49.999<sup>3</sup> (\$0.50), were used then the cost would have been \$8'250.000 (assuming all assignments were all approved).

**2. Costa and Kahn (2013) (4 points).** Read the paper [Costa and Kahn \(2013\)](#). This paper builds off the electricity market studies of [Schultz et al. \(2007\)](#) and [Alcott and Rogers \(2014\)](#). Write your responses to the following questions as a one-to-two-page composition. (a) State the research question of this paper in the form of a question and one sentence? (b) The data for this study came from at least two sources. Name the sources, and describe the data. (c) Define and describe the control group and the treatment group in this study. What was the treatment? (d) Beyond the previous work of [Schultz et al. \(2007\)](#), what extra layer of participant heterogeneity did Costa and Kahn control to answer their research question? (e) What was Costa and Kahn's finding?

Increasing literature on the consequences of political ideology has focused on the role it has in shaping preferences. A study that addresses this point is Costa and Kahn's paper "Energy Conservation "Nudges" and Environmentalist Ideology." In this paper, Costa and Kahn explore the relationship between political ideology and the responses to non-market mechanisms designed to reduce energy consumption. Specifically, the authors used the data of an experiment carried out by Positive Energy in 2008, to evaluate if people from different political ideologies respond different to energy conservation nudges, and If they do respond differently, how political ideology shape the consumption of energy.

The experiment Costa and Kahn (2013) used was done in a western utility district. The experiment consisted in sending a two-page report called the "Home Electricity Report" (HER) to a treatment group and not sending the report to a control group. Each report gave two pieces of information in the first page to each household in the treatment group: i) the absolute consumption level of electricity of the household and ii) a comparison of its consumption with 100

---

<sup>3</sup> I used these 2 premium qualifications based on the encouragement of women to take it and a household income that guarantee the person has a cellphone and a computer.

neighbors living in similar-sized homes, with same heat type and with neighbors who were in the bottom 20<sup>th</sup> percentile of electricity usage. In the back page, the report presented a comparison between the household's electricity usage in the current month relative to the same month the prior year. Additionally, the report gave "green stars" to those households that decrease their energy consumption in comparison with the previous month. The first report also included an additional message telling the household if they were doing "great," "good" or they had "room for improvement" in their energy consumption saving.

Before explaining the assignment of the households to the treatment and control groups, it is necessary to clarify the selection criteria that was used to choose the households. The authors mention that the HER experiment selected households from 85 census tracts with a high density of single-family homes. To be eligible to be part of the experiment the household should meet the following criteria: i) current electric account activate for at least one year; ii) not be living in apartment buildings, and iii) living in a house with square footage between 250 and 99,998 square feet.

After filtering by the eligibility requirements, 49,000 households were selected. From these 49,000 households, groups of contiguous census blocks were randomly assigned to the treatment or control group. According to the authors, the HER experiment selected what they called a "block batch" of five contiguous census blocks and randomly assigned them to the treatment group and the next contiguous "block batch" was assigned to the control group. This process was repeated until 35,000 households were assigned to the control group and treatment group, and the remaining 14,000 households were assigned to the control group.<sup>4</sup>

To estimate the effect of political ideology on energy consumption saving, the authors used X datasets. First, their primary data set consisted of residential billing data from January 2007 to October 2009. This dataset contained information regarding kilowatt hours purchased per billing cycle, the number of days measured each billing cycle, whether the house used electric heat and whether the household was enrolled to purchase energy from renewable sources. The second dataset the authors used contained climate conditions to control for differences in weather conditions.

The third dataset used in this paper included information on when the household began to receive the report, square footage of the house and the age of the house. By merging the billing data to the treatment and control data, their final dataset contained 81,722 observations, with 48,058 households in the control group and a treatment group divided into two subgroups: 24,028 households that received a monthly report and 9,636 that received a quarterly report.

A fourth dataset used by the authors contained the voter registration and marketing data for March 2009 to the individuals in the primary dataset. This dataset gave the authors information about the party affiliation and whether the individuals in those households donated to environmental organizations. By linking this data to the information mentioned above, the authors were able to link half of their sample to the voter registration data, allowing them to evaluate the impact of political preferences in the consumption of energy. This dataset was crucial in the analysis done by Costa and Kahn (2013) because having a measure of political preference allowed them to make an additional contribution to the study done by Schultz et al. (2007). However, the authors failed to show conclusive results that show that a normative assessment has different effects also depending on the political ideology of the household.<sup>5</sup>

Costa and Kahn (2013) measured ideology in different forms, by political party affiliation, donations to environmental organizations, or if the household purchased renewable energy. Their results show an overall robust effect of political ideology on energy consumption saving. The authors estimate seventh regressions. The results presented in table 3

---

<sup>4</sup> The authors explain that the reason to select contiguous block groups was taken because the implementation contractor believed that peer effects could lead to greater energy saving. However, in a foot note Costa and Kahn explain that according to 2009 Home Energy Use survey the HER report did not generate discussion and if there were peer effects those operate through implicit social pressure.

<sup>5</sup> In their paper Schultz et. al (2007) evaluate a similar experiment done in San Marcos, California in 2009. The results of Schultz. Et al (2007) reveal that a normative assessment of the energy consumption affects future energy consumption.

show that conservatives decrease their average daily consumption by 1.7% while liberals decrease it by 2.4% in response to the treatment. Additionally, when using purchasing of green energy, they found that the household that purchases this type of energy decreased their consumption by 0.9% in comparison to those households that do not purchase green energy. Finally, the results related to the third measure of ideology, donating to environmental organizations, show that the household that donates money to those organizations decreased their consumption by 1.1% in comparison to those that do not donate.

When allowing for heterogeneous treatment effects, the results for some of the ideology measures are more robust than others. For example, when the authors control for a fraction of liberals in a block group the only measure that remains statistically significant is donation to environmental organizations, and joint variable of donating and paying for green energy remains statistically significant too. Other controls that are used by the authors are: characteristics of the house (bigger, older houses), household income, value of the house, renter status, quintile/quartiles distribution of liberal households, the period in which the household is treated (to test for persistence over time). The authors did not find a significant effect of renter status but found significant effects on the other variables.<sup>6</sup> For the household income and home value, the researcher found that it does not affect the response of liberals to the treatment.

Another estimation that the authors do is related to the decision of the treatment households to keep receiving the report. They found that conservatives are more likely to opt out of receiving the report, and these results are also robust when they allow for heterogeneous treatment effects. The authors also presented results using joint variables combining multiple measurements of political ideology (i.e., registered conservative, pays for renewable energy and donates to environmental groups vs. registered liberal, pays for renewable energy and donates to environmental groups). In conclusion, the main results behind this paper are that nudges need to be targeted to be most effective, specifically energy conservation nudges.

### References

Costa, D. L., & Kahn, M. E. (2013). Energy conservation “nudges” and environmentalist ideology: Evidence from a randomized residential electricity field experiment. *Journal of the European Economic Association*, 11(3), 680-702.

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological science*, 18(5), 429-434.

**3. Analytical exercise (3 points).** This is exercise #19 at the end of [Salganik \(2018, Ch. 4\)](#) with a slight addition. A good reference for how to answer this question is [Salganik \(2018, pp. 203-209\)](#), the sections entitled “Potential Outcomes Framework” and “Precision.” A new experiment aims to estimate the effect of receiving text message reminders on vaccination uptake. One hundred and fifty clinics, each with 600 eligible patients, are willing to participate. There is a fixed cost of \$100 for each clinic you want to work with, and it costs \$1 for each text message that you want to send. Further, any clinics that you are working with will measure the outcome (whether someone received a vaccination) for free. Assume that you have a budget of \$1,000.

(a) Under what conditions might it be better to focus your resources on a small number of clinics and under what conditions might it be better to spread them more widely?

In his book “Bit by Bit,” Salganik explains that an important assumption to estimate the effect of a treatment using causal inference and experiments is the Stable Unit Treatment Value Assumption (SUTVA). This assumption implies two things. First, the only thing that affects the outcome of individual  $i$  is whether she has been treated or not. And second, that the only relevant treatment is the one that the specific researcher is delivering.

---

<sup>6</sup> House characteristics: increases of 0.1 in the fraction of liberals decreases the consumption by 0.4%; quintiles: “greater liberal share relative to the lowest quintile is associated with households’ reducing their electricity consumption in response to the treatment” (Costa and Kahn 2013, p. 693).

If we observed that there are no reasons to believe that the SUTVA assumption will be violated in a small number of clinics it would be acceptable to choose this small number. However, if we believe that we will have spillover effects (i.e., patients tend to live closer to the clinics they go, so they have a higher probability to be clustered in the same neighborhoods if I choose a small number of clinics. This proximity can increase the probability that the patients in my treatment and control groups talk among them and then treated individual  $i$  can talk to the not treated individual  $j$ , which creates spillover effects).

Moreover, if I am in an area in which several studies are done, it would be a good idea to spread the clinics more widely. For instance, in some regions of developing countries researchers tend to carry several experiments. In some cases when a group of researchers arrives at a region, a previous study has been done, and the persons of the study can be affected by previous or current treatment. (i.e., in an extremely poor region in a country in Africa a group of researchers developed a randomized control trial, the evaluation of the treatment depended on what the community thought was extremely important. The result of the study was that these persons living in extreme poverty considered that the most important thing was global warming. After doing some research, the social scientists discovered that another experiment related to climate change was done in the same region and that affected the results of the study). In light of this, if we are doing the experiment in a period in which several other reminders of vaccination are active, or we believe that other “treatments” could be affecting our selected population, then we should spread the clinics more widely.

Finally, if the number of people that received the treatment affects the effect, then we need to consider if applying the treatment to a higher population increases the probability that spills over exists.

(b) What factors would determine the smallest effect size that you will be able to reliably detect with your budget?

According to Salganik (2018), our capacity to determine the smallest effect size depends on the amount of variation. We want to minimize the standard error to increase the efficiency of our estimations and detect the smallest effect size. Following the strategy proposed by List, Sadoff and Wager (2010), I would need to reduce the variance of the unobserved components to decrease the effect size. To do this, I have to divide my experimental units into subgroups or blocks using observable variables to do that division.

Some of the observable variables that I would use to do the block division are age (age ranges: less than 5, 5-15, 15-20, 21-25, 26-30, 31-35, 36-40, 41-45, 46-50, 51-55, 56-60, and 61+); gender (woman, man or other); pre-existing diseases (i.e. disease that affect the pre-disposition of getting a particular disease related to the vaccination reminder); if the individual is a woman if she is pregnant (Regan et al. 2017).

This segmentation of the population that I will be studying will allow me to reduce the noise on my outcome data in a not highly costly way and therefore detect a relatively small effect using the budget I was assigned.

References

List, A. J., SadoFF, S., & Wagner, M. (2010). WSo You Want to Run an Experiment, Now What? Some Simple Rules of Thumb for Optimal Experimental DesignX.

Regan, A. K., Bloomfield, L., Peters, I., & Effler, P. V. (2017). Randomized controlled trial of text message reminders for increasing influenza vaccination. *The Annals of Family Medicine*, 15(6), 507-514.

Salganik, M. J. (2018). *Bit by bit: social research in the digital age*. Princeton University Press.