Essays on Field Experiments in Behavioral Economics

by

Linfeng Li

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Information)
in the University of Michigan
2022

Doctoral Committee:

Professor Tanya Rosenblat, Chair
Professor Yan Chen
Professor Maggie Levenstein
Professor David Miller

Linfeng Li
llinfeng@umich.edu
ORCID iD: 0000-0002-1119-8740

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

This dissertation contains three essays in behavioral and experimental economics based on three field experiments. Chapter 1 provides an overview. The discussions in Chapter 2 and Chapter 3 are both framed around the provision of digital public goods. Chapter 4 covers an audit study conducted on Indeed.com.

Chapter 2 examines Wikipedia's gender gap using a student population. Upon self-replicating the experiment for three consecutive years, we do not find gender differences in the entry stage. This resolves the concern that the gender gap in editing Wikipedia is introduced when new editors are first invited to edit on Wikipedia and calls for a more targeted effort to enhance the retention of female editors on Wikipedia.

Chapter 3 documents a large-scale field experiment in which domain experts (economists) were invited to provide metadata information for their datasets hosted on a public data and code repository, openICPSR. We find that reminding the experts in this experiment of the private or public benefits of contributing their metadata is not as effective as simply explaining why the metadata are being collecting, the benefits of which include the enhanced findability of the dataset and potential future reuse opportunities.

Chapter 4 is based on an audit study conducted before the pandemic at a time when the unemployment rate was at an historic low and gig work opportunities were beginning to become a popular on-demand alternative work opportunity for the general population. Targeted at the population of low-skilled workers, this experiment documents that having gig-work experience during unemployment spells does not significantly reduce the likelihood of receiving a callback during a job search process.

CHAPTER 1

Introduction

Field experiments are known to be reliable in testing the otherwise untestable hypotheses in a realistic setting. Typically, compliance and possible spillover of treatments are the main reasons to ensure the validity of the random assignment of treatment conditions, which directly impacts the internal validity of the experiment. This dissertation presents three field experiments that share one common feature: over thousands of treated units, the treatments are precisely assigned at the participant level, and outcome data are collected at the same granularity.

1.1 Overview

Chapter 2, coauthored with Yan Chen, David Miller, Muriel Niederle and Chang Ge, was motivated by the observation that there is a large and persistent gender gap on Wikipedia; i.e., female editors make up less than 10% among all editors who contribute to Wikipedia. To understand why there is a gender gap among these editors, we attempt to reconstruct the gender gap by inviting undergraduate students to edit on Wikipedia. In this experiment, each student was assigned to edit two articles on Wikipedia, which were randomly assigned. In preparation for the experiment, we built two collections of Wikipedia articles about economic concepts and biographies for economists, each of which contained hundreds of articles. We administered the article assignments by directly emailing the articles to the participating students. In a given iteration of the experiment, all students in the experiment received a unique pair of links, and they could choose either one to edit as part of a bonus point activity. Upon repeating the experiment three times, we instead found a robust null result where female students were as likely as male students to edit Wikipedia articles. Furthermore, to understand the mechanism that drives the decision to edit (or not to edit), we collected a debriefing survey in which each student not only reviewed exactly what assignment he/she received at the beginning of the experiment but also reported the level of expertise and competency he/she thought were needed to edit these Wikipedia articles. We found that the perceived levels of expertise and competency elicited from the debriefing survey predicts the editing decision in a significant manner.

In Chapter 3, coauthored with Yan Chen, Maggie Levenstein and Lars Vilhuber, we investigate what motivates domain experts to contribute to metadata. The experts invited to participate in this experiment were the authors of studies that underwent data migration, but the metadata for the studies were missing due to the lack of relevant infrastructure before the migration. With a relatively straightforward encouragement design, we used three types of motivational messages to solicit metadata contributions and found that simply reminding experts about why the metadata was being collected was the most effective way to elicit contributions. To prepare for the email campaign, we looked up a total of 4,503 economists by name to obtain their up-to-date email addresses. To collect the metadata for the studies included in the experiment, we built a customized survey interface that featured a precise rendering of the study information for each author. Since the survey links were uniquely generated for each author and distributed to each individual through a hyperlink field in the body of the treatment email, we were able to track contribution decisions at the participant level.

In Chapter 4, coauthored with Tawanna Dillahunt and Tanya Rosenblat, we explore the labor market outcome of gig-work experience. By the time that the study took place in 2019, approximately 20% of the U.S. unemployed population had been out of the labor force for more than 6 months. The rise of the gig economy has changed the landscape of nontraditional employment opportunities for predominantly low-skilled long-term unemployed workers. This particular type of on-demand work can be used to fill unemployment gaps and offers both little to no training costs and flexible hours. Therefore, we explore whether driving as a form of gig work helps to mitigate the negative effects of long-term unemployment for low-skilled job seekers with employment gaps and how employers evaluate workers who have held nontraditional jobs. Using an audit study of 1006 job applications, we evaluated whether a set of resumes "enhanced" with experience driving for a real-time ridesharing service received more callbacks than baseline resumes with an employment gap. We found no evidence that driving as a form of gig work increased the callback rates for the applicants. In fact, we observed that in comparison to the callback rates for men, those for women were slightly lower. This study suggests that driving "gigs" might not be a substitute for traditional employment on resumes for low-skilled workers. In this study, we investigate methods that help us to understand why real-time ridesharing services are not a substitute for traditional jobs in regard to bridging employment gaps and solutions to overcome them.

1.2 Method review

In my dissertation, I have created a few notable tools that are worth mentioning due to their methodological value.

In Chapter 2, I use the MediaWiki action API (application programming interface) to collect the complete editing history of all participants in the experiment who had registered for an account on Wikipedia. The key identifier for this query was the Wikipedia username, and this infrastructure helped to keep track of any subsequent edits that the participants made outside the scope of the experiment. Through this querying tool, we verified that very few participants in the Wikipedia editing experiment actually returned to edit another Wikipedia article after the experiment. Another notable property of Wikipedia articles is that each page is uniquely identified with a "page id". Furthermore, each revision made to the article is also associated with a unique "revision id". Through the MediaWiki action API, I was able to obtain a precise measurement of the Wikipedia articles assigned to each participant. Specifically, with the "revision id" reflecting the version first seen by the participant when they were assigned a pair of assigned Wikipedia articles, I was able to take down precise measurements of the articles assigned to each participant.

In Chapter 3, despite the simplicity of the design, the authors of the migrated studies are shown to have formed a densely connected coauthor network, where the most productive author was involved in a total of 17 studies. Concerns about the spillover and contamination of treatments could easily arise if one author was treated in relation to more than one study or if his/her coauthors were assigned to a different treatment condition. One simple way to address this concern is to randomly draw from the collection of all studies and keep a set of studies where each author shows up exactly once. Given the combinatorics nature of the coauthor network, maximizing the total number of treatment units is an NP-hard problem. To achieve an efficient treatment assignment where the goal is to include as many authors and studies in the experiment as possible, I adopted the state-of-the-art community detection algorithm from network science literature (Traag, Waltman, and van Eck, 2019) and integrated it into a network truncation procedure, where large, connected components in the network are partitioned into components with no overlapping edges. At the end of the network truncation exercise, we had recovered a noticeably larger number of treatment units for the experiment than was possible with simple random draw attempts.

In Chapter 4, I navigate the new challenges posed by the concentration of job postings, where the employers switched to recruiting talent from a few major "job sites", and the job seekers were also convinced to apply for jobs on these "job sites". First, as an improved way to sample job postings, the concentrated posting of jobs made it possible for the experimenters to target representative positions on a daily basis. By using a daily scraper that was aware of all the jobs currently available on the job site, I was able to achieve a strictly balanced random assignment schedule in scheduling the application of jobs using our fictitious accounts. However, the creation of these fictitious accounts made the audit study far less scalable than it would be otherwise. As explained further in Chapter 4, we ended up using eight distinctive job-seeker profiles on Indeed.com to conduct the experiment. Back when job applications were sent through mail services, a printer alone was able to fully randomize all fields as needed in an audit study. This was the first challenge I encountered when attempting to scale up the experiment. Additionally, despite the limited scope of the experiment, which targeted only one "job site", it was extremely hard to automate the job application process. Different employers posted a wide range of questions, which would have made it tedious and error-prone to try to automate the job application process. I ended up applying to all the jobs manually with the help of one research assistant. This second obstacle was hard to circumvent and may have been best addressed with an army of research assistants. Last, I wrote monitoring scripts that kept track of responses received as email and text messages. By making use of the Gmail API and Twilio API, I was able to monitor the callback decisions in real time. The script also sent notifications to the employers stating that the fictitious applicant was no longer interested in the position.

1.3 Discussion

All three experiments presented in this dissertation have been conducted according to the best practices recommended in Christensen, Freese, and Miguel (2019), where all the code for preparing and analyzing the experiment are tracked by version control software. In Chapter 2 and Chapter 3, all hypotheses are pre-registered on the AEA RCT Registry (Li and Chen, 2020; Chen et al., 2020b), following the suggestion from Coffman and Niederle (2015). In this section, I reflect on the experiences following the best practices for conducting transparent and reproducible social science research.

First, although we exerted extensive effort to motivate the hypotheses based on literature and established theoretical results, we were not able to find empirical support for a large fraction of the preregistered hypotheses. In Chapter 2, we did not find support for the main hypothesis that female students are less likely to edit on Wikipedia. In fact, for the student population that we recruited for the experiment, female students were found to be more likely to conduct edits on Wikipedia, based on our records. This reversal was not expected and may at best be interpreted as a special attribute of the specific experimental setting we employed; female students were simply more conscientious and therefore were more likely to conduct an edit on Wikipedia as part of their coursework. In Chapter 3, contrary to what was registered in the preregistration, the participants treated with the control message were more likely to contribute to the metadata by a significant margin. This result is equally puzzling, as the participants in the experiment were economists and thus should be the most sensitive group of experts in terms of responding to various incentives introduced under different treatment conditions. Ultimately, we arrived at the unexpected conclusion that attention played an important role when soliciting metadata contributions. In short, participants who received a shorter message (the control message) were more likely to make contributions to the metadata. Despite the surprise that quite a number of hypotheses were unsupported, since we preregistered the set of hypotheses for each study, we were able to honestly report the findings in the form of a registered report. Additionally, by preregistering the experiment design, the process of conducting the experiments was made very straightforward; since all the key elements of the experiments were spelled out in the preregistration documents, there was little ambiguity when executing the experiment design at all stages.

Second, another important lesson we learned was about self-replication. In Chapter 2, we repeated the same experiment design for three iterations in total. The second and third iterations can be treated as self-replication, as they were introduced after the initial iteration in which we observed a promising gender gap in editing the assigned Wikipedia articles; i.e., women were less likely to edit on Wikipedia when they were assigned to edit articles about economics concepts. Concerns about the statistical power of the experiment played an important role in deciding to conduct the second iteration of the experiment as an exact replication of the first run. Mainly, the gender gap observed in the first iteration did not hold under a regression framework; the authors considered this is only plausible due to the lack of statistical power. In short, the second and third replications were introduced to double the sample size and thus served the purpose of a robustness check. Notably, although the same group of researchers conducted the experiment under the same experiment design in both the subsequent iterations, we were not able to self-replicate the significant gender gap that was observed in the first iteration. In the second iteration, due to the spread of COVID-19, all the students were taking the course remotely. This change constituted a drastically different environment for the experiment; exams were conducted exclusively in a remote manner, and the students were subject to more pronounced stress when compared to the amount of stress present during the first iteration of the experiment. By the third iteration, the mode of teaching remained different from that in the first iteration; under a hybrid teaching environment, the students enrolled in the course experiment may have participated either remotely or in-person. Arguably, the setting for the third iteration of the experiment also differed from that of the first iteration. In conclusion, as documented in Chapter 2, there is a timeliness element present in all the replication attempts; despite our best effort to achieve exact replication, the changes made to the external environment most likely introduced idiosyncratic variations that were hard to isolate.

To conclude this discussion with a review of the implication of timeliness with regard to the field experiments, Chapter 4 provides an in-depth discussion of what was learned the hard way by trying to replicate the well-known audit study method that was first pioneered in (Bertrand and Mullainathan, 2004). Being a versatile experimental method, the audit study approach has been widely adopted as a tool to uncover systematic distribution in the labor market. In conducting the experiment documented in Chapter 4, however, we no longer had the luxury of printing as many variations of resumes as we needed and sending them out through the postal service. By the time we conducted the experiment, jobs were primarily being posted online where applicants are expected to first register for an account on the platform and then complete a number of forms in order to submit their job application. Although I was able to conduct the audit study on a small scale using a collection of specialized tools for scraping job postings and responding to callbacks, due to unforeseeable changes related to both the websites on which jobs are posted and email/SMS service providers, in a few years from now, these feasibility concerns will need to be addressed afresh since all of the required specialized tools are very much specific to the context at the time at which the experiment is conducted.

References

Bertrand, Marianne and Sendhil Mullainathan (2004). "Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination". In: *American Economic Review* 94.4, pp. 991–1013.

Chen, Yan et al. (2020b). "Motivating Metadata Contributions for Data Re-Use and Reproducibility". In: *AEA RCT Registry*. DOI: 10.1257/rct.6159-1.0.

Christensen, Garret et al. (2019). *Transparent and Reproducible Social Science Research: How to Do Open Science*. University of California Press. DOI: doi:10.1525/9780520969230.

Coffman, Lucas C. and Muriel Niederle (Sept. 2015). "Pre-Analysis Plans Have Limited Upside, Especially Where Replications Are Feasible". In: *Journal of Economic Perspectives* 29.3, pp. 81–98. DOI: 10.1257/jep.29.3.81.

Li, Linfeng and Yan Chen (2020). "Gender Inequality in Contributions to Wikipedia". In: *AEA RCT Registry*. DOI: 10.1257/rct.6753.

Traag, V. A. et al. (Mar. 26, 2019). "From Louvain to Leiden: Guaranteeing Well-Connected Communities". In: *Scientific Reports* 9.1 (1), p. 5233. DOI: 10.1038/s41598-019-41695-z.

CHAPTER 2

Gender (In)equality in Contributions to Wikipedia

2.1 Introduction

Wikipedia, often praised as "a sum of all human knowledge" (Mesgari et al., 2015), is an online encyclopedia created by volunteer editors as a public good. A surprising finding about this crowdsourced production is that Wikipedia editors are predominantly men. Ford and Wajcman (2017), along with the existing body of literature they surveyed, reported that less than 10% of Wikipedia editors are women. Despite knowledge of this fact, the understanding of the cause of this gap is limited.

One conjecture about the cause of this gender gap is that women have lower internet skills. Hargittai and Shaw (2015) explored the impact of internet skills on students' likelihood of editing Wikipedia. In a survey of first-year college students enrolled in a mandatory university-wide writing class, they found that female students were less likely to edit a Wikipedia article and that "the gender gap in editing is exacerbated by a similarly significant Internet skill gap." In a later study, Shaw and Hargittai (2018) investigated the contribution pipeline to Wikipedia by surveying a national representative sample and found that "a gender gap emerges only at the later stages in the pipeline[1], in terms of who knows the site can be edited and who has edited the site."

Although the Internet skill explanation is plausible, it does not explain why wikiHow, which uses the same MediaWiki software as Wikipedia, has no gender gap among its contributors. In fact, more than 50% of contributors to wikiHow are women (Johansson, 2013). We conjecture that Wikipedia is an online encyclopedia and therefore potential editors need to perceive themselves as having domain expertise and authority before participating in editing, whereas wikiHow is a how-to site that everyone who knows how to do something can feel comfortable editing.

[1] In early stages of the pipeline, there was no gender gap in having heard of Wikipedia and having visited the site.

In this paper, we investigate Wikipedia's gender gap by conducting a field experiment, where we recruit male and female students in an intermediate microeconomics class and assign Wikipedia articles for them to edit. By observing whether female students would disproportionally choose not to edit on Wikipedia, we can document whether gender gap is introduced at the entry stage.

Our research design is inspired by recent research on gender differences in the contribution of ideas (Coffman, 2014). In a laboratory experiment designed to study factors that predict an individual's decision to contribute her ideas in a team decision-making environment, Coffman randomizes individuals into groups of two and presents them with questions from a variety of domains, including matters that are stereotypically perceived as male-typed and those that are stereotypically perceived as female-typed. She measures an individual's knowledge of the material, her willingness to contribute answers in a group decision-making setting, and the beliefs about her own and her partner's ability. She finds that the decision to contribute ideas to the group strongly depends on the interaction of gender and the gender stereotype associated with the decision-making domain. Conditional on measured ability, individuals are much less willing to contribute ideas in areas that are gender incongruent. Even very knowledgeable women undercontribute in male-typed domains. These results show that even in an environment where other group members show no bias, women in male-typed areas may be less influential.

In our experiment, building on the natural variations on Wikipedia where there are EconConcept articles and biography articles about economists, we randomly assign two articles of either type for the students to edit. Assuming that there exists a gender gap based on the random assignment for the type of the Wikipedia articles, we can test the hypothesis that women are less likely than men to perform edits when assigned to edit pages related to economic concepts.

We conducted the experiment across three consecutive years and found that *(i)* When introducing students to Wikipedia article editing related to their field of study, we documented that there is no identifiable gender gap at this entry stage. *(ii)* Furthermore, based on the aggregate page-rating information we collected through a debriefing survey, we can attribute the differences in editing decisions to the perceived difficulty of the assigned pages.

2.2 Literature review

Since the first official documentation of Wikipedia's gender gap in Glott, Schmidt, and Ghosh (2010), a growing body of literature has been generated around the topic. Ford and Wajcman (2017) provides a comprehensive review of empirical work on the matter. More concerned with the behavioral attributes of the editor, Bear and Collier (2016) surveyed the English-speaking samples in the Global Wikipedia Survey conducted in 2008 (Glott, Schmidt, and Ghosh, 2010) who were labeled occasional contributors. Based on this selective sample with a total of 43,912 English-speaking participants in the US, a total of 1,598 participants responded to the follow-up survey. In this small sample, female participants made up a total of 17.15% of the respondents, which was comparable to the gender breakdown of the original survey. In an effort to explain why women are less likely to edit on Wikipedia, based on survey measurements about confidence in expertise, Bear and Collier concluded that "women reported less confidence in their expertise, expressed greater discomfort with editing and reported more negative responses to critical feedback compared to men." In this experiment, to further explore the role that confidence plays in the entry decision to edit Wikipedia, we introduce an incentive-compatible measure of confidence and explore the interaction of perceived difficulty of editing on the target page and the self-reported confidence levels.

Research in multiple fields has shown that preferences and cognitive ability measured in laboratory environments predict participant behavior in the field. The risk preferences measured in laboratories are correlated with interpersonal conflicts (Lahno et al., 2015), exposure to violence (Voors et al., 2012), and the adoption of agricultural technologies (Ward and Singh, 2015). Prosocial preferences measured in laboratories can predict people's donations in the field (Benz and Meier, 2008; Laury and Taylor, 2008), trustworthiness measured in laboratories can predict people's trustworthiness in financial situations (Karlan, 2005). On the other hand, laboratory-measured cognitive ability predicts real-world decision outcomes such as missing flights and drunk driving (Juanchich et al., 2015), while competitiveness can predict people's salary and achievement in the real world (Buser, Niederle, and Oosterbeek, 2014; Reuben, Sapienza, and Zingales, 2015).

In this experiment, we employ a battery of games to elicit the preferences of the subjects and take measurements of their cognitive abilities. Preferences include risk preferences, willingness to trust and trustworthiness, and conditional willingness to contribute to public goods. We also measure competitiveness as well as the ability to think strategically and solve complex constrained optimization problems.

2.3 Experiment Design

Participants in this experiment were recruited exclusively from Econ 401, an intermediate microeconomics course at the University of Michigan. Econ 401 is one of the two core economic theory courses at the university and is a prerequisite for a large number of upper-level electives in the Economics Department. Students enrolled in Econ 401 are usually in their sophomore or junior year, with an intention to declare a major in Economics. To incentivize student participation, the instructor introduces the experiment as a series of bonus point activities, with a cumulative weight of 1% of the total total points available in the course.

The experiment was repeated in winter 2019, fall 2020, and fall 2021. We recruited a total of1,126 students, where female students made up 38.72 % of the student population. One typical iteration of the experiment was composed of three main components: (1) a battery of games and a survey conducted before the main part of the experiment, where we recruited the participants and obtained their self-reported gender in the survey; (2) a Wikipedia editing activity, where students received a pair of individualized Wikipedia links to edit; and (3) a debriefing survey, where we collected additional confidence measurements about both the pages assigned for editing and the participants themselves.

In this section, we document the procedures for each component of the experiment. Iteration-specific implementations are documented in Appendix 2.A.

2.3.1 Preexperiment games and survey

After the instructor announced the bonus activity in the class, we invited the students to complete an online activity, including an informed consent form, a demographic and behavioral traits survey, and a number of games.

This online activity began with a list of incentivized games that enabled us to measure each student's risk preference, trust, trustworthiness, ability to think strategically (p-beauty contest), and ability to solve constrained optimization problems (Knapsack game). These games are popular in lab-in-the-field experiments, as they each have been shown to predict behavior in the field (Karlan, 2005). For the fall 2021 iteration of the experiment, we added a public goods game in which participants report their individual contribution in both a conditional and an unconditional contributor role (Fischbacher and Gachter, 2010). The online activity concluded with an exit survey, where we collected demographic information (including major and gender) along with other self-reported behavioral trait measurements.

The full list of survey questions and links to the experiment interface is summarized in Appendix 2.A.2. The students had a week to complete the online survey and games.

2.3.2 The Wikipedia Editing Activity

After students completed the online games and surveys, we randomly assigned them to two experimental conditions, stratified by gender and their existing Econ 401 exam scores. With equal probability, a typical student in the experiment received two Wikipedia articles on either (1) economist biographies or (2) economic concepts that the instructor considered relevant to the course. Following the Wiki Education Foundation's best practice guidance for teaching with Wikipedia, we decomposed the Wikipedia editing assignment into two parts. In Part 1, we asked the students to register for a Wikipedia account by emailing them a short introduction email. All participants received the same introduction email. In Part 2, we introduced all participants to the Wikipedia editing task by sending each participant a personalized email that contained links to the pair of assigned Wikipedia articles. Each student was assigned a unique set of Wikipedia articles to edit.

In the personalized email, each assigned article was accompanied by a link pointing to a credible external source serving as a reference. This controlled for the difficulty in editing the assigned pages. In particular, we embedded individual specific Wikipedia links and reference links in the body of the email to achieve precise treatment assignment. To send these personalized emails to the participants, we used MailChimp, a marketing, automation & email platform. Templates for the email used for each treatment are provided in the Appendix. 2.A.3.

2.3.2.1 The Economist Biography Condition

For students randomized into the economist biography condition, each student received two Wikipedia articles on economist biographies, one for a male economist and the other for a female economist[2]. In the same email, we also provided a pair of links for the economist's professional website as a reference.

To control for the quality of the original Wikipedia page, we use Wikipedia's internal *content assessment class* as the ground truth for page quality. These ratings are generated by members of WikiProjects[3]. The ratings include the following six classes in increasing order: *Stub*, *Start*, *C*, *B*, *A*, *Good Article* (GA) and *Featured Article* (FA). The criteria range from "slightly more than a dictionary definition" for the *Stub* class to "a definitive source for encyclopedic information" for the *Featured Article* class. An article is considered "complete" if it reaches the A-class.

In our collection of Wikipedia articles about economists, the highest quality rating was C-class, with a total of 6 articles. The remaining biography articles were of Start-class or lower. All articles in our collection had room for improvement; therefore, we used all of them for our experiment.

[2]To collect for such set of Wikipedia articles that are gender-balanced, we searched the names of economists on Wikipedia by drawing names both from the List of Fellows of the Econometric Society and the Top 10% Female Economists List from REPEC.

[3]WikiProjects are formed by a group of contributors who collectively improve articles on a certain topic. For the same Wikipedia article, multiple WikiProject teams can give different content assessment ratings based on the completeness of the article pertaining to their specific topic of interest.

2.3.2.2 The economics concepts condition

For students assigned to the econ concepts condition, we sent each student a pair of Wikipedia articles on economic concepts covered in the class. In the same email, we also provided a link to the same concept covered in the online version of the New Palgrave Dictionary as a reference.[4]

Limited both by the scope of the course and the scarcity of economics concepts articles that are actually available on Wikipedia, we were able to obtain a collection of 147 Wikipedia articles for the econ concepts condition. For each iteration of the experiment, we generated unique pairs of links for all participants in the econ concepts condition. We reused the links in our link assignment no more than three times across all participants in one iteration of the experiment.

2.3.3 Debriefing survey

After the due date for Part 2 of the Wikipedia editing activity, we sent a debriefing survey to all participants in the experiment. The debriefing survey had three proposes. First, according to our IRB application, we had to explain the full set of treatments to all participants. Building on this necessary disclosure of experimental conditions, we solicited the page ratings from the participants directly; each participant rated two pages that were originally assigned to him/her and also rated two more pages from the other treatment condition. Last, we elicited two individual confidence measures from all participants who completed the debriefing survey. The complete set of questions used in the debriefing survey is available in Appendix 2.A.5.

2.3.4 Incentives

In this experiment, students received bonus points as an incentive to participate; they could earn up to 1% of the total points available from all their class activities. We awarded a maximum of 12 bonus points in total.

First, students could earn up to 4 points from the preexperiment games. and survey, where:

$$PreExp\_Bonus\_Point_i = 1 + 3 \times \frac{PreExp\_Point_i - \min(PreExp\_Points)}{\max(PreExp\_Points) - \min(PreExp\_Points)}$$

This continuous transformation was intended to provide an incentive scheme that ensured that all games in the preexperiment segment were properly incentivized.

[4] Students had access to the online dictionary through a proxy server maintained by the university library.

For the Wikipedia editing task, 2 bonus points were awarded for registering for a Wikipedia account and submitting proof of having done so (Part 1). Students could earn a maximum of 6 points for conducting edits on the assigned page (Part 2). Additionally, 1 point was granted if the participant submitted proof of the edit, and another 5 points were issued based on the quality of the edits, which was rated by three independent raters on an scale of 1–5. Last, by participating in the debriefing survey, participants could earn another 4 bonus points.

2.4 Results

In the Winter 2019 semester, 462 registered students were on the roster at the beginning of the semester. A total of 405 participants eventually consented to participate. We invited 403 of these individuals to participate in our experiment, as two participants consented until the end of the semester but were dropped from the study for logistical reasons. The overall participation rate was 87.23%, and the effective sample size was $n = 401$[5]. In the Fall 2020 semester, 432 students were registered in the course, and a total of 333 students consented to participate in the experiment. The overall participation rate was 77.08%. We treated all $n = 333$ students by assigning them either a pair of Wikipedia articles about economics concepts or a pair of Wikipedia articles that consisted of biographies of economists. In the Fall 2021 semester, 392 students were registered at the start of the experiment, and we invited all the students to participate in the experiment. A total of 345 students completed the consent form. However, based on the subsequent requests of the students, we ended up treating all 392 students by inviting them to edit Wikipedia articles[6]. Figure 2.1 summarizes the number of participants that engaged in each stage of the experiment.

2.4.1 Overview of Results

In Figure 2.2, we document the fraction of students who joined and edited Wikipedia articles by gender and treatment condition over the three iterations. In the Winter 2019 semester, when women were assigned to edit Wikipedia articles, those assigned to edit pages on economic concepts were significantly less likely to edit their assigned articles. To be exact, we found a significant difference in the likelihood of editing; i.e., female participants were less likely to edit their assigned Wikipedia article when they were assigned articles on economic concepts

[5]We dropped two transgender participants from the sample analysis. They were assigned articles to edit, similar to all the other students. In our survey, we presented a multiple choice question about gender with five options: Female, Male, Transgender, Other, Prefer not to say. Participants could choose multiple options at the same time. The two participants we dropped explicitly chose only the "Transgender" category.

[6]In the first two iterations of the experiment, we received multiple requests from students who did not participate in the Preexperiment Games and Survey in which they asked for their pairs of articles to edit.

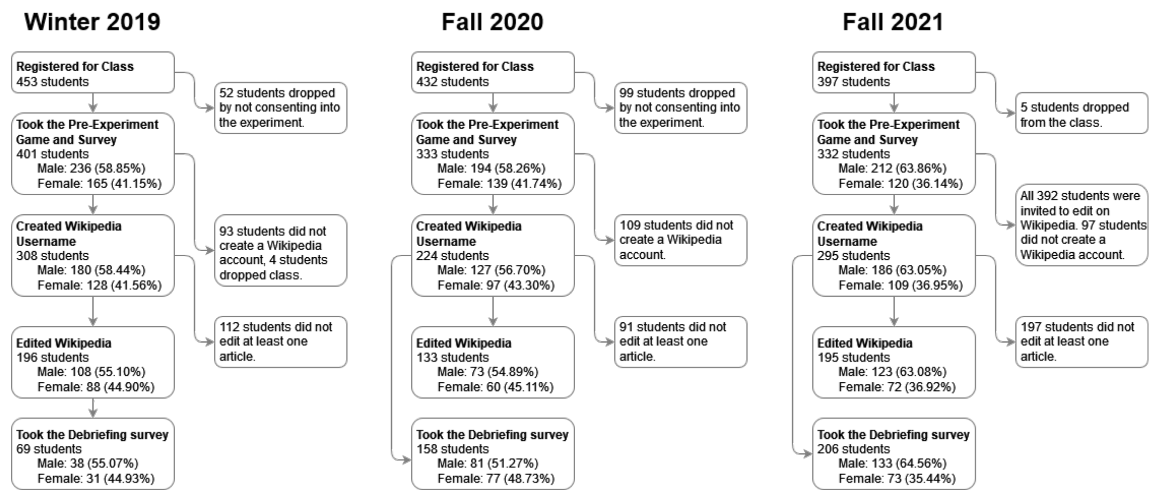


Figure 2.1: Participants' Responses in Each Stage of the Experiment

than when they were assigned biographical articles, with a difference in likelihood of 22.31%. There was no such difference in likelihood to edit for the male participants. However, this finding was not replicated in the 2020 and 2021 iterations, where the likelihood of joining Wikipedia and editing the assigned articles were mostly indistinguishable among the different treatment groups and sexes.

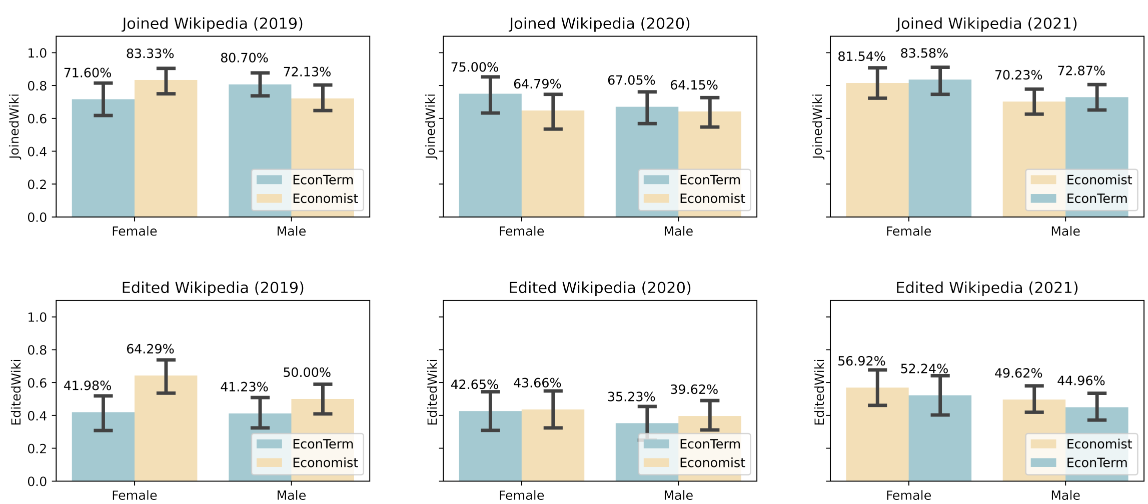


Figure 2.2: Fractions of Participants Joining/Editing Wikipedia

Note: 95% confidence intervals obtained using bootstrap method were plotted.

2.4.2 Registered Report

In this section, we test the set of hypotheses introduced in our preanalysis plan (Chen et al., 2020a). In this registered report, we use the pooled data collected over the three iterations of the experiment.

2.4.2.1 Hypotheses

Hypothesis 1 *Students assigned to edit Wikipedia articles on economist biographies are more likely to make edits than their counterparts assigned to edit Wikipedia articles on economic concepts.*

Hypothesis 2 *Women are less likely than their male counterparts to edit a Wikipedia article on an economics concept.*

Hypothesis 3 *Women and men are equally likely to edit economists' biographies on Wikipedia.*

Hypothesis 4a *Participants who are more confident about their grasp of economic concepts are more likely to edit articles on economic concepts.*

Hypothesis 4b *Participants who are more confident about their standing in the class are more likely to edit articles on economic concepts.*

Hypothesis 5 *When controlling for the average confidence, competence, and expertise level needed to edit the assigned Wikipedia article, the gender difference in editing assigned articles on economic concepts disappears.*

2.4.2.2 Measurements from Preexperiment Games

In this experiment, all participants who were recruited to edit Wikipedia articles completed a battery of games[7] before they were assigned Wikipedia articles to edit. From these games, we obtained a number of measurements covering preferences and cognitive abilities. In the following analysis, we include the following set of variables as control variables: Trust_Invest, KB_correct_count, MPL_risk_aversion and p_Beauty_InitialGuess.

In the trust game, Trust_Invest was elicited as the initial unconditional investment amount when the participant was the first mover in the game. This variable measured the baseline trust the participant had toward a randomly paired opponent. In the knapsack game, KB_correct_count captured the number of rounds in which the participant correctly solved the

[7] The full set of games is listed in the Appendix 2.A.2.

knapsack question. A higher count indicated that the participant performed better in solving such a complex problem Murawski and Bossaerts (2016). In the multiple price list game, MPL_risk_aversion denoted the switching point from a safe to a risky lottery. Participants who switched later down the list had a large MPL_risk_aversion value, denoting that they were more risk averse than other participants (Holt and Laury, 2002). In the p-beauty contest game, we kept only the initial guess, p_Beauty_InitialGuess, as it has been shown to be the most representative measure for level-k thinking (Schnusenberg and Gallo, 2011). In the context of intermediate economics courses, this variable serves as a proxy for the students' performance in the course.

2.4.2.3 Measurements of the Assigned Wikipedia Articles

Each participant in the experiment was assigned two Wikipedia articles and could choose to edit either one. To control for the quality and length of the assigned articles, we included page_length and stub_quality as covariates in our analysis.

Notably, the page_length variable is defined as the average page length in kilobytes of the pair of assigned articles. To explain the construction for the stub_quality variable, we considered two aspects; from the participant's perspective, the page with lower quality sets a lower bar for editing. However, since high-quality pages are rare, there would not be enough variation within each quality category if we introduced a set of dummy variables for each page quality grade.

2.4.2.4 Pooled Analysis

In this section, we pool the data from all three iterations and report the direct testing of the five hypotheses. To account for the unforeseeable common shocks introduced by the new coronavirus disease (COVID-19), we estimated all the regression models for the pooled results by controlling for the yearly dummy variables. As an overview, we did not find a gender gap when introducing students to editing Wikipedia articles for the first time. We did find evidence that participants are less likely to complete such an edit when they are assigned to edit articles on economic concepts instead of biographies.

In Table 2.1, based on the pooled data gathered over the three iterations, we found support for Hypothesis 1. Based on the reported predicted margins for the EconConcept dummy variable, participants assigned to edit economic concept articles were 8% less likely to edit their assigned Wikipedia articles than their counterparts.

To test Hypothesis 2, based on the reported interaction term in Table 2.1, we did not find a significant difference in the likelihood of editing when male and female participants were assigned to edit articles on economic concepts. This outcome is reaffirmed in Table 2.2 Column (1-4), which shows that the gender dummy does not have a significant predictive margin for the participants who were assigned to edit articles on economic concepts.

We rejected Hypothesis 3 based on the subsample analysis of those who were assigned to edit biography articles, as shown in Table 2.2 Column (5) and (6); female participants were 8% more likely than male participants to edit their assigned biography page. However, upon controlling for risk preferences and other behavioral attributes elicited from the preexperiment games, the female dummy did not predict any significant difference in likelihood to edit. That is, holding all other behavioral attributes and preferences fixed, female participants were as likely as male participants to edit their assigned Wikipedia articles.

Table 2.1: Edited on Wikipedia

	(1) H1	(2) H1	(3) H1	(4) H1	(5) H2H3	(6) H2H3	(7) H2H3	(8) H2H3
EconConcept	-0.075 (0.030)**	-0.074 (0.033)**	-0.081 (0.031)***	-0.069 (0.035)**	-0.076 (0.029)***	-0.076 (0.033)**	-0.081 (0.031)***	-0.070 (0.035)**
Female					0.070 (0.030)**	0.070 (0.030)**	0.030 (0.033)	0.029 (0.033)
Female×EconConcept					-0.042 (0.061)	-0.039 (0.061)	-0.009 (0.064)	-0.014 (0.064)
page_length		0.000 (0.001)		0.000 (0.002)		0.001 (0.001)		0.000 (0.002)
Trust_Invest			-0.015 (0.010)	-0.017 (0.010)			-0.014 (0.010)	-0.016 (0.010)
KB_correct_count			0.005 (0.019)	0.007 (0.019)			0.007 (0.019)	0.009 (0.019)
LD_Curiosity			0.000 (0.002)	0.000 (0.002)			0.000 (0.002)	0.000 (0.002)
MPL_risk_aversion			0.013 (0.008)*	0.013 (0.008)			0.013 (0.008)	0.012 (0.008)
p_Beauty_InitialGuess			0.001 (0.001)	0.001 (0.001)			0.001 (0.001)	0.001 (0.001)
page qualities (dummy)	no	yes	no	yes	no	yes	no	yes
ExpYear (dummy)	yes	yes	yes	yes	yes	yes	yes	yes
Observations	1126	1125	995	993	1126	1125	995	993

Notes: The table reports different specifications for Probit regression. Subjects without complete game records are dropped in Column (3), (4), (7) and (8). Average marginal effects are reported; the marginal effects for interaction terms are the difference between the average marginal effect of the dummy EconConcept for female and male; the standard errors are calculated using the Delta method (Ai and Norton 2003), and hypotheses are tested using the Wald test.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.2: Edited on Wikipedia — Split sample by treatment condition

	(1) EconConcept	(2) EconConcept	(3) EconConcept	(4) EconConcept	(5) Biography	(6) Biography	(7) Biography	(8) Biography
Female	0.049	0.048	0.031	0.022	0.088	0.090	0.028	0.032
	(0.043)	(0.043)	(0.048)	(0.048)	(0.042)**	(0.042)**	(0.045)	(0.044)
page_length		-0.001		-0.001		0.008		0.008
		(0.002)		(0.002)		(0.004)**		(0.004)**
Trust_Invest			0.001	-0.001			-0.028	-0.028
			(0.015)	(0.015)			(0.014)**	(0.014)**
KB_correct_count			0.008	0.011			0.006	0.002
			(0.028)	(0.028)			(0.026)	(0.026)
LD_Curiosity			0.000	0.001			-0.001	-0.001
			(0.002)	(0.002)			(0.002)	(0.002)
MPL_risk_aversion			0.003	0.002			0.022	0.023
			(0.011)	(0.011)			(0.011)**	(0.011)**
p_Beauty_InitialGuess			0.001	0.001			-0.000	-0.000
			(0.001)	(0.001)			(0.001)	(0.001)
page qualities (dummy)	no	yes	no	yes	no	yes	no	yes
ExpYear (dummy)	yes	yes	yes	yes	yes	yes	yes	yes
Observations	547	547	481	479	579	578	514	514

Notes: The table reports different specifications for Probit regression. Subjects without complete game records are dropped in Column (3), (4), (7) and (8). Average marginal effects are reported; the marginal effects for interaction terms are the difference between the average marginal effect of the dummy EconConcept for female and male; the standard errors are calculated using the Delta method (Ai and Norton 2003), and hypotheses are tested using the Wald test. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

To test Hypothesis 4, in the debriefing survey administered after the due date for the WikiEdit activities, we included an economic concepts quiz composed of eight questions. After the participants answered the quiz, we directly asked them how confident they felt about their performance on the quiz both in terms of absolute and relative performance[8]. In Table 2.3, we see that the absolute confidence measure has a significant predictive margin, but the relative confidence measure does not. Elicited as an incentivized guess of their correctly answered quiz questions, participants who reported having answered one additional question correct were 2.3% more likely than other participants to edit their assigned Wikipedia articles. This predictive margin is considerably small in magnitude compared to the baseline likelihood to edit, which is above 40%. Coupled with the observation null result for relative confidence measures shown in Column (5-8) Table 2.3, we can conclude that choosing to edit on Wikipedia has less to do with one's relative standing across peers in a population and more to do with one's absolute confidence in the grasp of the subject matter. Furthermore, in Table 2.4, the self-reported GPA for the course does not predict whether participants will edit their assigned articles. This outcome agrees with the observation we made based on Table 2.3, where the relative standing does not matter much in the individual decision to edit on Wikipedia.

[8]See Appendix 2.A.5.1 for the set of questions we used to elicit the confidence measures.

Table 2.3: Edited Wikipedia: Pooled data on confidences

	(1) H4a	(2) H4a	(3) H4a	(4) H4a	(5) H4a	(6) H4a	(7) H4a	(8) H4a
EconConcept	-0.089	-0.121	-0.068	-0.099	-0.096	-0.128	-0.076	-0.107
	(0.047)*	(0.052)**	(0.048)	(0.053)*	(0.047)**	(0.052)**	(0.048)	(0.053)**
Female	0.026	0.024	0.003	0.001	0.027	0.026	0.005	0.003
	(0.048)	(0.049)	(0.049)	(0.051)	(0.048)	(0.049)	(0.049)	(0.051)
EconConcept×Female	-0.085	-0.085	-0.093	-0.093	-0.093	-0.093	-0.101	-0.101
	(0.095)	(0.095)	(0.096)	(0.097)	(0.095)	(0.095)	(0.096)	(0.097)
Confidence_Absolute	0.027	0.026	0.027	0.027				
	(0.011)**	(0.011)**	(0.012)**	(0.012)**				
Confidence_Relative_TapHalf					0.113	0.104	0.113	0.106
					(0.079)	(0.079)	(0.081)	(0.082)
page_length		0.003		0.002		0.003		0.002
		(0.002)		(0.002)		(0.002)		(0.002)
Trust_Invest			-0.020	-0.022			-0.021	-0.022
			(0.016)	(0.016)			(0.016)	(0.016)
KB_correct_count			-0.041	-0.039			-0.036	-0.034
			(0.027)	(0.027)			(0.027)	(0.027)
LD_Curiosity			0.002	0.002			0.002	0.002
			(0.002)	(0.003)			(0.003)	(0.003)
MPL_risk_aversion			0.018	0.018			0.020	0.019
			(0.014)	(0.014)			(0.014)	(0.014)
p_Beauty_InitialGuess			-0.000	-0.000			-0.000	-0.000
			(0.001)	(0.001)			(0.001)	(0.001)
page qualities (dummy)	no	yes	no	yes	no	yes	no	yes
ExpYear (dummy)	yes	yes	yes	yes	yes	yes	yes	yes
Observations	379	376	369	366	379	376	369	366

Subjects without complete game records are dropped in Column (3), (4), (7) and (8). Average marginal effects are reported; the marginal effects for interaction terms are the difference between the average marginal effect of the dummy EconConcept for female and male; the standard errors are calculated using the Delta method (Ai and Norton 2003), and hypotheses are tested using the Wald test.

Table 2.4: Edited Wikipedia: Pooled Data on Self-reported GPA

	(1) H4b	(2) H4b	(3) H4b	(4) H4b
EconConcept=1	-0.091	-0.128	-0.069	-0.105
	(0.048)*	(0.052)**	(0.048)	(0.053)**
Female=1	0.036	0.036	0.018	0.018
	(0.048)	(0.050)	(0.050)	(0.051)
EconConcept×Female	-0.116	-0.109	-0.123	-0.118
	(0.096)	(0.097)	(0.097)	(0.098)
Predict_GPA	0.003	-0.002	0.011	0.005
	(0.046)	(0.046)	(0.048)	(0.048)
page_length		0.003		0.003
		(0.002)		(0.002)
Trust_Invest			-0.011	-0.012
			(0.016)	(0.016)
KB_correct_count			-0.037	-0.035
			(0.027)	(0.028)
LD_Curiosity			0.002	0.002
			(0.002)	(0.003)
MPL_risk_aversion			0.021	0.021
			(0.014)	(0.014)
p_Beauty_InitialGuess			-0.001	-0.001
			(0.001)	(0.001)
page qualities (dummy)	no	yes	no	yes
ExpYear (dummy)	yes	yes	yes	yes
Observations	368	365	358	355

Subjects without complete game records are dropped in Column (3), (4), (7) and (8). Average marginal effects are reported; the marginal effects for interaction terms are the difference between the average marginal effect of the dummy EconConcept for female and male; the standard errors are calculated using the Delta method (Ai and Norton 2003), and hypotheses are tested using the Wald test.

Predict_GPA is elicited by asking participants to predict their letter grade for the course. It is then mapped into numerical grade points in $[0, 4]$.

Last, in Table 2.5, we provide evidence of how average page ratings impact the participants' likelihood of editing their assigned Wikipedia articles. For a pair of articles that was rated to have higher avg_competence and avg_expertise ratings, the predicted margins are significantly different from zero. Based on the aggregated responses from the debriefing survey, avg_competence captures how likely participants were to consider themselves among the top half of Wikipedia editors in regard to the page they were prompted to edit. If all participants rated themselves among the top half of editors for the pair of links they were assigned, then individual participants were 15% more likely to perform the edit than in cases where they were assigned to a pair of articles that no one felt competent to edit. This is a sizable difference. Similarly, avg_expertise captures how much expertise was needed to edit the page. Intuitively, if the expertise requirement increased by one grade, then participants were almost 5% less likely to perform the edit.

Note that since we collected the page ratings through the debriefing survey, the analysis is based on Fall 2020 and Fall 2021 data. For the Winter 2019 iteration, since the debriefing survey was conducted a year later, the participation rate for the survey is very low. In Table 2.5, due to the restricted sample size, we no longer see a gender gap such that female participants were less likely to edit on Wikipedia. Since there was no gender gap to begin with, we did not have evidence to reject Hypothesis 5.

Table 2.5: Edited Wikipedia: Pooled Data on Average Page Attribute

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	H5	H5	H5	H5	est5	est6	est7	est8
EconConcept	-0.051	-0.044	-0.056	-0.038	-0.037	-0.031	-0.044	-0.028
	(0.039)	(0.043)	(0.041)	(0.045)	(0.041)	(0.044)	(0.043)	(0.047)
Female	0.063	0.067	0.019	0.019	0.063	0.065	0.015	0.013
	(0.040)	(0.041)	(0.043)	(0.043)	(0.040)	(0.040)	(0.042)	(0.042)
Female $\times$EconConcept	0.014	0.014	0.014	0.007	0.030	0.029	0.034	0.027
	(0.081)	(0.081)	(0.085)	(0.085)	(0.081)	(0.081)	(0.084)	(0.084)
avg_competence					0.140	0.137	0.185	0.183
					(0.069)**	(0.069)**	(0.073)**	(0.073)**
avg_confidence					0.046	0.047	0.064	0.065
					(0.032)	(0.032)	(0.034)*	(0.034)*
avg_expertise					-0.050	-0.050	-0.049	-0.048
					(0.026)*	(0.026)*	(0.026)*	(0.026)*
page_length		0.001		0.001		0.001		0.001
		(0.002)		(0.002)		(0.002)		(0.002)
Trust_Invest			-0.024	-0.027			-0.024	-0.027
			(0.014)*	(0.014)*			(0.014)*	(0.014)**
KB_correct_count			-0.008	-0.003			-0.011	-0.005
			(0.022)	(0.022)			(0.022)	(0.022)
LD_Curiosity			0.001	0.001			0.001	0.001
			(0.002)	(0.002)			(0.002)	(0.002)
MPL_risk_aversion			-0.004	-0.003			-0.003	-0.002
			(0.011)	(0.011)			(0.011)	(0.011)
p_Beauty_InitialGuess			-0.000	-0.000			-0.000	-0.000
			(0.001)	(0.001)			(0.001)	(0.001)
page qualities (dummy)	no	yes	no	yes	no	yes	no	yes
ExpYear (dummy)	yes	yes	yes	yes	yes	yes	yes	yes
Observations	641	640	584	581	641	640	584	581

Subjects without complete game records are dropped in Column (3), (4), (7) and (8). The average page attributes are collected by having each of the subjects rate the originally assigned articles in the debriefing survey. For more details, please refer to Appendix 2.A.5.2 and Appendix 2.A.5.3. Average marginal effects are reported; the marginal effects for interaction terms are the difference between the average marginal effect of the dummy EconConcept for female and male; the standard errors are calculated using the Delta method (Ai and Norton 2003), and hypotheses are tested using the Wald test.

2.5 Discussion

The strength of our study lies in self-replication, specifically, that we conducted the experiment for three consecutive years. In Year 1 (2019), we observed a small gender gap that held only when comparing how likely female students were to edit economic concept vs. biography articles. However, after controlling for the treatment assignments, the gender effect disappeared in regression analyses. This finding encouraged us to run a robustness check in Years 2 and 3, mainly to increase the sample size. After conducting the set of registered analyses on year-specific data for 2020 and 2021, we obtained null results from a regression analysis, similar to what we initially obtained in Year 1. Since the year-specific results did not provide new insights, we have omitted them for simplicity. Notably, Year 2 (2020) was special, as everyone was participating remotely due to COVID-19, and Year 3 (2021) was a mix of remote and in-person instruction, where similar concerns still applied.

We found that students were less likely to edit on Wikipedia when they were assigned to edit economic concept articles. Based on our random assignment of economic concept and biography articles, this relationship is causal. This outcome is not surprising, as we discovered in the debriefing survey that compared to editing a biographical article, editing an article related to economic concepts is perceived as a more demanding task that requires more confidence as an editor and a better grasp of the domain knowledge.

We can conclude that our findings are robust and support the broad recruitment of editors from the university student population, as proposed by the Wiki Education Foundation. Notably, in our experiment, a seemingly minimal incentive actually introduced equality among male and female students in regard to editing Wikipedia articles. By equally offering a small stake, we demonstrated that the intervention was quite powerful by itself.

One limitation is that we only had one round of interaction with the students. Based on the current design, we cannot explain the possible differential retention of editors by gender. It was rare for the participants to reuse their reported Wikipedia accounts to perform additional edits on Wikipedia outside of the experiment. This finding calls for future research to target a different group to study the retention aspect of the gender gap.

By inviting only Econ 401 students to participate in the experiment, this study is also limited in that the findings may not be easily generalized to the broader population. However, inviting college students to edit Wikipedia articles has become a common practice in recent years, largely thanks to the efforts led by the Wiki Education Foundation. With our discovery that there is no gender gap in the entry stage based on our examined student population, we can ease the concern that the gender gap may be exacerbated by recruiting new editors from university campuses.

One additional limitation is that we used bonus points as incentives in this experiment, which is not comparable to the incentive structures for most Wikipedia editors. For the student population, issuing bonus points is known to be a very effective mechanism to inspire students to do extra coursework. By nature, this approach works equally well for both male and female students. In reality, however, Wikipedia editors are motivated by a wide variety of incentives, which are not bound to have equal attractiveness for editors belonging to different gender groups. If we were to conduct the experiment again in the future, we could control for the incentives introduced through bonus points by completely removing such an incentive from the Wikipedia-editing task. This would help us conclude whether recruiting from the student population without offering incentives leads to an inclusive group of editors.

Bibliography

Ahmed, Syed Ishtiaque et al. (2016). "Peer-to-peer in the Workplace: A View from the Road". In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. CHI '16. San Jose, California, USA: ACM, pp. 5063–5075. DOI: 10.1145/2858036.2858393.

Ai, Chunrong and Edward C Norton (2003). "Interaction terms in logit and probit models". In: *Economics letters* 80.1, pp. 123–129.

Akerlof, George A. and Rachel E. Kranton (2000). "Economics and Identity". In: *The Quarterly Journal of Economics* 115.3, pp. 715–753. DOI: 10.1162/003355300554881.

Andreoni, James and B Douglas Bernheim (2009). "Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects". In: *Econometrica* 77.5, pp. 1607–1636.

Ariely, Dan et al. (2009). "Doing good or doing well? Image motivation and monetary incentives in behaving prosocially". In: *The American Economic Review* 99.1, pp. 544–555.

Autor, David and Susan Houseman (2005). *Do Temporary Help Jobs Improve Labor Market Outcomes for Low-skilled Workers? Evidence from Random Assignments*. National Bureau of Economic Research.

Babcock, Linda et al. (2017). "Gender Differences in Accepting and Receiving Requests for Tasks with Low Promotability". In: *American Economic Review* 107.3, pp. 714–47. DOI: 10.1257/aer.20141734.

Bahler, Kristen (2017). "Unemployment Is Really Low. So Why Can't These People Find Jobs?" In: ed. by Money. [Accessed: 2019-08-15].

Bear, Julia B. and Benjamin Collier (Mar. 2016). "Where Are the Women in Wikipedia? Understanding the Different Psychological Experiences of Men and Women in Wikipedia". In: *Sex Roles* 74.5, pp. 254–265. DOI: 10.1007/s11199-015-0573-y.

Bénabou, Roland and Jean Tirole (2006). "Incentives and Prosocial Behavior". In: *American Economic Review* 96.5, pp. 1652–1678. DOI: 10.1257/aer.96.5.1652.

Benz, Matthias and Stephan Meier (Feb. 2008). “Do people behave in experiments as in the field?—evidence from donations”. In: *Experimental Economics* 11.3, pp. 268–281. DOI: 10.1007/s10683-007-9192-y.

Bergstrom, Theodore et al. (1986). “On the private provision of public goods”. In: *Journal of Public Economics* 29.1, pp. 25–49.

Bernanke, Ben S. (2004). “Editorial Statement”. In: *The American Economic Review* 94.1, pp. 404–404.

Bertrand, Marianne and Esther Duflo (2016). *Field Experiments on Discrimination*. Working Paper 22014. National Bureau of Economic Research. DOI: 10.3386/w22014.

Bertrand, Marianne and Sendhil Mullainathan (2004). “Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination”. In: *American Economic Review* 94.4, pp. 991–1013.

Bielby, William T and James N Baron (1986). “Men and women at work: Sex segregation and statistical discrimination”. In: *American Journal of Sociology* 91.4, pp. 759–799.

Bishop, Bradley Wade et al. (2019). “Scientists’ Data Discovery and Reuse Behavior: (Meta)Data Fitness for Use and the FAIR Data Principles”. In: *Proceedings of the Association for Information Science and Technology* 56.1, pp. 21–31. DOI: 10.1002/pra2.4.

Blondel, Vincent D. et al. (Oct. 9, 2008). “Fast Unfolding of Communities in Large Networks”. In: *Journal of Statistical Mechanics: Theory and Experiment* 2008.10, P10008. DOI: 10.1088/1742-5468/2008/10/P10008. arXiv: 0803.0476.

Borgman, Christine L. et al. (Feb. 2018). “Digital Data Archives as Knowledge Infrastructures: Mediating Data Sharing and Reuse”. In.

Bowers, John et al. (1995). “Workflow from within and without: Technology and cooperative work on the print industry shopfloor”. In: *Proceedings of the Fourth European Conference on Computer-supported Cooperative Work ECSCW’95*. Springer, pp. 51–66.

Buser, Thomas et al. (May 2014). “Gender, Competitiveness, and Career Choices ∗”. In: *The Quarterly Journal of Economics* 129.3, pp. 1409–1447. DOI: 10.1093/qje/qju009.

Carlsson, Magnus (2011). “Does hiring discrimination cause gender segregation in the Swedish labor market?” In: *Feminist Economics* 17.3, pp. 71–102.

Chapman, Adriane et al. (Jan. 2020). “Dataset Search: A Survey”. In: *The VLDB Journal* 29.1, pp. 251–272. DOI: 10.1007/s00778-019-00564-x.

Chen, Daniel L et al. (2016a). “oTree—An open-source platform for laboratory, online, and field experiments”. In: *Journal of Behavioral and Experimental Finance* 9, pp. 88–97.

Chen, Daniel L. et al. (Mar. 2016b). “oTree—An Open-Source Platform for Laboratory, Online, and Field Experiments”. In: *Journal of Behavioral and Experimental Finance* 9, pp. 88–97. DOI: 10.1016/j.jbef.2015.12.001.

Chen, M. Keith et al. (2017). *The value of flexible work: Evidence from Uber drivers*. Tech. rep. National Bureau of Economic Research.

Chen, Roy and Yan Chen (2011). “The Potential of Social Identity for Equilibrium Selection”. In: *The American Economic Review* 101.6, pp. 2562–2589.

Chen, Yan (2008). “Incentive-Compatible Mechanisms for Pure Public Goods: A Survey of Experimental Research”. In: *The Handbook of Experimental Economics Results*. Ed. by Charles Plott and Vernon Smith. Vol. 1. Amsterdam: North-Holland, pp. 625 –643.

Chen, Yan et al. (Apr. 2020a). *Motivating Experts to Contribute to Digital Public Goods: A Personalized Field Experiment on Wikipedia*. SSRN Scholarly Paper ID 3588132. Rochester, NY: Social Science Research Network. DOI: 10.2139/ssrn.3588132.

Chen, Yan et al. (2020b). “Motivating Metadata Contributions for Data Re-Use and Reproducibility”. In: *AEA RCT Registry*. DOI: 10.1257/rct.6159-1.0.

— (2020c). *Motivating Metadata Contributions for Data Re-Use and Reproducibility*. DOI: 10.1257/rct.6159-1.0.

Christensen, Garret et al. (2019). *Transparent and Reproducible Social Science Research: How to Do Open Science*. University of California Press. DOI: doi:10.1525/9780520969230.

Coffman, Katherine Baldiga (Nov. 1, 2014). “Evidence on Self-Stereotyping and the Contribution of Ideas”. In: *The Quarterly Journal of Economics* 129.4, pp. 1625–1660. DOI: 10.1093/qje/qju023.

Coffman, Lucas C. and Muriel Niederle (Sept. 2015). “Pre-Analysis Plans Have Limited Upside, Especially Where Replications Are Feasible”. In: *Journal of Economic Perspectives* 29.3, pp. 81–98. DOI: 10.1257/jep.29.3.81.

Comenetz, Joshua (2016). “Frequently occurring surnames in the 2010 Census”. In: *United States Census Bureau*.

Cook, Cody et al. (2018). *The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers*. Tech. rep. National Bureau of Economic Research.

Cosley, Dan et al. (2007). “SuggestBot: using intelligent task routing to help people find work in wikipedia”. In: *Proceedings of the 12th international conference on Intelligent user interfaces*. Downloaded on February 23, 2003 at http://www.communitytechnology.org/nsf_ci_report/, pp. 32 –41.

Cox, D. R. (1958). *Planning of Experiments*. A Wiley Publication in Applied Statistics. John Wiley & Sons.

Daniels, Morgan et al. (2012). "Managing Fixity and Fluidity in Data Repositories". In: *Proceedings of the 2012 iConference on - iConference '12*. Toronto, Ontario, Canada: ACM Press, pp. 279–286. DOI: 10.1145/2132176.2132212.

Dillahunt, Tawanna R. (2014). "Fostering social capital in economically distressed communities". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, pp. 531–540.

Dillahunt, Tawanna R. and Amelia R. Malone (2015). "The Promise of the Sharing Economy among Disadvantaged Communities". In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. CHI '15. Seoul, Republic of Korea: ACM, pp. 2285–2294. DOI: 10.1145/2702123.2702189.

Dillahunt, Tawanna R. et al. (2016a). "Designing for Disadvantaged Job Seekers: Insights from Early Investigations". In: *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*. DIS '16. Brisbane, QLD, Australia: ACM, pp. 905–910. DOI: 10.1145/2901790.2901865.

Dillahunt, Tawanna R. et al. (2016b). "Do Massive Open Online Course Platforms Support Employability?" In: *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. CSCW '16. San Francisco, California, USA: ACM, pp. 233–244. DOI: 10.1145/2818048.2819924.

Dillahunt, Tawanna R. et al. (Dec. 2017). "The Sharing Economy in Computing: A Systematic Literature Review". In: *Proceedings of the ACM on Human-Computer Interaction* 1.CSCW, 38:1–38:26. DOI: 10.1145/3134673.

Dillahunt, Tawanna R. et al. (2018). "Designing Future Employment Applications for Underserved Job Seekers: A Speed Dating Study". In: *Proceedings of the 2018 Designing Interactive Systems Conference*. ACM, pp. 33–44.

Dokko, Jane et al. (2015). "Workers and the Online Gig Economy. A Hamilton Project Framing Paper". In: *The Hamilton Project: Advancing Opportunity, Prosperity, and Growth*.

Dombrowski, Lynn et al. (2017). "Low-Wage Precarious Workers' Sociotechnical Practices Working Towards Addressing Wage Theft". In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. CHI '17. Denver, Colorado, USA: ACM, pp. 4585–4598. DOI: 10.1145/3025453.3025633.

Eckel, Catherine C. and Philip J. Grossman (2005). "Managing Diversity by Creating Team Identity". In: *Journal of Economic Behavior & Organization* 58.3, pp. 371–392.

Farrell, Diana et al. (2018). "The Online Platform Economy in 2018: Drivers, Workers, Sellers, and Lessors". In: *JPMorgan Chase Institute*.

Fischbacher, Urs and Simon Gachter (Mar. 2010). “Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Goods Experiments”. In: *American Economic Review* 100.1, pp. 541–556. DOI: 10.1257/aer.100.1.541.

Ford, Heather and Judy Wajcman (Aug. 1, 2017). “‘Anyone Can Edit’, Not Everyone Does: Wikipedia and the Gender Gap”. In: *Social Studies of Science*.

Gaddis, S. Michael (2018). *Audit studies: Behind the scenes with theory, method, and nuance*. Vol. 14. Springer.

Gebel, Michael (2013). “Is a temporary job better than unemployment? A cross-country comparison based on British, German, and Swiss panel data”. In: *SOEP paper* 543.

Glott, Ruediger et al. (2010). *Wikipedia Survey–Overview of Results*. Tech. rep.

Granovetter, Mark S. (1995). *Getting a job: A study of contacts and careers*. Chicago: University of Chicago Press.

Greenberg, Jane et al. (Oct. 24, 2001). “Author-Generated Dublin Core Metadata for Web Resources: A Baseline Study in an Organization”. In: *International Conference on Dublin Core and Metadata Applications* 0.0 (0), pp. 38–45.

Gregory, Kathleen (July 13, 2020). “A Dataset Describing Data Discovery and Reuse Practices in Research”. In: *Scientific Data* 7.1 (1), p. 232. DOI: 10.1038/s41597-020-0569-5.

Groves, Theodore and John O. Ledyard (1987). “Incentive Compatibility since 1972”. In: *Information, Incentives and Economic Mechanisms: Essays in Honor of Leonid Hurwicz*. Ed. by Theodore Groves et al. Minneapolis: University of Minnesota Press, pp. 48–111.

Hargittai, Eszter and Aaron Shaw (Apr. 3, 2015). “Mind the Skills Gap: The Role of Internet Know-How and Gender in Differentiated Contributions to Wikipedia”. In: *Information, Communication & Society* 18.4, pp. 424–442. DOI: 10.1080/1369118X.2014.957711.

Heerwegh, Dirk and Geert Loosveldt (2006). “An Experimental Study on the Effects of Personalization, Survey Length Statements, Progress Indicators, and Survey Sponsor Logos in Web Surveys”. In: *Journal of Official Statistics* 22.2, p. 191.

Hemphill, Libby et al. (Mar. 2022). “How Do Properties of Data, Their Curation, and Their Funding Relate to Reuse?” In: *Journal of the Association for Information Science and Technology*, asi.24646. DOI: 10.1002/asi.24646.

Hendry, David G. et al. (2017). “Homeless Young People, Jobs, and a Future Vision: Community Members’ Perceptions of the Job Co-op”. In: *Proceedings of the 8th International Conference on Communities and Technologies*. C&T ’17. Troyes, France: ACM, pp. 22–31. DOI: 10.1145/3083671.3083680.

Hendry, David G. et al. (2017). “U-District Job Co-op: constructing a future vision for homeless young people and employment”. In: *Information Technology & People* 30, pp. 602–628.

Holt, Charles A. and Susan K. Laury (2002). “Risk Aversion and Incentive Effects”. English. In: *The American Economic Review* 92.5, pp. 1644–1655.

Hui, Julie et al. (Nov. 2018). “Making a Living My Way: Necessity-driven Entrepreneurship in Resource-Constrained Communities”. In: *Proceedings of the ACM on Human-Computer Interaction* 2.CSCW, 71:1–71:24. DOI: 10.1145/3274340.

Hui, Julie S. et al. (2018). “IntroAssist: A Tool to Support Writing Introductory Help Requests”. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, p. 22.

Ichino, Andrea et al. (2005). “Temporary work agencies in Italy: A springboard toward permanent employment?” In: *Giornale degli economisti e annali di economia*, pp. 1–27.

Johansson, Louise (2013). “A Study of the Motivation Behind Collaborative Knowledge Production and the Formation of Community in Web 2.0, Using the Case Study of wikiHow.Com”. MA thesis.

Jørgensen, Bent (1987). “Exponential Dispersion Models”. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 49.2, pp. 127–145. DOI: 10.1111/j.2517-6161.1987.tb01685.x.

Juanchich, Marie et al. (May 2015). “Cognitive Reflection Predicts Real-Life Decision Outcomes, but Not Over and Above Personality and Decision-Making Styles”. In: *Journal of Behavioral Decision Making* 29.1, pp. 52–59. DOI: 10.1002/bdm.1875.

Karlan, Dean (2005). “Using Experimental Economics to Measure Social Capital and Predict Financial Decision”. In: *American Economic Review* 95.5, pp. 1688–1699.

Krueger, Alan B. et al. (2014). “Are the long-term unemployed on the margins of the labor market?” In: *Brookings Papers on Economic Activity* 2014.1, pp. 229–299.

Labor Statistics, Bureau of (2019a). “Economic News Release: Employment Situation Summary”. In: [Accessed: 2019-04-03].

— (2019b). “Occupational Employment and Wages Summary”. In: [Accessed: 2019-04-03].

Lahno, Amrei M. et al. (Oct. 2015). “Conflicting risk attitudes”. In: *Journal of Economic Behavior & Organization* 118, pp. 136–149. DOI: 10.1016/j.jebo.2015.03.003.

Laury, Susan and Laura Taylor (2008). “Altruism spillovers: Are behaviors in context-free experiments predictive of altruism toward a naturally occurring public good”. In: *Journal of Economic Behavior Organization* 65.1, pp. 9–29.

Ledyard, John (1995). “Public goods: A survey of experimental research”. In: *The Handbook of Experimental Economics*. Ed. by John H. Kagel and Alvin E. Roth. Vol. 1. Princeton, New Jersey: Princeton University Press.

Leider, Stephen et al. (2009). “Directed Altruism and Enforced Reciprocity in Social Networks”. In: *The Quarterly Journal of Economics* 124.4, pp. 1815–1851.

Li, Linfeng and Yan Chen (2020). “Gender Inequality in Contributions to Wikipedia”. In: *AEA RCT Registry*. DOI: 10.1257/rct.6753.

Mavletova, Aigul and Mick P. Couper (2015). “A Meta-Analysis of Breakoff Rates in Mobile Web Surveys”. In: *Mobile research methods: Opportunities and challenges of mobile research methodologies*, pp. 81–98.

Mesgari, Mostafa et al. (2015). ““The Sum of All Human Knowledge”: A Systematic Review of Scholarly Research on the Content of Wikipedia”. In: *Journal of the Association for Information Science and Technology* 66.2, pp. 219–245. DOI: 10.1002/asi.23172.

Mittereder, Felicitas and Brady T West (June 2021). “A DYNAMIC SURVIVAL MODELING APPROACH TO THE PREDICTION OF WEB SURVEY BREAKOFF”. In: *Journal of Survey Statistics and Methodology*, smab015. DOI: 10.1093/jssam/smab015.

Murawski, Carsten and Peter Bossaerts (Oct. 2016). “How Humans Solve Complex Problems: The Case of the Knapsack Problem”. In: *Scientific Reports* 6.1. DOI: 10.1038/srep34851.

Newman, Mark (Sept. 2017). *Networks: An Introduction*. 2 edition. Oxford ; New York: Oxford University Press.

Newman, Mark EJ and Michelle Girvan (2004). “Finding and Evaluating Community Structure in Networks”. In: *Physical review E* 69.2, p. 026113.

Niederle, M. and L. Vesterlund (Aug. 2007). “Do Women Shy Away From Competition? Do Men Compete Too Much?” In: *The Quarterly Journal of Economics* 122.3, pp. 1067–1101. DOI: 10.1162/qjec.122.3.1067.

Peytchev, Andy (2009). “Survey Breakoff”. In: *Public Opinion Quarterly* 73.1, pp. 74–97.

— (2011). “Breakoff and Unit Nonresponse across Web Surveys”. In: *Journal of Official Statistics* 27.1, p. 33.

Pienta, Amy M. et al. (Nov. 2010). “The Enduring Value of Social Science Research: The Use and Reuse of Primary Research Data”. In.

Piwowar, Heather A. et al. (Mar. 21, 2007). “Sharing Detailed Research Data Is Associated with Increased Citation Rate”. In: *PLoS ONE* 2.3. Ed. by John Ioannidis, e308. DOI: 10.1371/journal.pone.0000308.

Pons, Pascal and Matthieu Latapy (2006). “Computing Communities in Large Networks Using Random Walks”. In: p. 28.

Rege, Mari and Kjetil Telle (2004). “The Impact of Social Approval and Framing on Cooperation in Public Good Situations”. In: *Journal of Public Economics* 88.7, pp. 1625–1644.

Reuben, Ernesto et al. (Nov. 2015). *Taste for Competition and the Gender Gap Among Young Business Professionals*. Tech. rep. DOI: 10.3386/w21695.

Saccardo, Silvia et al. (Apr. 2018). “On the Size of the Gender Difference in Competitiveness”. In: *Management Science* 64.4, pp. 1541–1554. DOI: 10.1287/mnsc.2016.2673.

Salehi, Niloufar and Michael S. Bernstein (2018). “Ink: Increasing Worker Agency to Reduce Friction in Hiring Crowd Workers”. In: *ACM Transactions on Computer-Human Interaction (TOCHI)* 25.2, p. 10.

Samuelson, Paul A. (1954). “The Pure Theory of Public Expenditure”. In: *Review of Economics and Statistics* 36.4, pp. 387–389.

Santos, Luiz et al. (Sept. 2016). “FAIR Data Points Supporting Big Data Interoperability”. In: p. 10.

Sariisik, Merve (2018). *Identity Discrimination in the Sharing Economy: A Field Experiment*. University of Michigan Working Paper.

Schnusenberg, Oliver and Andrés Gallo (2011). “On Cognitive Ability and Learning in a Beauty Contest”. In: *Journal for Economic Educators* 11.1, pp. 13–24.

Shaw, Aaron and Eszter Hargittai (Feb. 1, 2018). “The Pipeline of Online Participation Inequalities: The Case of Wikipedia Editing”. In: *Journal of Communication* 68.1, pp. 143–168. DOI: 10.1093/joc/jqx003.

Steinbrecher, Markus et al. (June 2015). “Why Do Respondents Break Off Web Surveys and Does It Matter? Results From Four Follow-up Surveys”. In: *International Journal of Public Opinion Research* 27.2, pp. 289–302. DOI: 10.1093/ijpor/edu025.

Suchman, Lucy A. (1987). *Plans and situated actions: The problem of human-machine communication*. Cambridge University Press.

Suzuki, Ryo et al. (2016). “Atelier: Repurposing expert crowdsourcing tasks as micro-internships”. In: *Proceedings of the 2016 CHI conference on human factors in computing systems*. ACM, pp. 2645–2656.

Thomer, Andrea K. et al. (Feb. 2022). “The Craft and Coordination of Data Curation: Complicating ”Workflow” Views of Data Science [PREPRINT]”. In: DOI: 10.7302/4017.

Traag, V. A. et al. (Mar. 26, 2019). “From Louvain to Leiden: Guaranteeing Well-Connected Communities”. In: *Scientific Reports* 9.1 (1), p. 5233. DOI: 10.1038/s41598-019-41695-z.

Van Belle, Eva et al. (2017). “Why is unemployment duration a sorting criterion in hiring?” In: *IZA Discussion Paper No. 10876. Available at SSRN:* `https://ssrn.com/abstract=2998986`.

Vesterlund, Lise (2015). “Using experimental methods to understand why and how we give to charity”. In: *The Handbook of Experimental Economics*. Ed. by John H. Kagel and Alvin E. Roth. Vol. 2. Princeton, New Jersey: Princeton University Press.

Vilhuber, Lars (May 2019). “Report by the AEA Data Editor”. In: *AEA Papers and Proceedings* 109, pp. 718–29. DOI: `10.1257/pandp.109.718`.

Voors, Maarten J et al. (Apr. 2012). “Violent Conflict and Behavior: A Field Experiment in Burundi”. In: *American Economic Review* 102.2, pp. 941–964. DOI: `10.1257/aer.102.2.941`.

Wakita, Ken and Toshiyuki Tsurumi (May 8, 2007). “Finding Community Structure in Mega-Scale Social Networks: [Extended Abstract]”. In: *Proceedings of the 16th International Conference on World Wide Web*. WWW ’07. Banff, Alberta, Canada: Association for Computing Machinery, pp. 1275–1276. DOI: `10.1145/1242572.1242805`.

Ward, Patrick S. and Vartika Singh (June 2015). “Using Field Experiments to Elicit Risk and Ambiguity Preferences: Behavioural Factors and the Adoption of New Agricultural Technologies in Rural India”. In: *The Journal of Development Studies* 51.6, pp. 707–724. DOI: `10.1080/00220388.2014.989996`.

Wheeler, Earnest and Tawanna R. Dillahunt (2018). “Navigating the Job Search as a Low-Resourced Job Seeker”. In: *Proceedings of the 36th Annual ACM Conference on Human Factors in Computing Systems*. CHI ’18. Montreal, QC, Canada: ACM. DOI: `10.1145/3173574.3173622`.

Wilkinson, Mark D. et al. (Mar. 2016). “The FAIR Guiding Principles for Scientific Data Management and Stewardship”. In: *Scientific Data* 3.1, p. 160018. DOI: `10.1038/sdata.2016.18`.

Yaraghi, Niam and Shamika Ravi (2017). “The current and future state of the sharing economy”. In: *Available at SSRN:* `https://ssrn.com/abstract=3041207`.

Zhang, Xiaoquan Michael and Feng Zhu (2011). “Group Size and Incentives to Contribute: A Natural Experiment at Chinese Wikipedia”. In: *American Economic Review* 101.4, pp. 1601–15.

Appendices

2.A Experiment Procedures

In total, we conducted three iterations of the experiment across three semesters. Over all iterations, the treatment emails were administered through the same service provider. However, there were differences, which are described as follows. In 2019, we used MobLab[9]as the platform to administer the PreExp Games and Survey, and we conducted the debriefing survey in early 2020. For subsequent iterations, we controlled all steps of the experiment where we recreated the collection of games in oTree Chen, Schonger, and Wickens (2016a).

2.A.1 Timeline of the Experiment

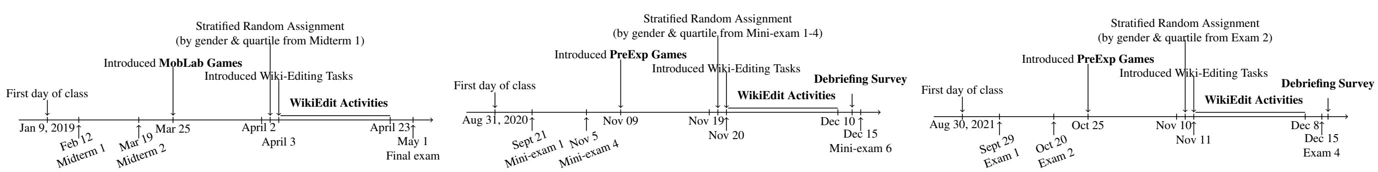


Figure 2.3: Annotated Timeline for Three Iterations of the Experiment

2.A.2 Pretreatment Measurements: Consent Form, Games and Survey

We deployed an oTree application (using Chen, Schonger, and Wickens, 2016a) to collect the pretreatment measurements[10]. The code is available on Github[11], and we are also hosting

[9]MobLab Inc. provides online services that allow instructors to teach concepts in economics and business through real-time interactive behavioral games. In this experiment, we chose games that could be played in an asynchronous manner.

[10]For the first iteration of the experiment conducted in April 2019, we used an identical set of games developed and hosted by MobLab Inc.

[11]https://github.com/llinfeng/WikiEdit_PreExpApp_Public

a live-demo page on Heroku. Please visit *click here* to see a demo session, where thirty participation links will become available upon toggling the "Show/hide" button[12].

Students received a personalized email with a unique link directing them to their individual experiment session. Through this interface, students were first prompted to sign an informed consent form. Then, they were invited to play the following list of games:

1. Risk preference measurement using a multiple price list (Holt and Laury, 2002).
2. A trust game where each participant plays the roles of both investor and recipient using the strategy method.
3. A competitiveness game with three stages. The layout of the stages follows Niederle and Vesterlund (2007), and the elicitation of competitiveness measurement follows Saccardo, Pietrasz, and Gneezy (2018).
4. A P-beauty contest, which repeats 5 rounds with $p = 2/3$. Each round includes the empirical distribution of Michigan students collected from the pilot session conducted in the Winter 2019 semester;
5. A three-round knapsack game with varying backpack capacity and item values (Murawski and Bossaerts, 2016).
6. (Fall 2021 only) A public goods game in which participants decide their individual contribution in as either a conditional or unconditional contributor. Each participant is paired with three randomly chosen subjects from data collected in Fischbacher and Gachter (2010).

As the last step, students were guided to the following survey. We presented the survey questions on five separate pages in our oTree application. The survey questions in each part fell into the described sections below.

[12] Note that the demo interface may take two minutes to respond. Once fully loaded, The demo interface should behave normally. To clarify, participants did not see any of the "Debug info" sections that appear toward the end of each demo page. The "Debug info" sections are unavoidable when we host the oTree application for unlimited public access to the admin portal.

2.A.2.1 PreExp Games

In this section, we compile the set of screenshots into two columns. To follow the flow of the experiment, please read from top to bottom in the left column. Then, proceed with the column on the right.

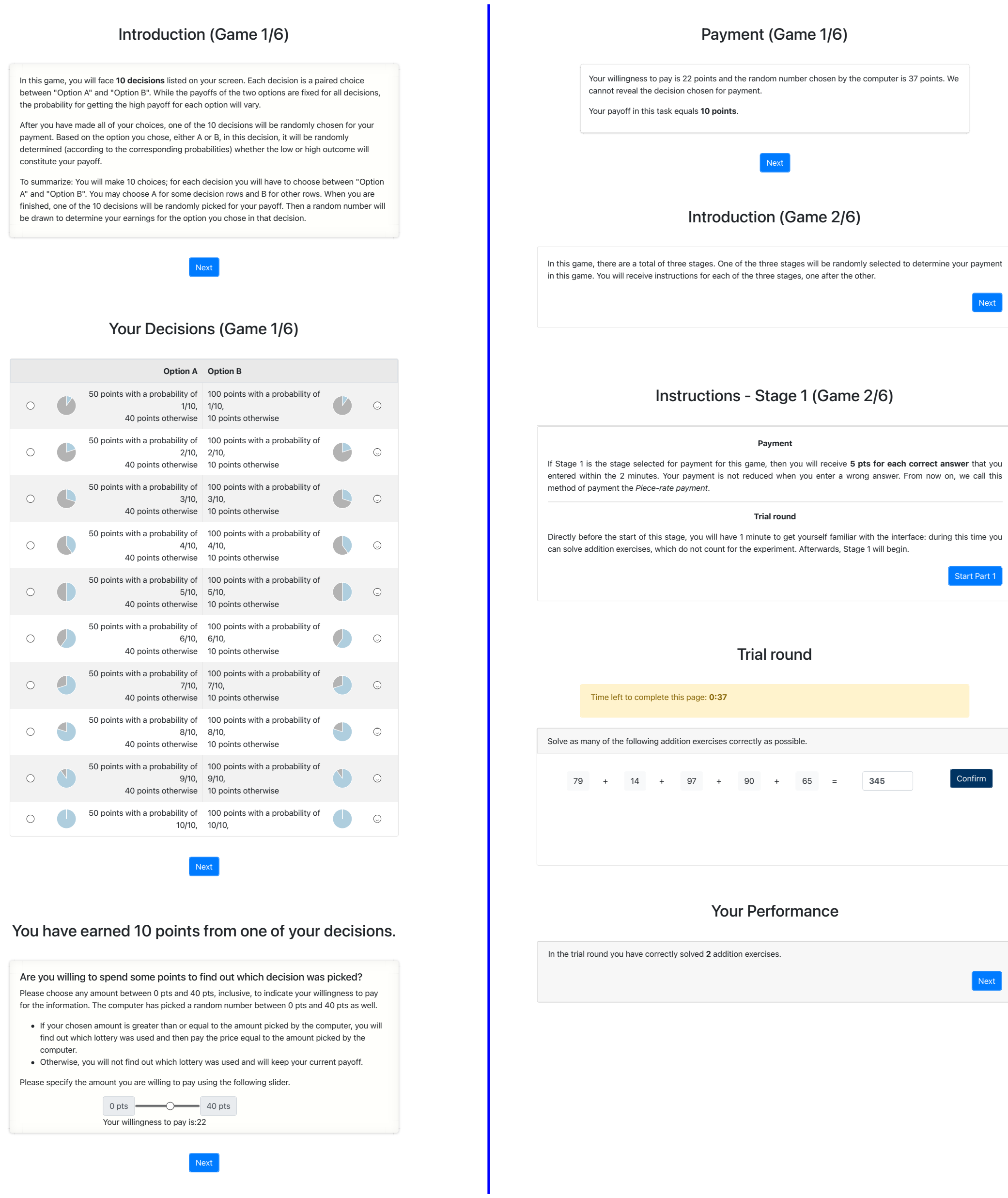


Figure 2.4: PreExp Games Screenshots — Part 1

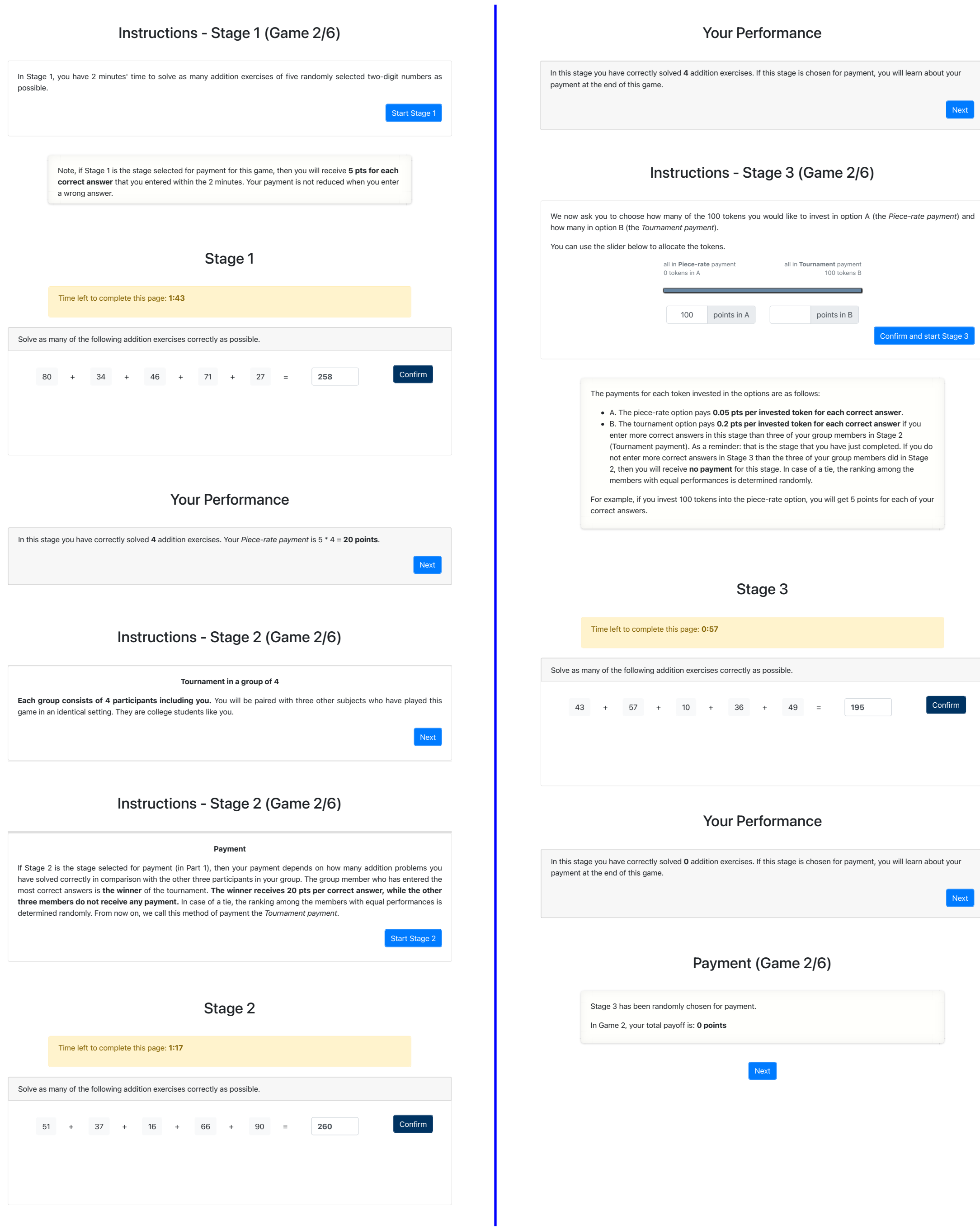


Figure 2.5: PreExp Games Screenshots — Part 2

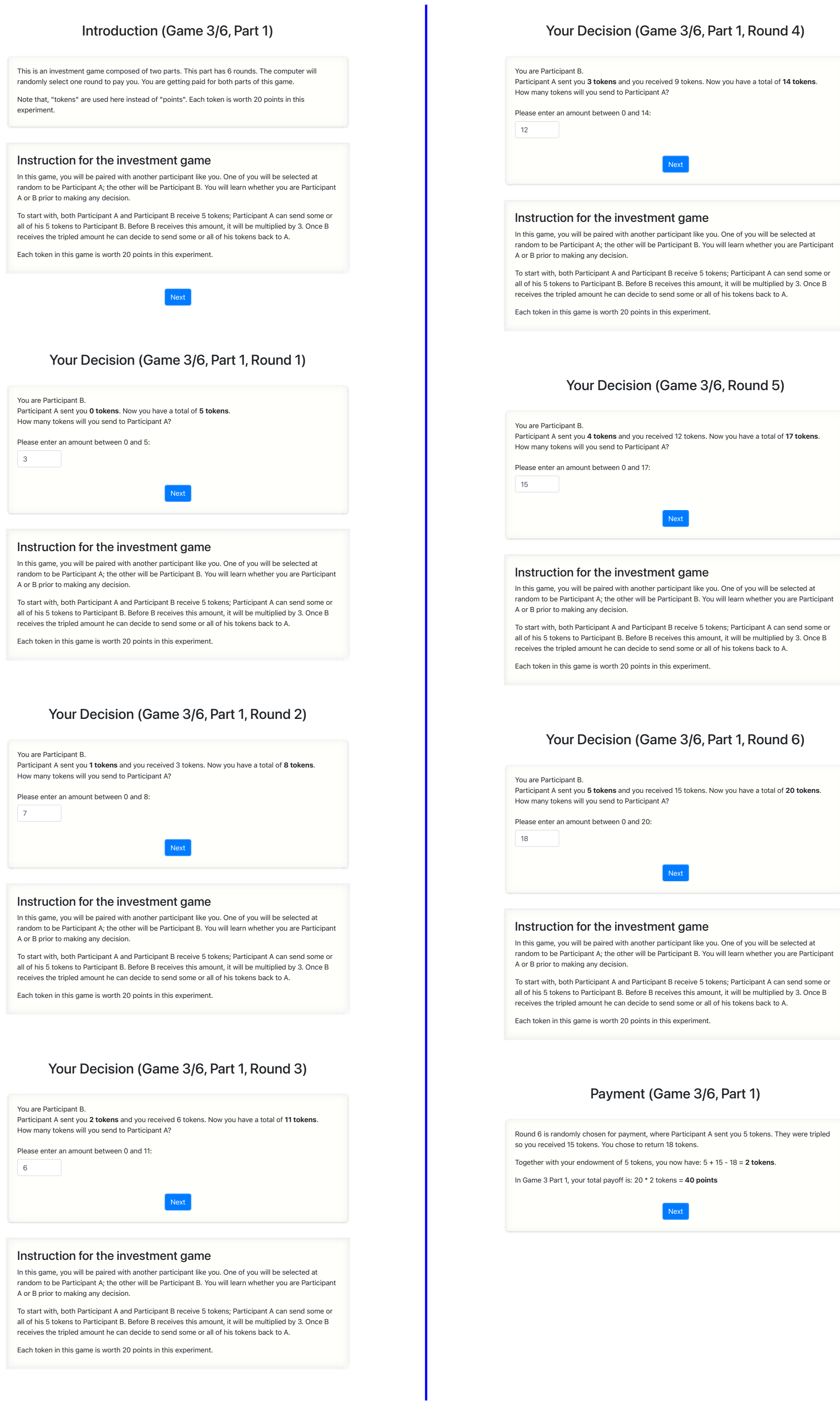


Figure 2.6: PreExp Games Screenshots — Part 3

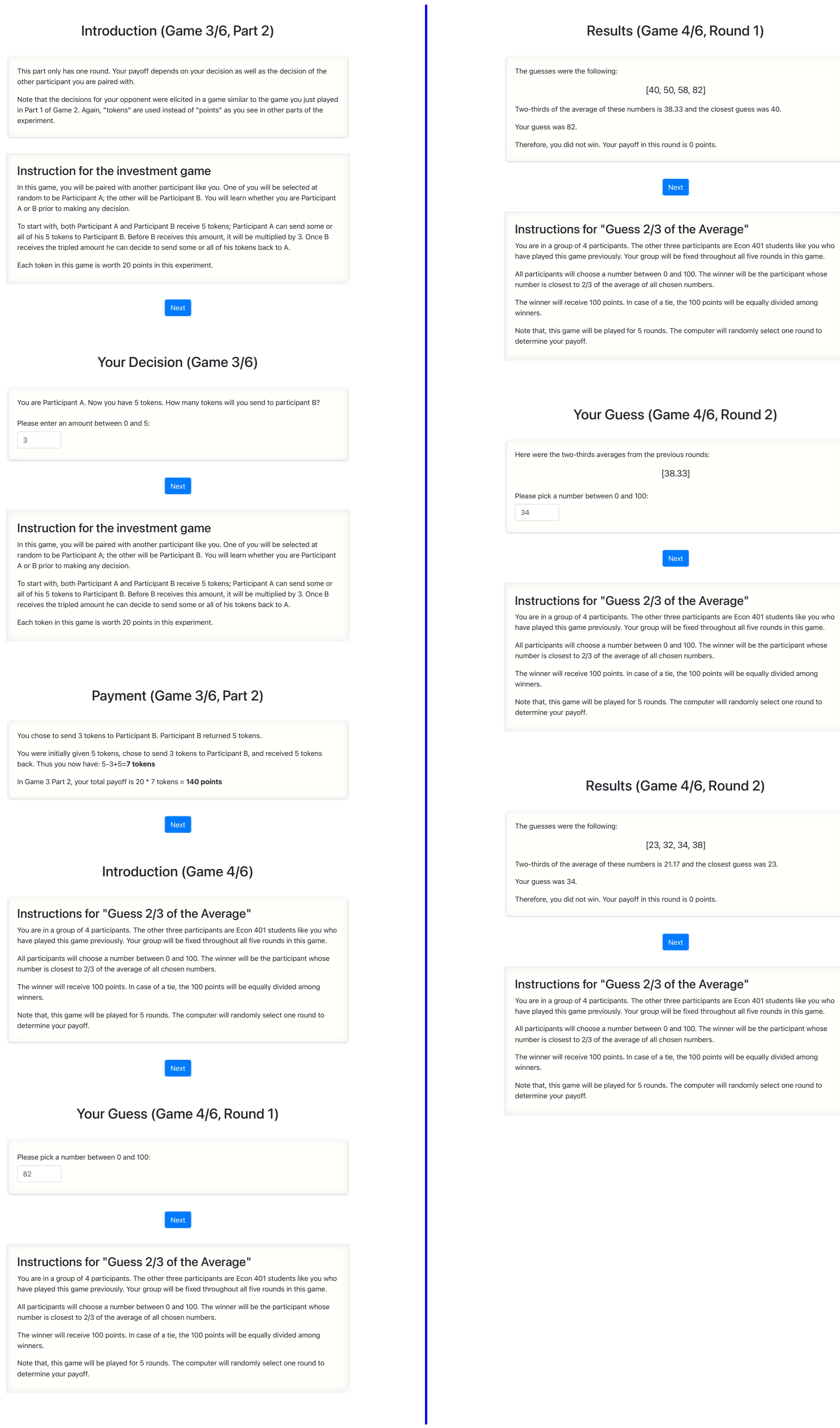


Figure 2.7: PreExp Games Screenshots — Part 4

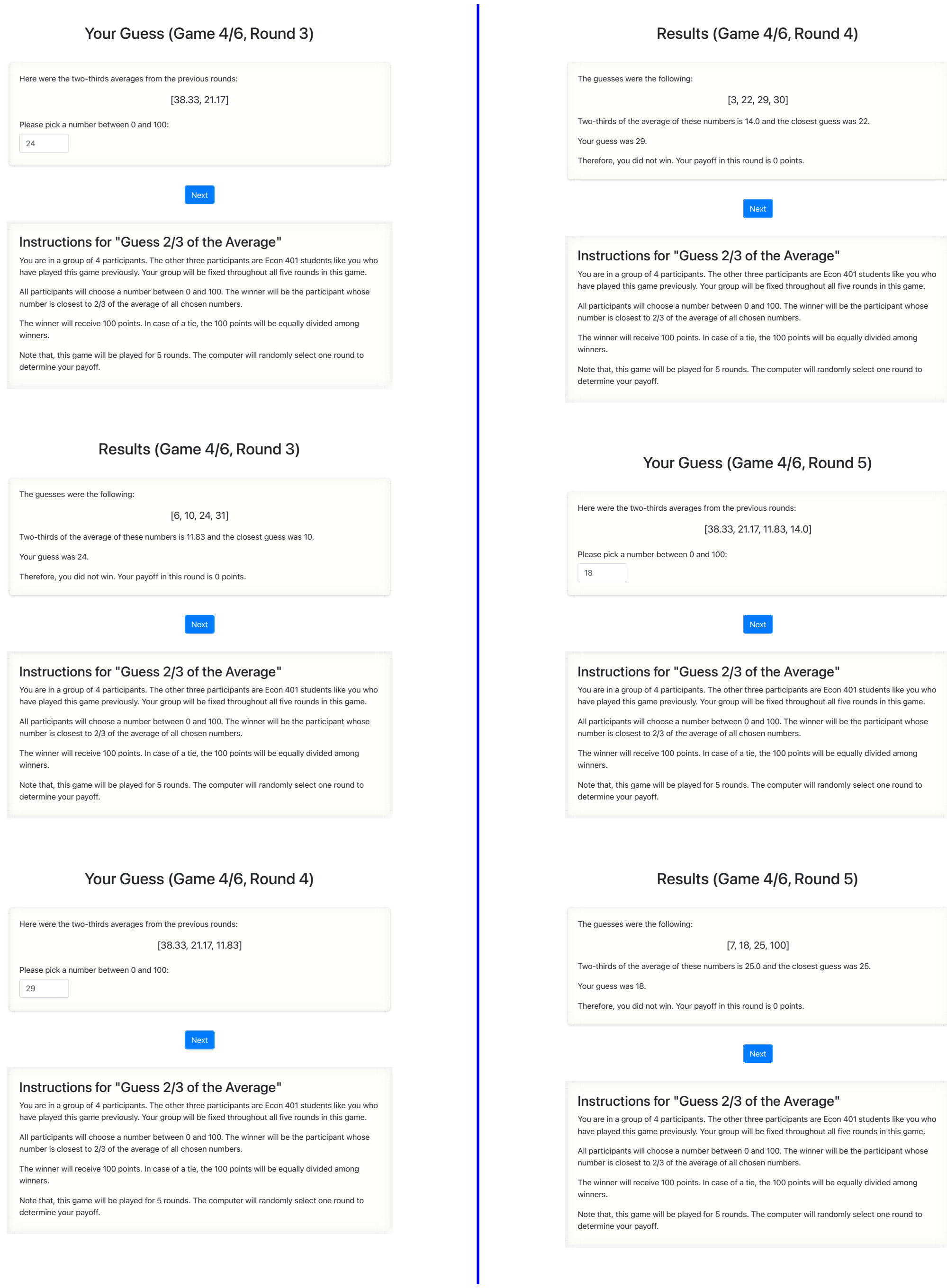


Figure 2.8: PreExp Games Screenshots — Part 5

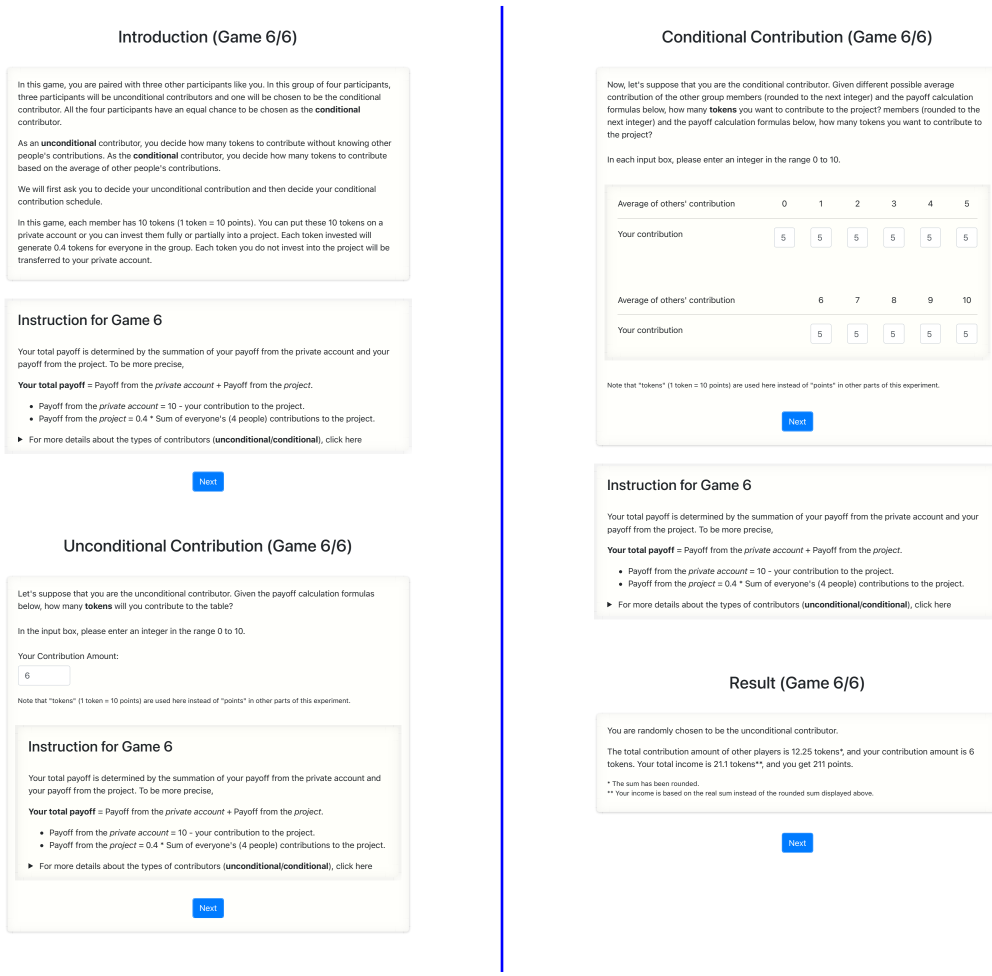


Figure 2.9: PreExp Games Screenshots — Part 7

2.A.2.2 Survey: Part 1

1. How competitive do you consider yourself?

   Please choose a value on the scale below, where the value 0 means ”not competitive at all” and the value 10 means ”very competitive”.

2. Generally speaking, would you say that most people can be trusted or that you can’t be too careful in dealing with people?

   - Most people can be trusted
   - Can’t be too careful
   - Other, depends
   - Don’t know

2.A.2.3 Survey: Part 2

3. Do you generally procrastinate (e.g., do you delay doing an unpleasant task that needs to get done) or would you rather get an unpleasant task done right away?

   Please select one from the following, where the value 0 means: “procrastinate whenever possible” and the value 10 means: “never procrastinate”.

4. Do you have a clear vision for what you want your future to look like or do you feel you don’t yet have a plan?

   Please select one from the following, where the value 0 means: “no particular plan yet” and the value 10 means: “know exactly what you want to do”.

5. Would you agree that you often think about your future goals and what you need to do to achieve them?

   Please select one from the following, where the value 0 means: “never think about the future and never plan ahead” and the value 10 “always think about the future and plan ahead”.

2.A.2.4 Survey: Part 3

6. Are you generally a person who is willing to take risks or do you try to avoid taking risks?

Please select one from the following, where the value 0 means “not at all willing to take risks” and the value 10 means “very willing to take risks”

7. Individuals may have different risk tolerance depending on the situation. How would you rate your risk tolerance in the following 6 situations?

   Please select one from the following, where the value 0 means: “very risk averse” and the value of 10 “very willing to take risks”.

   7.1 While driving?

   7.2 For investments?

   7.3 During freetime and sports?

   7.4 In your professional career?

   7.5 With your health?

   7.6 At trust in strangers?

2.A.2.5 Survey: Part 4

In this part, all options are listed as a “multiple-choice, multiple-answer” field where more than one option can be checked.

8. Which of the following do you identify with?

   - Female
   - Male
   - Transgender
   - Prefer not to say

9. What is your ethnicity?

   - American Indian or Alaskan Native
   - Asian or Pacific Islander
   - Black
   - Hispanic
   - White
   - Other

- Decline to answer

10. What is your major?

    - Economics
    - Mathematics
    - Statistics
    - Philosophy
    - Business

2.A.2.6 Survey: Part 5

11. What is your approximate cumulative GPA?

12. What grade do you expect to receive for Econ 401?

2.A.3 Email Templates

Throughout the email templates below, *variable texts* are represented by *|VarName|*. To populate all batches of individualized emails, for each email template, we created a spreadsheet with all the variable texts in the column, with each row representing a different participant. MailChimp, the email service provider we used, was in charge of substituting the *variable texts* by the intended values for each out-going email.

2.A.3.1 Email Template for Assigning Economist Biography Pages

```
Dear *|FNAME|*,
  As announced in the lecture, we will introduce you to
Wikipedia Editing activities. Once you finish each part,
please submit your response to the corresponding Canvas
assignment.

          Part 1: Creating a Wikipedia account

  To begin, please watch this  3-minute video tutorial on
how to edit Wikipedia articles.
  Then proceed to register for a Wikipedia account by
following this link. When creating a Wikipedia account,
```

we suggest that you use a combination of your first and last name as your username, e.g., DavidMiller. If your ideal username is taken, please use a different variation, e.g., DavidMiller2019.

Once you have created a user account, please submit your username and a screenshot of your user page to the corresponding Canvas assignment. (If you already had a Wikipedia account, please still submit your username and account page.)

Part 2: Picking and Editing a Wikipedia article

You may choose from the following Wikipedia articles:

- **Link to the Wikipedia page for Professor *|NAME1|*:** *|WLINK1|*
- **Link to the Wikipedia page for Professor *|NAME2|*:** *|WLINK2|*

To assist your edits, we would like to provide you with the following sources for each of the assigned Wikipedia pages. Please DO NOT cut and paste directly from the source and please DO respect the Honor Codes at the U of M.

- **For Professor *|NAME1|*, *|PRONOUN1|* homepage is:** *|SOURCE1|*
- **For Professor *|NAME2|*, *|PRONOUN2|* homepage is:** *|SOURCE1|*

Once you have finished editing, please submit a screenshot of the ''View History'' page as proof of your edits to the corresponding Canvas assignment (**1 bonus point**). Three trained raters will independently evaluate your edits on a scale of 1-5. We will use the median rating for your quality score. The total bonus points you can earn from Part 2 will be the sum of your quality score (**up to 5 bonus points**) and the 1 point for uploading the screenshot for the ''View History'' page.

As a reminder, this activity is part of a research study on contributions to public information goods, such as Wikipedia, designed by Professors Yan Chen and David Miller at the University of Michigan.

If you have any questions concerning this activity, please email us at econ-401-bonus-activity@umich.edu.

All the best,

-David

2.A.3.2 Email Template for Assigning Pages on Economic Concepts

Dear *|FNAME|*,

As announced in the lecture, we will introduce you to Wikipedia Editing activities. Once you finish each part, please submit your response to the corresponding Canvas assignment.

Part 1: Creating a Wikipedia account

To begin, please watch this 3-minute video tutorial on how to edit Wikipedia articles.

Then proceed to register for a Wikipedia account by following this link. (If you already have a Wikipedia account, you can skip the previous step; just log into your existing account.) When creating a Wikipedia account, we suggest that you use a combination of your first and last name as your username, e.g., DavidMiller. If your ideal username is taken, please use a different variation e.g., DavidMiller2019. Once you have created a user account, please submit your username and a screenshot of your user page to the corresponding Canvas assignment. You can earn **up to 2 bonus points** from this assignment.

(If you already had a Wikipedia account, please still submit your username and account page.)

Part 2: Start Editing a Wikipedia article

You may choose to edit one article from the following pair of Wikipedia articles:

- **Link to the Wikipedia article of ``*|WTERM1|*'':** *|WLINK1|*
- **Link to the Wikipedia article of ``*|WTERM2|*'':** *|WLINK2|*

To assist your edits, we would like to provide you with the following sources for each of the assigned Wikipedia pages. These references are chosen from the New Palgrave Dictionary of Economics. Please DO NOT cut and paste directly from the source and please DO respect the Honor Codes at the U of M.

- **For the Wikipedia article of ``*|WTERM1|*'', please refer to** ``*|NPTERM1|*'';
- **For the Wikipedia article of ``*|WTERM2|*'', please refer to** ``*|NPTERM2|*''.

(You will need to log in your Umich account to access the reference article.)

Once you have finished editing, please submit a screenshot of the ``View History'' page as proof of your edits to the corresponding Canvas assignment (**1 bonus point**). Three trained raters will independently evaluate your edits on a scale of 1-5. We will use the median rating for your quality score. The total bonus points you can earn from Part 2 will be the sum of your quality score (**up to 5 bonus points**) and the 1 point for uploading the screenshot for the ``View History'' page.

```
   As a reminder, this activity is part of a research
study on contributions to public information goods, such
as Wikipedia, designed by Professors Yan Chen and David
Miller at the University of Michigan.
   If you have any questions concerning this activity,
please email us at econ-401-bonus-activity@umich.edu.
   All the best,
   -David
```

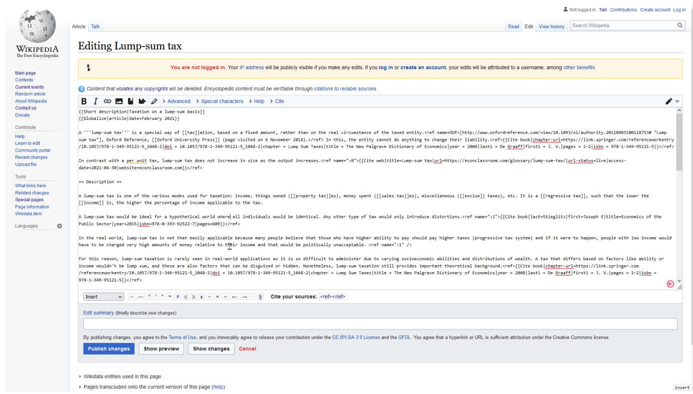
2.A.4 Editing on Wikipedia

In this section, we document the steps participants take to edit a page on Wikipedia. As shown in Figure 2.10 (a), the participant first needs to choose an article to edit based on the initial assignment sent through email. Once the participant makes a decision and clicks on the "Edit" tab at the top right corner of the Wikipedia page, the editing interface will load. An example of the editing interface is shown in Figure 2.10 (b). Last, after the participant submits his/her edits, we can examine what exactly was edited by visiting the "History" tab of the edited page and entering a "diff-view". In such a diff-view, two arbitrary historical versions of the Wikipedia page can be compared against each other. An example of the difference view is provided in Figure 2.10 (c), where new paragraphs that were added are marked with a leading plus(+) sign, and deleted texts are marked with a leading minus(−) sign. From what we observe, more than half of the participants who edited actually finished their edits in multiple sittings, where they may have edited multiple times and thus introduced multiple historical versions of the page. Still, with the diff-view shown in Figure 2.10 (c), we can expand the comparison of revisions to fully capture the set of edits in this case.

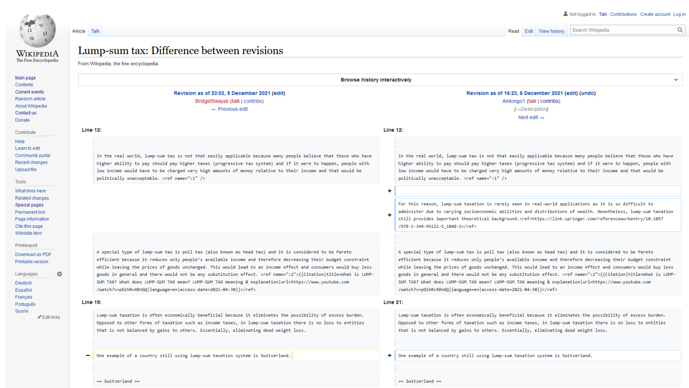
2.A.5 Postexperiment Survey

We sent the debriefing survey to all participants. Participants responded to the debriefing survey after they finished the main activities. By this point, the participants knew that their completion of the Wikipedia editing activities had been confirmed based on the bonus point they had been awarded. When surveying the Winter 2019 cohort, we provided monetary rewards capped at \$15 for each individual. For the Fall 2020 and Fall 2021 iterations, we transformed all questions proportionally into bonus points. The relative weight for each survey question was preserved during this transformation. Since the debriefing survey was administered within the same semester in both the Fall 2020 and Fall 2021 groups, we had a much higher participation

(a) Excerpt from the treatment email: choose one article to edit



(b) Editing interface



(c) Diff-View showing the details about the specific edit

Figure 2.10: Screenshots of How to Conduct and Verify an Edit

rate. The debriefing survey was composed of the following components.

2.A.5.1 Part 1: Confidence Survey

Page 1: Economics Concept Quiz

This part contains 8 multiple-choice questions about economics concepts covered in Econ 401. Let us see how much you remember.

You will be paid *25 cents* for each correct answer.

1. Eight multiple-choice questions on economic concepts

Page 2: Reflection for the Concept Quiz

2. 2 belief elicitation questions

After the participants finished the questions, on the next "screen/page", we asked the participants to predict the following two items:

2-(1) **Absolute confidence:** Out of the 8 questions from the quiz, how many questions do you think you answered correctly? Please select from the following drop-down menu. You will be paid \$1 if your answer is accurate.

2-(2) **Relative confidence:**

Please report a probability $p \in [0, 1]$ that you believe your score will be among the top half of scores among students who complete the quiz. You will have a chance to win \$2 additional bonus, and reporting your true belief about p will maximize your chance to win.

For more details, click here $\Rightarrow$ (the following text are presented to explain how the payoff is determined based on the reported p value).

Algebraically, you are choosing p to answer the following question:

In this question, choose a probability p that would make you indifferent between the following two options:

(a) Receive \$2 if your score is among the top half of scores,

(b) Receive \$2 with probability $p \in \{0, 0.01, 0.02, \cdots, 0.99, 1\}$.

We will draw a random number $y \in \{0, 0.01, \cdots, 0.99, 1\}$

- If $y > p$, then you get a \$2 bonus with probability y,
- If $y \leq p$, then you will get a \$2 bonus if your score is among the top half of scores.

The best you can do to maximize your expected payoff is to set p equal to your true estimate.

2.A.5.2 Part 2: Rate the Assigned Articles

The following questions were asked on a customized survey interface with a split-view design. In the right panel, we have embedded the Wikipedia article that were assigned. In the left panel, we have listed the survey questions, which are provided below. The full text of the survey interface starts after the following horizontal divider.

The Wikipedia article on the right is an old version of the page from the time it was assigned to you during the experiment (before you made any edits). Please answer the following questions.

3-(1) Wikipedia articles are written and updated by editors like you. Compared to the editors who wrote this Wikipedia article, please rate your relative competence in editing this Wikipedia article.

- Among the top half of the Wikipedia editors
- Among the bottom half of the Wikipedia editors

3-(2) How confident is an average Econ 401 student to edit and improve this Wikipedia page?

```
5 = completely confident to conduct edits
4 = highly confident in their edits
3 = moderately confident in their edits
2 = slightly confident in their edits
1 = not confident at all in their edits
0 = cannot conduct any edit due to lack of confidence;
```

Other Econ 401 students are also answering the following three questions about this same article. If your answer agrees with the majority of the raters (including you), you will get 50 cents.

3-(3) How much expertise would a person need to edit and improve this Wikipedia article?

```
5 = have completed a PhD in economics
4 = have completed a Master's degree in economics
3 = have taken an intermediate microeconomics class,
such as econ 401;
2 = have taken an introductory microeconomics class,
such as econ 101;
1 = have had no economics training;
```

3-(4) Think of all Econ 401 students ranked by their cumulative exam scores, what is the appropriate standing for the student to be qualified to edit and improve this article?

```
5 = No Econ 401 student is qualified
4 = Top 25%
3 = Above median
2 = Top 75%
1 = Any Econ 401 student
```

3-(5) How many questions in the Econ Concept quiz should an Econ 401 student answer correctly to edit and improve this article?

```
5 = No Econ 401 student is qualified
4 = Top 25%
3 = Above median
2 = Top 75%
1 = Any Econ 401 student
```

2.A.5.3 Part 3: Rate Assigned Articles for Others

In addition to having participants review the articles that were initially assigned to them, it is helpful to know what they think about the articles that were assigned to other participants. In this part, we have replaced the Wikipedia articles presented in the right panel and asked the questions provided below. The texts for the survey interface are located below the horizontal divider.

The Wikipedia article on the right is an old version of the page from the time it was assigned to one of your classmates during the experiment (before anyone made any edits). Please answer the following questions.

4-(1) Wikipedia articles are written and updated by editors like you. Compared to the editors who wrote this Wikipedia article, please rate your relative competence in editing this Wikipedia article.

- Among the top half of the Wikipedia editors
- Among the bottom half of the Wikipedia editors

4-(2) How confident is an average Econ 401 student to edit and improve this Wikipedia page?

```
5 = completely confident to conduct edits
4 = highly confident in their edits
3 = moderately confident in their edits
2 = slightly confident in their edits
1 = not confident at all in their edits
0 = cannot conduct any edit due to lack of confidence;
```

Other Econ 401 students are also answering the following three questions about this same article. If your answer agrees with the majority of the raters (including you), you will get 50 cents.

4-(3) How much expertise would a person need to edit and improve this Wikipedia article?

```
5 = have completed a PhD in economics
4 = have completed a Master's degree in economics
3 = have taken an intermediate microeconomics class,
such as Econ 401;
2 = have taken an introductory microeconomics class,
such as Econ 101;
1 = have had no economics training;
```

4-(4) Think of all Econ 401 students ranked by their cumulative exam scores, what is the appropriate standing for the student to be qualified to edit and improve this article?

```
5 = No Econ 401 student is qualified
4 = Top 25%
3 = Above median
2 = Top 75%
1 = Any Econ 401 student
```

4-(5) How many questions in the Econ Concept quiz should an Econ 401 student answer correctly to edit and improve this article?

```
5 = No Econ 401 student is qualified
4 = Top 25%
3 = Above median
2 = Top 75%
1 = Any Econ 401 student
```

2.A.5.4 Part 4: Rate the Edits

In this part, we asked the participants to evaluate the edits made by participants in previous iterations[13]. For each participant, we present one high-quality edit and one low-quality edit. We asked the participants to evaluate each edit by answering the following questions.

5-(1) How long would it take you to complete the edits that you have seen? Note, these edits have been conducted by Econ 401 students like you. Please fill in the number of minutes: ___.

5-(2) How many new references were added in the edits? Please fill in the number of references: ___.

5-(3) Do you think this edit would stay? Edits that are not up to the standard are likely to be reverted by other Wikipedia editors. Other editors can improve upon these edits as well, sometimes overwriting the existing edits.

(a) I believe these edits are poorly written and will not have a chance to stay on Wikipedia.

(b) I am *certain* that the edits will stay.

This will be answered using a slider.

5-(4) How confident was the Econ 401 student when he/she performed the edit?

```
1 = the editor does not appear to be confident in the
edit at all;
...
5 = the editor must have felt natural to conduct edits
on Wikipedia
```

[13]In the Winter 2019 iteration, we used edits from the same iteration.

This will be answered using a slider.

5-(5) How would you rate the *overall quality* of the edits? Please give a rating between 1-5 based on the quality of the edit as well as the *effort level*. Here, 1 denotes "minimum effort and poor quality" and 5 denotes "considerable effort and high quality". *This set of edits has been rated by three trained raters. If your answer agrees with the majority of the raters, you will get 50 cents.*

2.A.5.5 Part 5: Predict the Likelihood That an Edit Will Stay Published

We presented "a typical Wikipedia page" about an economic concept or biography and asked the participants to *predict* the likelihood that an edit would stay published on such a Wikipedia page.

6. How would you predict your letter grade? Please choose from the drop down list. (A+, A, A-, B+, B, B-, C, D)

   If you guessed correctly, you will get up to $2. ($2 if correct, $1 if off by one letter grade, and 0 otherwise.)

7. If a student received an A (top 40%)/B(next 40%)/C(next 16%)/DE (among the bottom 4%) in Econ 401, what's the likelihood that their edits would stay until now? Please report a probability for each scenario (slider between 0 - 100%).

   - For Econ Concept articles:
     - The edits of A-students, ...
     - The edits of B-students, ...
     - The edits of C-students, ...
     - The edits of lowest scoring students (D or lower), ...
   - For Economists' biography articles
     - The edits of A-students, ...
     - The edits of B-students, ...
     - The edits of C-students, ...
     - The edits of lowest scoring students (D or lower), ...

2.A.6 Change in Page Quality During the Experiment

First, since the experiment was conducted with three iterations over three years, one natural concern is that the assigned Wikipedia pages were also improved over time. Thus, the participants who participated in the experiment in the first iteration may have found the assigned pages easier to edit than did those who participated in the third iteration. One related concern is that in cases where page quality improvement can be attributed to the edits from the participants in the experiment, other participants who were assigned the same article may have found it more challenging to edit. In this section, we document a comprehensive data collection exercise that helped address the concern regarding the possible endogeneity issue with page quality.

First, for all the Wikipedia articles that we assigned in the experiment, we kept track of all the changes in page quality by compiling a full history of the revisions made on the “Talk page for each article. Mechanically, to make changes to a Wikipedia article, one has to change the content on the “Talk page by making a change to the page quality rating. Using the records of all revisions of the “Talk pages, we parsed out the assigned page quality scores and time tags for the corresponding revisions.

Throughout all the changes in page quality, we subsequently checked whether the change in page quality could be attributed to any of the edits from our participants. Based on the full set of Wikipedia usernames we collected from the participants in the experiment, we compiled a full record of the revisions that enumerated all the edits made by our participants. Then, for each Wikipedia article with a change in quality, we checked to see if there were revisions made on the page by our participants that were close enough in time to the time tag for the change in page quality. Using our best discretion, we did not find evidence that any of the changes in page quality were attributable to edits made by the participants in this experiment. This resolves the concern that participants over multiple iterations of the experiment were assigned with an increasingly difficult set of Wikipedia articles for the editing task.

CHAPTER 3

Motivating Contributions to Public Goods of Uncertain Future Values

3.1 Introduction

Metadata is data about data. Traditionally, metadata is contributed by data curators and data archives, where extensive efforts have been made to guarantee the correctness of the metadata. Once populated, the metadata about the studies and datasets are indexed by search engines, making future reuse possible. High-quality metadata increase the likelihood that a dataset will be found and reused in the future. However, it is often unclear when and how the metadata will be utilized. Metadata can therefore be viewed as a public good of uncertain future value. Little is known about people's motivations to contribute to public goods of uncertain future value. In this study, we investigate motivations for metadata provision using a field experiment.

Domain experts' contribution to metadata is invaluable, as they have deep expertise rooted in their scholarly works. However, experts may have little interest in enhancing metadata for studies they have published, which makes the datasets difficult to find and create barriers for data reuse and reproducibility. In our field experiment, we investigate what motivates scholars to contribute to metadata for their scholarly works. Specifically, we investigate potential motivations for experts to contribute metadata about data and code. First, an expert may care about the findability of her study. By enhancing the metadata, she helps modern search engines *find* the study and related data more easily. Findability is a key element of best practices for managing data to be FAIR (findable, accessible, interoperable, and reusuable) (Wilkinson et al., 2016). Articulating the connection between metadata and findability might reduce the uncertainty of the value for this type of digital public good. Having datasets that are easily discoverable facilitates reuse. A scholar might care about her data being used for two reasons. First, she might be motivated to provide metadata for *private benefits*, such as increased citations. Second, she may care about *social benefits*, such as helping others in their research.

When an expert provides enhanced metadata for a data deposit, it may generate different benefits for various groups of people at different levels of proximity to the expert. Locally, metadata enhancements by one expert will benefit all coauthors of the study. Studies on directed altruism suggest that the provision of help increases as social distance decreases (Leider et al., 2009). At an intermediate level, close neighbors in the coauthor network who are experts in the same field might also benefit from enhanced metadata. As close academic neighbors who share common topical or methodological interests, experts in the same field are more likely to reuse the data from well-documented and reusable studies from experts other than those from studies with less accessible or usable data. (Gregory, 2020). Globally, improving study-level metadata increases the *findability* of the data (Chapman et al., 2020). Researchers in the same or neighboring disciplines are more likely to find and cite better documented studies than less well-documented studies.

There are costs to enhancing metadata. For example, for one published study, it may take up to 30 minutes to populate all the study-level metadata fields that are available on openICPSR[1]. These costs can be borne by one author, a journal, a data archive, a research funding organization, or some combination of these stakeholders.

Chapman et al. (2020) offers a survey of the state of the art *dataset search* architectures from both the commercial and research domains. Despite advances in machine learning and NLP (natural language processing), dataset search still relies heavily on structured metadata fields. Greenberg et al. (2001) provides evidence that resource experts can provide high-quality metadata with the help of simple webforms and that resource experts are willing to provide metadata to enhance the findability of their studies. Santos et al. (2016) argues that metadata stand at the core of operationalizing an architecture that serves both data owners and data users in a FAIR (findability, accessibility, interoperability, and reusability) manner. To the best of our knowledge, there is no established quantitative evidence on the impact of high-quality metadata on citation outcomes. Piwowar, Day, and Fridsma (2007) demonstrates a sizable correlation between the number of citations (which experts value) and whether datasets are openly accessible.

Our study aims to find mechanisms that increase metadata provision. We do so by reducing the uncertainty of the future value of metadata and by making two different kinds of benefits more salient. We find that simply reducing the uncertainty of the future value of metadata is sufficient to attract a high level of metadata contribution from experts.

[1]openICPSR is a self-publishing repository for social, behavioral, and health sciences research data, available at https://www.openicpsr.org/

3.2 Literature Review

Metadata is a pure public good, as it is both nonrivalrous and nonexcludable. Economists have long examined individuals' motivations for contributing to public goods. Neoclassical theories of public goods provision predict that rational individuals have an incentive to undercontribute because they do not internalize the positive externalities of their contributions on others (Samuelson, 1954; Bergstrom, Blume, and Varian, 1986). Numerous lab and field experiments have been conducted to test and contextualize these models. We refer readers to Ledyard (1995) for a survey of laboratory experiments using a voluntary contribution mechanism in a wide range of environments and to Vesterlund (2015) for a more recent survey of laboratory and field experiments on charitable giving.

Economists have developed several perspectives to understand what might mitigate this undercontribution to the production of public goods. The mechanism design perspective relies on incentive-compatible tax-subsidy schemes enforced by a central authority.[2] Therefore, they cannot be directly applied to contexts where contribution is voluntary. In these contexts, a social norms and identity perspective applies insights from theories of social identity to the study of economic problems (Akerlof and Kranton, 2000). This research shows that when people feel a stronger sense of common identity with a group, they exert more effort and make more contributions to reach an efficient outcome (Eckel and Grossman, 2005; Chen and Chen, 2011). Finally, image motivations (Bénabou and Tirole, 2006), such as the desire to be liked and respected by others and by one's self, might lead to pro-social behavior as well (Andreoni and Bernheim, 2009; Ariely, Bracha, and Meier, 2009; Rege and Telle, 2004).

A robust finding from prior public goods experiments is that increasing the private benefit from the public good increases contributions. Researchers have investigated the effect of social impact on contributions to the digital public goods. (Zhang and Zhu, 2011) examine the natural experiment created by the Chinese government's blocking of Wikipedia. This policy reduced the size of the readership and led to a 43% decrease in the level of contributions by overseas Wikipedia editors who were not blocked during that time. This paper indicates that a reduction in the social impact of the public good discourages contributions. Our experiment design is partly based on these findings.

Another area that we examine is the effect of personalization on voluntary contributions to public goods. For example, Cosley et al. (2007) deploy an intelligent task-routing agent, SuggestBot, which asks Wikipedia editors to improve articles similar to those they have worked on before. They find that personalized recommendations lead to nearly four times as

[2] See Groves and Ledyard (1987) for a survey of the theoretical literature and Chen, 2008 for a survey of the experimental literature.

many actual edits as random suggestions do.

The increasing availability of data, along with the increasing expectations of data sharing by funders, has increased the gap between the need for high-quality metadata, which makes these data useful, and the quantity and quality of metadata voluntarily provided by the research community. Mandated data sharing is an attempt to address this gap through a central authority and implicit taxes using the threat of withdrawing federal funding. This approach, however, does not ensure that the quality of the metadata provided will be sufficient to make the data useful to third parties. Consistent with the general undersupply of public goods, researchers generally do not provide high-quality metadata[3]. It is widely accepted that high-quality metadata is necessary for successful data reuse (see, for example, Borgman, Scharnhorst, and Golshan (2018), Bishop et al. (2019)). Empirical research has established that curatorial actions improve the quality of metadata and increase data discovery and reuse (Hemphill et al., 2022; Daniels et al., 2012; Pienta, Alter, and Lyle, 2010). However, there is little research on what motivates people to voluntarily improve the quality or quantity of the metadata they provide. Given the value of and growing need for such contributions, this paper explores those motivations.

3.3 Experiment Design

In this section, we present our field experiment design to explore factors that motivate experts to contribute metadata to their existing publications. In this section, we present our research site and sample selection strategies, treatment designs and experimental procedures.

3.3.1 Research Site and Sample Selection

Our study is set within the context of a journal data and code repository. The American Economics Association (AEA) is a leader in social science efforts to make empirical research published in its journals more transparent and reproducible. Beginning in 2004, the American Economic Review, one of the "top five" journals in economics, required that authors submit the data, code and readme files (a "replication package") of their articles as a ZIP file to the editorial office upon acceptance of their paper (Bernanke, 2004). This policy was subsequently extended to the American Economic Journal launched in 2009. Upon publication, the

[3]Thomer et al. (2022) quantifies the large number of hours primarily devoted to metadata, documentation, and quality checks (Table 2) that are required to transform data deposited at the Inter-university Consortium for Political and Social Research (ICPSR) into FAIR data. Specifically, curating the metadata for studies deposited at the ICPSR took a total of 2,669 hours from February 2017 to December 2019.

replication packages were published on the AEA journal website. Because of the technological constraints of using ZIP files, these datasets typically had sparse metadata.

Starting in July 2019, as part of a revision and enhancement of its data and code availability policy (Vilhuber, 2019), the AEA started requiring that new replication packages be deposited to a dedicated section of a trusted repository managed by the ICPSR, called "openICPSR." The openICPSR interface allows for richer metadata to be associated with deposits. In October 2019, the AEA migrated the entire historical archive of 3,073 data and code supplements collected under the previous policy into the AEA Data and Code Repository hosted at openICPSR.[4] This migration ensured that generalist and specialized search engines could more easily find the data included in the replication packages. However, due to the paucity of the metadata inherent in the original ZIP files, the metadata fields for the migrated deposits were very sparse (limited to JEL codes), which therefore reduced the benefits of migration. Participants in our experiment are the authors of articles published in AEA journals whose replication packages were migrated to openICPSR.

Many articles in economics are coauthored, making it challenging to implement a sensible treatment message due to potential spillovers onto coauthors. Of the total of 3,073 studies that were migrated, 2,452 (79.87%) were coauthored. To avoid treatment spillovers, we identified a set of articles in which no two articles have a common author (the sets of authors of the chosen studies are mutually exclusive). For more details about this coauthor network, please refer to the Appendix 3.D.2.

3.3.2 Experimental Conditions

To compare the efficacy of different motivations for metadata contributions, our study included three experimental conditions, a control condition (C), a private benefit condition (T_p), and a private and social benefit condition (T_{p+s}), implemented as personalized emails to our participants. Each email had the same subject line, namely, "Your AEA dataset migration." The emails also shared the same opening paragraphs that provided the background information that data deposits previously hosted by AEA had been migrated to openICPSR. Participants were invited to provide metadata for their data deposit that was associated with the selected AEA publication through a customized experiment interface through which we collected metadata.

```
Since July 16, 2019, the American Economic Association
has used the AEA Data and Code Repository at openICPSR
as the default archive for its supplements. The migration
```

[4] https://www.openicpsr.org/openicpsr/aea/. The ICPSR is the largest curated social science data repository in the world.

```
increases the findability of your data through a variety
of federated search interfaces such as Google Dataset
Search, the openICPSR search interface, and the general
ICPSR search interface.

To further enhance the findability of your data, we ask
that you spend up to 20 minutes to provide additional
metadata for your AEA data deposit through a user-friendly
web interface. The information will be batch-imported
back to the original openICPSR deposit.
```

The treatments were implemented by including additional paragraphs in the emails. Specifically, for the private benefit condition (T_p), we added an extra paragraph explaining that providing metadata increases the findability and the number of citations of the dataset and the article, thus emphasizing the private benefit of providing metadata. This paragraph was designed to reduce the uncertainty of the future value of metadata contributions.

```
Analyses of search and usage of ICPSR's data catalog indicate
that most datasets are discovered because searches pick
up metadata that includes citation to published articles
and key concepts (geography, methods). Enhancing the metadata
for your dataset will increase the likelihood that your
publication and data are found and cited.
```

In the private and social benefit condition (T_{p+s}), we added information on the social benefit of the enhanced metadata for graduate students and others. Specifically, the additional text reads as follows:

```
Analyses of search and usage of ICPSR's data catalog indicate
that most datasets are discovered because searches pick
up metadata that includes citation to published articles
and key concepts (geography, methods). Enhancing the metadata
for your dataset will increase the likelihood that your
publication and data are found and cited, making it more
useful to graduate students and others.
```

The closing paragraph was the same for all three conditions. The paragraph concluded with an emphasis on the private benefits of providing enhanced metadata. Participants were invited to a customized experiment interface to populate metadata. The experiment interface was identical across the control and treatments.

- `If you are interested to proceed, please click here to provide additional metadata for your study titled ``<Article 1:Title>.''`
- `Please click here if you think your coauthors are better suited for providing metadata. We will opt you out of future communications.`
- `Please click here if you are not interested in providing metadata and would like to opt out of future communications altogether.`

`For articles with more than one author, each coauthor is receiving an identical email with an individualized link. Thank you for your effort!`

`Sincerely,`

`Lars Vilhuber`

`AEA Data Editor`

At the end of the email, we provided three links for participants to choose from regarding their options. They could opt to do one of the following: (1) provide the metadata for their study; (2) let their coauthor(s) provide the metadata; or (3) opt out. Note that we provided two distinct ways for the authors to opt out. One acknowledges the role of their coauthor(s), and the other does not. Last, telling each author that their coauthor(s) received an identical message reduced the burden on authors to coordinate with each other prior to responding to the message.

We provided a simple interface to make it easier for the authors to provide study-level metadata. The interface was implemented in oTree (Chen, Schonger, and Wickens, 2016b), where each author was assigned a unique participation link to access the interface. Once loaded into the interface, the authors were immediately prompted to provide metadata after the completion of the informed consent form. The interface featured a two-panel design, with a screenshot of the current status of the data deposit on the right and a list of input fields on the left to collect metadata. Unlike the openICPSR website where authors have to sign in and publish a new version of their deposit to update their metadata, in our interface, the authors had immediate access to the metadata fields without the need to create an account, thereby reducing transaction costs. For more details about the experiment interface, please refer to the Appendix 3.A.

3.3.3 Randomization Procedure

In this experiment, the unit of randomization was at the data deposit level. All coauthors of the chosen data deposit were contacted under the same treatment arm. We chose a subset of the migrated data deposits such that none of the chosen deposits had overlapping authors[5]. We treated each participant by sending one email regarding exactly one data deposit.

Table 3.1 reports the summary statistics of the pretreatment characteristics, broken down into the three experimental conditions. Panel A presents the article attributes of the included articles, and Panel B provides demographic information for the authors. Columns (1) through (3) report the mean and standard errors. We performed χ^2 tests on joint orthogonality across the treatments and reports the associated p-values in column (4). As none of the reported p-values is less than or equal to 0.05, we argue that our block-random assignment produced balanced experimental groups along observable characteristics.

3.4 Hypotheses

The two primary outcome variables include the link(s) that participants clicked and the quantity of metadata contributions they provided. Our first hypothesis is based on the findability information in the treatment emails, which reduces the uncertainty and increases the expected future value of the public good.

Hypothesis 1 *Participants in the treatments are less likely to opt out of the study than those in the control condition.*

Between the two treatments, we expect that the marginal increase in participation (i.e., the decline in opting out) from the private benefit of findability and citation will be larger than that from the social benefit of helping others, based on a field experiment designed to motivate domain experts to contribute to Wikipedia (Chen et al., 2020a).

Hypothesis 2 *The marginal increase in the participation rate from the private benefit treatment is greater than that from the additional social benefit.*

Conditional on reaching the metadata contribution page (i.e., opting in), we expect that the effort level will follow the same order as that specified in our first two hypotheses.

Hypothesis 3 *At the data deposit level, the quantity of metadata provided will be greatest under the Social&Private Benefit treatment, followed by that under the Private Benefit treatment, which, in turn, is followed by that in the control condition.*

[5]For details of the exact procedure, please refer to Appendix 3.D.

Table 3.1: Attributes of Included Articles and Participants, by Conditions

	Control (1)	T_p (2)	T_{p+s} (3)	p-value (4)
Panel A: Article attributes				
Number of references	26.682	25.532	26.879	0.394
	(0.794)	(0.683)	(0.778)	
Journals:				
AEA Papers and Proceedings	0.047	0.066	0.045	0.290
	(0.010)	(0.011)	(0.009)	
AEJ: Applied Economics	0.109	0.119	0.146	0.192
	(0.014)	(0.015)	(0.016)	
AEJ: Economic Policy	0.144	0.117	0.160	0.144
	(0.016)	(0.015)	(0.017)	
AEJ: Macroeconomics	0.097	0.140	0.107	0.088
	(0.013)	(0.016)	(0.014)	
AEJ: Microeconomics	0.055	0.064	0.045	0.451
	(0.010)	(0.011)	(0.009)	
American Economic Review	0.507	0.441	0.451	0.084
	(0.023)	(0.023)	(0.023)	
AER: Insights	0.000	0.008	0.004	0.135
	(0.000)	(0.004)	(0.003)	
Journal of Economic Literature	0.006	0.000	0.004	0.246
	(0.004)	(0.000)	(0.003)	
Journal of Economic Perspectives	0.035	0.045	0.037	0.684
	(0.008)	(0.009)	(0.009)	
Observations	487	487	486	
Panel B: Participant attributes				
Female	0.240	0.226	0.211	0.300
	(0.013)	(0.013)	(0.013)	
Years since obtaining PhD	16.889	17.915	17.685	0.109
	(0.358)	(0.370)	(0.359)	
Observations	1007	1013	1003	

[1] Columns (1) through (3) report the average values, and column (4) reports the p-value testing the joint orthogonality across treatments. Standard errors are provided in parentheses.
[2] Article attributes used for block-random assignment are omitted.

Finally, based on Babcock et al. (2017), we expect that women are more likely to accept tasks with low promotability, such as providing local public goods. We expect that women in our sample will be less likely to click the link to indicate that their coauthor(s) will take over the task of metadata contribution.

Hypothesis 4 *For articles with more than one author, women are less likely to click the link to let their coauthors contribute metadata.*

3.5 Results

In this section, we first provide an overview of the participation and completion rate for the experiment. Then, we examine the two main outcomes in the experiment. At the individual level, we examine the intent to contribute at the extensive margin. At the study level, we examine the contribution rate, defined as the fraction of populated metadata fields by all coauthors in the study. Finally, we examine the delegation activities that took place in the context of coauthored the studies.

3.5.1 Overview of Metadata Contribution

According to our random assignment schedule, we treated a total of 3,023 participants that were drawn from 1,460 articles. Overall, we had a considerably high participation rate. In Table 3.2, for the baseline condition where we only mentioned findability in the email, the metadata contribution rate was 20.4% at the individual participant level and 40.2% at the study level. The study-level contribution-rate was higher since 70.89% of the papers in our sample were coauthored. According to the p-values reported in column (4), there is evidence that the contribution rate was different across the control and treatment conditions. Unexpectedly, the control condition attracted the most contributions. Figure 3.1 visualizes these results.

We provide summary statistics in Table 3.3. Starting at the first stage of the experiments, where participants interacted with hyperlinks embedded in the body of the treatment email, more participants in the control condition clicked on the link to participate. A joint orthogonality test conducted across three treatment conditions rejected the hypothesis that participants were equally likely to click the link to participate across all conditions at the 5% significance level. In line with what is plotted in Figure 3.1, participants in the treatment conditions were less likely to contribute to metadata than those in the control condition.

Table 3.2: Fractions of participants/studies that have contributed

	Full Sample	Control	T_p	T_{p+s}
Participant-level	18.0%	20.26%	15.99%	17.75%
(full sample)	(0.3842)	(0.4021)	(0.3667)	(0.3823)
Participant-level	23.87%	27.03%	21.18%	23.47%
(treated sample)	(0.4264)	(0.4444)	(0.4088)	(0.4241)
Study-level	35.75%	40.04%	32.24%	34.98%
(full sample)	(0.4794)	(0.4905)	(0.4679)	(0.4774)
Study-level	39.07%	43.62%	35.12%	38.46%
(treated sample)	(0.4881)	(0.4965)	(0.4779)	(0.4871)

Notes: Standard errors are provided in parentheses.

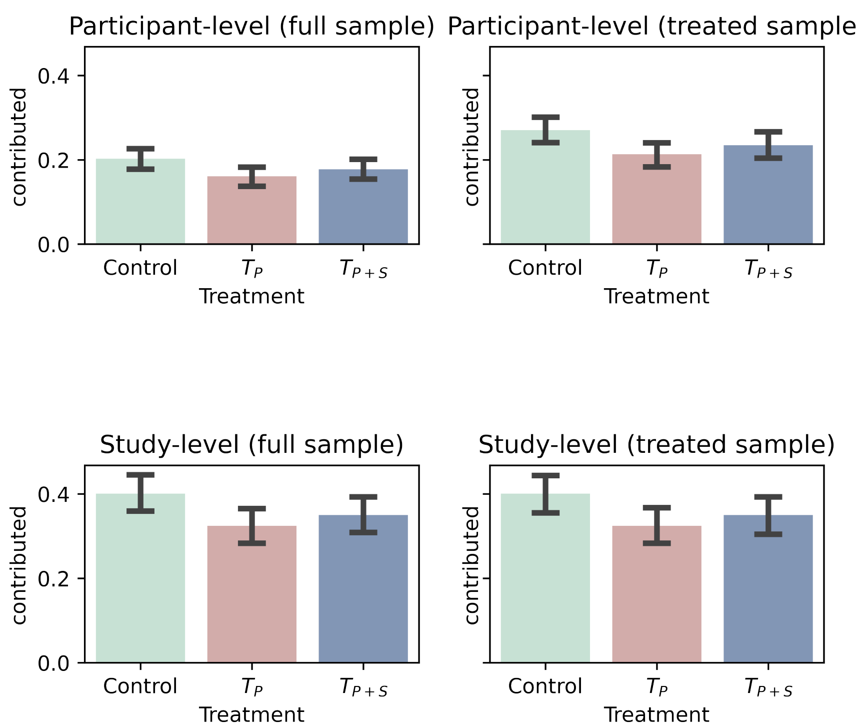


Figure Note: The error bars denote the 95% confidence intervals constructed using the bootstrap method.

Figure 3.1: Fractions of Contribution at the Participant and Study Level

Table 3.3: Summary Stats of Experimental Data

	Total	Control	T_p	T_{p+s}	p-value
Opened email	0.75	0.746	0.755	0.748	0.876
	(0.43)	(0.44)	(0.43)	(0.43)	
Intentions					
Link to participate	0.301	0.323	0.275	0.306	0.045
	(0.46)	(0.47)	(0.45)	(0.46)	
My coauthor will do	0.121	0.13	0.125	0.107	0.235
	(0.33)	(0.34)	(0.33)	(0.31)	
Not interested	0.149	0.146	0.147	0.154	0.876
	(0.36)	(0.35)	(0.35)	(0.36)	
No Response	0.483	0.459	0.505	0.484	0.111
	(0.5)	(0.5)	(0.5)	(0.5)	
Intended to participate (r_i)	0.083	0.104	0.051	0.094	0.175
	(0.72)	(0.73)	(0.71)	(0.71)	
Intended to contribute (c_i)	0.404	0.437	0.38	0.395	0.018
	(0.49)	(0.5)	(0.49)	(0.49)	
Contributed metadata	0.18	0.203	0.16	0.177	0.034
	(0.38)	(0.4)	(0.37)	(0.38)	
Demographics					
Female	0.239	0.252	0.235	0.229	0.300
	(0.35)	(0.36)	(0.35)	(0.35)	
Years since PhD	17.491	16.889	17.915	17.685	0.109
	(10.85)	(10.8)	(11.02)	(10.71)	
Coauthor Count	1.548	1.486	1.684	1.474	0.001
	(1.47)	(1.02)	(2.09)	(1.0)	
# Studies	1,460	487	487	486	
# Subjects	3,023	1,007	1,013	1,003	

Notes: The table reports means and standard deviations (in parentheses) for the variables used in the main analysis. The "Years since PhD" variable is constructed by counting the number of years since the PhD-granting year to year 2020. The "Coauthor count" variable does not include the author. That is, for two authors belonging to the same study, they each have a "coauthor count" of 1. The p-value column reports a joint orthogonality test on the treatment arms for each variable.

3.5.2 Intent to Contribute

In this section, we examine the willingness to contribute captured at the entry stage to the experiment, where participants chose either to click one out of three links in the treatment email or not to respond at all. The analysis was first carried out using the full ITT (intend-to-treat) sample. Second, with our capacity to track who actually opened the treatment emails, we focus the analysis on the treated participants.

3.5.2.1 Participation of the IIT Sample

To account for the willingness to contribute, we consider a binary outcome variable c_i that denotes whether the participant intended to contribute to the metadata. Among the three clickable entries in the email, the first two are classified as $c_i = 1$, where the participant either deliberately chose to provide metadata using our interface or thought that the coauthor(s) were bettered suited for the task of providing metadata. If the participant deliberately opted out of email communication by selecting the third option or the participant did not respond, we classify the participation outcome as $c_i = 0$. In Table 3.4, we employ a series of probit models to examine Hypothesis 1 and Hypothesis 2 using the full ITT sample.

Table 3.4: Intent to contribute

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
treatments	-0.0502***		-0.0509***		-0.0567***		-0.0576***	
	(0.0194)		(0.0194)		(0.0204)		(0.0202)	
find_private (T_p)		-0.0577**		-0.0581***		-0.0591**		-0.0580**
		(0.0225)		(0.0225)		(0.0237)		(0.0235)
find_private_social (T_{p+s})		-0.0428*		-0.0437*		-0.0543**		-0.0572**
		(0.0224)		(0.0224)		(0.0237)		(0.0235)
female			-0.0351	-0.0349	-0.0309	-0.0308	-0.0281	-0.0281
			(0.0216)	(0.0216)	(0.0233)	(0.0233)	(0.0234)	(0.0234)
NumCoauthors							-0.0213***	-0.0213***
							(0.0059)	(0.0060)
Female×NumCoauthors							-0.0101	-0.0101
							(0.0178)	(0.0178)
Years_since_PhD							0.0013	0.0013
							(0.0010)	(0.0010)
Article_Attr	no	no	no	no	no	no	yes	yes
χ^2-stats (T_p vs T_{p+s})		0.429		0.4		0.04		0.001
Observations	3023	3023	3023	3023	2684	2684	2684	2684

Notes: The table reports different specifications for Probit regression. Average marginal effects are reported; the marginal effects for interaction terms are the difference between the average marginal effect of NumCoauthors for female and male subjects; the standard errors are calculated using the Delta method (Ai and Norton, 2003), and hypotheses are tested using the Wald test. Subjects without publicly available PhD graduation year are dropped in (5) and (6). Since the random assignment is conducted at the study level, the standard errors in parentheses are clustered at the study level; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable, $c_i \in \{0, 1\}$, indicates the intent to contribute, with $c_i = 1$ denoting an deliberate action to click on the contribution link, and 0 otherwise. Article attributes include the dummies for the year of publication, the journal outlet, the number of references, and the relative position of the article on the network.

In Table 3.4, since all the coefficients for the treatment dummies are negative, we do not have evidence to reject our one-sided test in Hypothesis 1. Actually, we observe a reversal of the treatment effect, where participants treated with either the T_p or T_{p+s} message were less likely to contribute to this experiment than other participants. On average, participants who received the treatment messages were less willing to contribute to metadata by 5%.

To test Hypothesis 2, we perform a series of Wald tests comparing the coefficients for the two treatment dummies. With a null hypothesis that the coefficient for the two treatment dummies are identical, the p-values for Wald tests for Columns (2), (4), (6) and (8) in **??** are $0.51, 0.53, 0.84$ and 0.97, respectively. That is, we do not find evidence that adding additional text mentioning the "social benefit" led to noticeable differences in the degree of willingness to contribute. This null result can be explained in part by how the "social benefit" element was added; i.e., at the end of the paragraph that introduced the "private benefit" of contributing metadata, a short sentence was added to the end of the treatment message to introduce the "social benefit" of improved metadata. The two treatment messages are hard to tell apart at first glance.

Furthermore, we observe a significant and negative predicted margin for the number of coauthors in a study. This is consistent with the idea that it would take only one participant to complete the metadata for a study. That is, the more coauthors there were for a study, the less likely a specific author was to opt to provide metadata. When we apply the same specification using the study-level data, we observe a positive predicted margin for the number of authors, again confirming this idea.

Last, both the reported predicted margins for the female dummy variable and the interaction term are insignificant in Table 3.4. This indicates that female participants were as likely to contribute to metadata as male participants, and there is no difference in terms of how the number of coauthors affects the likelihood of contributing based on gender.

3.5.2.2 Participation of the Treated Sample

Although we intended to treat a total of 3,023 participants based on the random assignment schedule, in reality, we were only able to send out emails to 3,003 participants. The 20 participants whom we failed to reach by email either were deceased or did not have publicly available email addresses. Furthermore, even among those who received an email from us, not all opened and read the emails. Based on the records collected from the email service provider across multiple waves of reminders, we identified the treated sample as those who had opened our emails at least one time across the whole experiment. In this section, we reproduce the results presented in Table 3.4 using the "treated sample".

First, Table 3.5 reports the balancedness check of the treatment assignments across article attributes and participant attributes in the treated sample with a total of 2,226 participants. In reporting the attributes of the studies, we include all studies with at least one author who opened the email. There are a total of 1,336 studies. Notably, in Table 3.5, our initial random assignment no longer holds for the treated sample when considering the variable of years since obtaining a PhD. We address this concern by explicitly controlling for this variable in the full model.

Table 3.5: Attributes of treated participants and studies with treated author

	Control (1)	Findability + Private (2)	Findability + Social (3)	p-value (4)
Panel A: Article attributes				
NumReferences	27.107	25.546	27.057	0.291
	(0.844)	(0.711)	(0.837)	
NumCoauthors	2.112	2.143	2.118	0.885
	(0.044)	(0.054)	(0.044)	
Journals:				
AEA Papers and Proceedings	0.045	0.060	0.050	0.559
	(0.010)	(0.011)	(0.010)	
AEJ: Applied Economics	0.110	0.119	0.145	0.256
	(0.015)	(0.015)	(0.017)	
AEJ: Economic Policy	0.148	0.123	0.158	0.304
	(0.017)	(0.016)	(0.017)	
AEJ: Macroeconomics	0.096	0.134	0.104	0.163
	(0.014)	(0.016)	(0.015)	
AEJ: Microeconomics	0.060	0.060	0.045	0.523
	(0.011)	(0.011)	(0.010)	
American Economic Review	0.506	0.456	0.452	0.209
	(0.024)	(0.024)	(0.024)	
AER: Insights	0.000	0.007	0.005	0.246
	(0.000)	(0.004)	(0.003)	
Journal of Economic Literature	0.004	0.000	0.005	0.365
	(0.003)	(0.000)	(0.003)	
Journal of Economic Perspectives	0.031	0.040	0.036	0.772
	(0.008)	(0.009)	(0.009)	
PubYear*	2014.293	2014.181	2014.403	0.657
	(0.168)	(0.175)	(0.169)	
Observations	447	447	442	
Panel B: Author attributes				
Female	0.240	0.231	0.216	0.542
	(0.016)	(0.015)	(0.015)	
Years since PhD	16.161	17.761	17.078	0.017
	(0.388)	(0.414)	(0.386)	
Observations	751	765	750	

[1] Columns (1) through (3) report average values and column (4) reports the p-value testing the joint orthogonality across treatments. Standard errors are provided in parentheses.
[2] Article attributes used for block-random assignment are marked with an astroid.

In Table 3.6, we examine the intent to contribute under the same specification as that described in Table 3.4 with the treated sample. First, all the coefficients for the treatment dummies are still significant and negative. In terms of magnitude, the negative coefficients are actually smaller than the coefficients from the full sample. With the treated sample, we observe a reversal of the treatment effect again and find that the negative impact of treatment messages is more pronounced among the treated sample.

Second, to address Hypothesis 2, we perform a series of Wald tests comparing the coefficients for the two treatment dummies. With a null hypothesis that the coefficient for the two treatment dummies are identical, the p-values for the Wald tests for Columns (2), (4), (6) and (8) in Table 3.6 are $0.55, 0.56, 0.81$ and 0.87, respectively. Again, we do not find evidence that adding additional text mentioning the "social benefit" increased/decreased the participation rate.

Table 3.6: Intent to contribute (treated sample)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
treatments	-0.0579**		-0.0585**		-0.0648***		-0.0668***	
	(0.0232)		(0.0231)		(0.0239)		(0.0237)	
find_private (T_p)		-0.0659**		-0.0661**		-0.0682**		-0.0691**
		(0.0269)		(0.0269)		(0.0278)		(0.0274)
find_private_social (T_{p+s})		-0.0499*		-0.0506*		-0.0614**		-0.0645**
		(0.0267)		(0.0267)		(0.0275)		(0.0273)
female			-0.0319	-0.0317	-0.0282	-0.0280	-0.0232	-0.0231
			(0.0249)	(0.0249)	(0.0263)	(0.0263)	(0.0267)	(0.0267)
NumCoauthors							-0.0289***	-0.0288***
							(0.0078)	(0.0078)
Female×NumCoauthors							-0.0004	-0.0003
							(0.0198)	(0.0198)
Years_since_PhD							0.0007	0.0007
							(0.0012)	(0.0012)
Article_Attr	no	no	no	no	no	no	yes	yes
χ^2-stats (T_p vs T_{p+s})		0.353		0.333		0.058		0.028
Observations	2266	2266	2266	2266	2067	2067	2067	2067

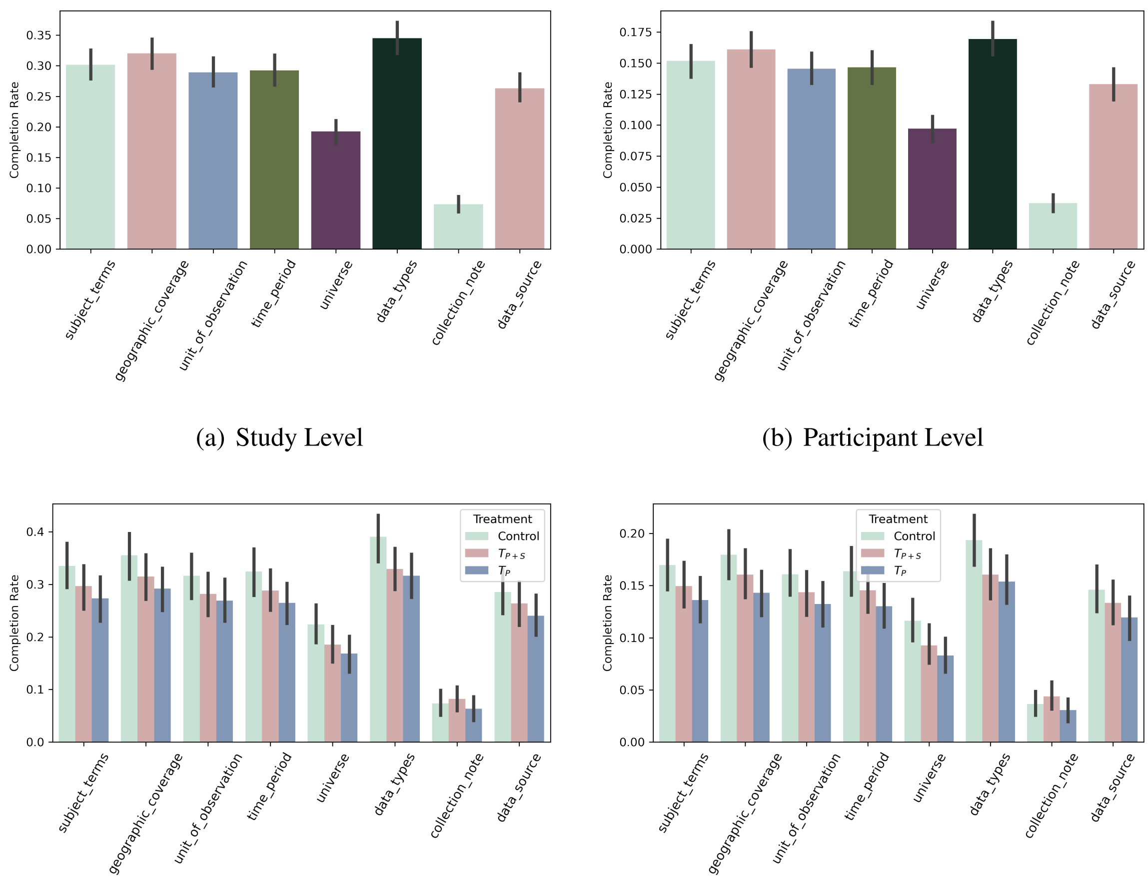
Notes: The table reports different specifications for Probit regression. Average marginal effects are reported; the marginal effects for interaction terms are the difference between the average marginal effect of NumCoauthors for female and male subjects; the standard errors are calculated using the Delta method (Ai and Norton, 2003), and hypotheses are tested using the Wald test. Subjects without publicly available PhD graduation year are dropped in (5) and (6). Since the random assignment is conducted at the study level, the standard errors in parentheses are clustered at the study level; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable, $c_i \in \{0, 1\}$, indicates the intent to contribute, with $c_i = 1$ denoting an deliberate action to click on the contribution link, and 0 otherwise. Article attributes include the dummies for the year of publication, the journal outlet, the number of references, and the relative position of the article on the network.

3.5.3 Contribution Quantity

3.5.3.1 Overview of Contribution to Metadata Fields

As detailed in Appendix 3.A.1, we asked the authors to contributed to a total of eight metadata fields. In Figure 3.2, we summarize the overall and treatment-specific completion rate across the eight metadata fields. Observations at the study level and at the author level are both populated to provide a full picture of the aggregate, as well as individual contribution to the metadata fields. Across the metadata fields, both at the study level and participant level, completion rates to the data types field are highest. An intuitive explanation is that, unlike other metadata fields that are text boxes that admit freeform text input, the "Data Types" field uses controled vocabulary such that participants only need to check a few boxes. As such, it is both easier and more accurate to choose the specific data types that apply to the data deposit. At the other extreme, the collection_note field attracted remarkably low contribution. This outcome is in line with the argument that controlled vocabulary enhances contribution, as the collection_note field is the least well-defined field across all eight metadata fields.

As shown in Figure 3.2 Panel (c) and Panel (d), when compared to the completion rate in the control group for the two treatment conditions, we observe a steady decline in the completion rate across all eight metadata fields. Although the differences in completion rate are not significant, the T_P treatment (findability text + private benefit only) yields the lowest completion rate among all three treatment conditions.



(a) Study Level

(b) Participant Level

(c) Study Level by treatment

(d) Participant Level by treatment

Figure 3.2: Completion Rate by Metadata Fields

3.5.3.2 Testing Metadata Contribution at the Study Level

At the study level, we pooled the individual contributions for each study and counted the number of populated metadata fields. In Table 3.7, contradictory to what we hypothesized in H3, we find evidence that both of the treatment messages reduced the contribution rate at the study level. Additionally, a series of F tests for the coefficients of the two treatment dummies revealed no difference between the two conditions; the p values for the corresponding F tests in Columns (2), (3), (5) and (6) are 0.58, 0.65, 0.73 and 0.81, respectively.

As a sanity check, in Table 3.7 Column (4) - (6), we see a small but significant coefficient for the number of coauthors in the study. If the number of coauthors increases by one, on average, the study will have an additional 1/5 of a metadata field populated[6].

Contradictory to what we hypothesized in Hypothesis 3, we found evidence that both of the treatment messages drove down the contribution outcome at the study level. Also, a series of F-tests for the coefficients of the two treatment dummies revealed no difference between the two conditions; the p-values for the corresponding F-tests in Columns (3), (4), (7) and (8) are 0.417, 0.405, 0.619 and 0.579, respectively. This indicates that there is no detectable difference between the two treatment messages.

As a sanity check, in Table 3.7 Column (5) - (8), we see a small but significant coefficient for the number of coauthors in the study. If the number of coauthors increases by one, on average, the study will have an additional 1/5 of a metadata field populated.

[6]Note that the data on contribution quantity feature a semicontinuous distribution with a mass at the origin, as 64.25% of the studies received zero metadata contribution. Such a large number of zeros would make the assumption of normality inappropriate. To overcome this issue, we fitted the data with an exponential dispersion model that assumes that the variance of the outcome is a power function of the mean (Jørgensen, 1987).

Table 3.7: Study (total) contribution quantity: treatment effect at the study level

	Dependent variable: contribution_quantity							
	OLS	Exp. Disp.	OLS	Exp. Disp.	OLS	Exp. Disp.	OLS	Exp. Disp.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
treatments	-0.340**	-0.160**			-0.357**	-0.164**		
	(0.164)	(0.077)			(0.166)	(0.079)		
find_private (T_p)			-0.417**	-0.200**			-0.405**	-0.192**
			(0.190)	(0.092)			(0.192)	(0.094)
find_private_social (T_{p+s})			-0.263	-0.121			-0.310	-0.138
			(0.190)	(0.090)			(0.191)	(0.092)
NumCoauthors					0.200**	0.081**	0.200**	0.081**
					(0.082)	(0.033)	(0.082)	(0.034)
female_fraction					-0.362	-0.185	-0.360	-0.184
					(0.229)	(0.118)	(0.229)	(0.118)
avg_Years_since_PhD					-0.010	-0.005	-0.010	-0.005
					(0.010)	(0.005)	(0.010)	(0.005)
Article_Attr	no	no	no	no	yes	yes	yes	yes
χ^2-stats (T_p vs T_{p+s})			0.659	0.695			0.247	0.309
Observations	1460	1460	1460	1460	1417	1417	1417	1417

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

The table reports different specifications of OLS and Exponential Dispersion Model. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.5.3.3 Treatment Effect on the Intensive Margin

We continue the discussion of the intensity of contribution at the individual level, where we examine the intensive margin for the impact of our treatment variables. In Table 3.8, we present a similar result as that reported in Table 3.7, where we observe a similar backfire result. Here, when treated with an additional paragraph of text in the treatment email, the participants ended up providing adequate information in fewer fields. Last, when testing for the difference in effect size between the two treatments conditions, there is no detectable difference according to the reported F-stats in Table 3.8.

Additionally, in Table 3.8, we report a negative and significant coefficient for the number of coauthors in the study. If the number of coauthors increases by one, on average, each author in the study will contribute less metadata content by 1/5 of a metadata field to be exact. Together with the positive coefficient reported in Table 3.7, we conclude that coauthors complement each other in metadata provision.

Table 3.8: Individual contribution quantity: treatment effects on the intensive margin

	Dependent variable: contribution_quantity							
	OLS	Exp. Disp.	OLS	Exp. Disp.	OLS	Exp. Disp.	OLS	Exp. Disp.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
treatments	-0.188**	-0.175**			-0.159*	-0.151*		
	(0.088)	(0.084)			(0.095)	(0.090)		
find_private (T_p)			-0.238**	-0.228**			-0.164	-0.167
			(0.100)	(0.100)			(0.109)	(0.106)
find_private_social (T_{p+s})			-0.137	-0.125			-0.153	-0.136
			(0.102)	(0.098)			(0.108)	(0.104)
female					-0.071	-0.072	-0.071	-0.071
					(0.115)	(0.106)	(0.115)	(0.106)
NumCoauthors					-0.195***	-0.292***	-0.195***	-0.293***
					(0.064)	(0.050)	(0.064)	(0.050)
Years_since_PhD					-0.023***	-0.025***	-0.023***	-0.025***
					(0.004)	(0.005)	(0.004)	(0.005)
Article_Attr	no	no	no	no	yes	yes	yes	yes
χ^2-stats (T_p vs T_{p+s})			1.048	0.999			0.011	0.077
Observations	3023	3023	3023	3023	2684	2684	2684	2684

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

The table reports different specifications of OLS and Exponential Dispersion Model. For OLS results, standard errors in parentheses are clustered at the study level; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.5.4 Testing Metadata Contribution Based on the Treated Sample

Among all the participants who opened the treatment email, since the subject line was identical across all treatment conditions, the treated sample is actually balanced in our setting. In Table

3.9 and 3.10, we find evidence that is consistent with what we discussed based on the full intent-to-treat sample, where both treatments drive down the contribution quantity, and the treatment with private benefit introduces a significant reduction in contribution quantity by 1/5 to 2/5 metadata fields. Such an unexpected result still holds both at the individual level and at the study level. Additionally, in line with the previous conclusion, the coefficient of the number of coauthors is positive for the study-level result and negative for the individual-level result. This again affirms the previous finding that coauthors are complementary to each other in terms of metadata provision.

Table 3.9: Study with authors who opened the email: treatment effect at the study level

	Dependent variable: contribution_quantity							
	OLS	Exp. Disp.	OLS	Exp. Disp.	OLS	Exp. Disp.	OLS	Exp. Disp.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
treatments	-0.360**	-0.155**			-0.369**	-0.157**		
	(0.175)	(0.076)			(0.176)	(0.078)		
find_private (T_p)			-0.454**	-0.200**			-0.423**	-0.184**
			(0.202)	(0.089)			(0.204)	(0.092)
find_private_social (T_{p+s})			-0.266	-0.112			-0.316	-0.130
			(0.203)	(0.088)			(0.204)	(0.090)
NumCoauthors					0.112	0.046	0.112	0.047
					(0.086)	(0.035)	(0.086)	(0.035)
female_fraction					-0.397	-0.183	-0.395	-0.181
					(0.248)	(0.116)	(0.248)	(0.116)
avg_Years_since_PhD					-0.012	-0.005	-0.012	-0.005
					(0.011)	(0.005)	(0.011)	(0.005)
Article_Attr	no	no	no	no	yes	yes	yes	yes
χ^2-stats (T_p vs T_{p+s})			0.864	0.908			0.275	0.321
Observations	1336	1336	1336	1336	1301	1301	1301	1301
Note:								*p<0.1; **p<0.05; ***p<0.01

The table reports different specifications of OLS and Exponential Dispersion Model.

Table 3.10: Individual contribution quantity: treatment effects on the intensive margin

	Dependent variable: contribution_quantity							
	OLS	Exp. Disp.	OLS	Exp. Disp.	OLS	Exp. Disp.	OLS	Exp. Disp.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
treatments	-0.268**	-0.188**			-0.234*	-0.165*		
	(0.114)	(0.082)			(0.121)	(0.087)		
find_private (T_p)			-0.332**	-0.239**			-0.234*	-0.172*
			(0.130)	(0.097)			(0.138)	(0.103)
find_private_social (T_{p+s})			-0.202	-0.138			-0.234*	-0.158
			(0.131)	(0.095)			(0.136)	(0.101)
female					-0.078	-0.062	-0.078	-0.062
					(0.142)	(0.102)	(0.142)	(0.102)
NumCoauthors					-0.238***	-0.248***	-0.238***	-0.248***
					(0.074)	(0.048)	(0.074)	(0.048)
Years_since_PhD					-0.029***	-0.023***	-0.029***	-0.023***
					(0.005)	(0.005)	(0.005)	(0.005)
Article_Attr	no	no	no	no	yes	yes	yes	yes
χ^2-stats (T_p vs T_{p+s})			1.058	1.005			0.0	0.016
Observations	2266	2266	2266	2266	2067	2067	2067	2067

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

The table reports different specifications of OLS and Exponential Dispersion Model. For OLS results, standard errors in parentheses are clustered at the study level; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.5.5 Delegation and Gender

In this section, we test Hypothesis 4 by revisiting the four possible outcomes at the first stage of the experiment where participants interacted with hyperlinks embedded in the body of the treatment email. Among the three actionable options in the treatment email, the second hyperlink offered authors the opportunity to delegate the metadata-contribution task to their coauthors. According to the summary statistics in Table 3.3, at least 10% of the participants chose this delegation option. Upon receiving the treatment email, participants could click on a link to provide the metadata, click on another link to opt out of the experiment by delegating the task to their coauthors, or click on a third link to opt out of the experiment entirely. If none of these clicking activities were detected, we labeled the author's intention as "No Response".

Since choosing the delegation option only makes sense when there are at least two authors in the study, in the results that follow, we disregard all single-authored papers. This approach leaves us with 2,598 participants who were coauthors of studies included in the experiment.

In Table 3.11, we present the results of two multinomial logistic models. In Column (3), before controlling for additional author and article attributes, female participants are approximately 4% less likely to choose the delegation option than male participants. However, in Column (7), the predictive margin for being female is no longer significant when predicting the delegation outcome. In turn, the likelihood for delegating to coauthors increases by the seniority of the author; i.e., with each year since an author obtained his/her PhD degree, there is a 0.5% increase in the likelihood that the author will delegate the metadata contribution task to his/her coauthors.

Table 3.11: Multinomial logistic models with delegation result

	(1) NoResponse	(2) NotInterested	(3) Delegate	(4) Intend-to-edit	(5) NoResponse	(6) NotInterested	(7) Delegate	(8) Intend-to-edit
female	0.0128	0.0191	-0.0393***	0.0074	0.0031	0.0196	-0.0210	-0.0018
	(0.0241)	(0.0170)	(0.0152)	(0.0197)	(0.0266)	(0.0188)	(0.0183)	(0.0216)
find_private (T_p)	0.0512**	-0.0115	-0.0084	-0.0313*	0.0414	-0.0022	-0.0192	-0.0201
	(0.0244)	(0.0178)	(0.0156)	(0.0189)	(0.0260)	(0.0196)	(0.0167)	(0.0215)
find_private_social (T_{p+s})	0.0204	0.0060	-0.0292*	0.0027	0.0221	0.0164	-0.0409**	0.0024
	(0.0238)	(0.0183)	(0.0153)	(0.0192)	(0.0255)	(0.0201)	(0.0159)	(0.0210)
NumCoauthors	0.0434***	-0.0258*	0.0139*	-0.0316***	0.0398***	-0.0212	0.0130	-0.0316***
	(0.0121)	(0.0150)	(0.0076)	(0.0093)	(0.0130)	(0.0154)	(0.0094)	(0.0117)
Years_since_PhD					-0.0002	0.0007	0.0048***	-0.0053***
					(0.0010)	(0.0007)	(0.0006)	(0.0009)
female$\times T_p$	-0.0415	-0.0307	0.0468	0.0254	-0.0182	-0.0166	0.0209	0.0139
	(0.0587)	(0.0405)	(0.0363)	(0.0472)	(0.0629)	(0.0449)	(0.0402)	(0.0508)
female$\times T_{p+s}$	-0.0099	-0.0163	-0.0182	0.0443	0.0225	0.0020	-0.0512	0.0267
	(0.0580)	(0.0404)	(0.0344)	(0.0486)	(0.0633)	(0.0463)	(0.0385)	(0.0523)
female×NumCoauthors	0.0501	-0.0400	0.0016	-0.0118	0.0526	-0.0440	0.0009	-0.0095
	(0.0278)	(0.0279)	(0.0080)	(0.0218)	(0.0286)	(0.0296)	(0.0127)	(0.0249)
Article_Attr	no	no	no	no	yes	yes	yes	yes
Observations	2598	2598	2598	2598	2286	2286	2286	2286

Standard errors in parentheses are clustered at the study level; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Notes: Two mlogit models were employed here. Since there are four potential outcomes at the first stage, the average marginal effects are reported for each outcome separately. Average marginal effects are reported; the marginal effects for interaction terms are the difference between the average marginal effect of the treatment dummy variable for females and that for males; standard errors are calculated using the Delta method (Ai and Norton, 2003); and hypotheses are tested using the Wald test.

3.6 Discussion

Across all treatment conditions in this study, we collected a total of 3,149 metadata entries from 544 authors and completed missing metadata fields in a total of 522 studies. Overall, the rate of responding to our treatment emails is very high; specifically, close to 40% of the participants interacted with the clickable entries available in the email. We attribute this high response rate to the fact that all emails were sent from an authoritative figure, namely, the AEA Data Editor.

From a metadata collection perspective, this study demonstrates that the "crowdsourcing approach" can be an effective way to solicit metadata contribution from experts who have conducted the studies themselves. While it may take traditional archivists hundreds of hours (Thomer et al., 2022) to accomplish this task at such a scale by contacting the authors of the original studies, we demonstrate that collecting metadata contributions is both feasible and productive at scale and can serve as a low-cost way to address missing metadata issues in the future.

To further comprehend the reversal of the treatment effects that went against our original hypotheses, we consider the following explanations. First, in line with the survey literature documenting the fact that people are more likely to break off as survey length increases (Mittereder and West, 2021; Mavletova and Couper, 2015; Heerwegh and Loosveldt, 2006; Peytchev, 2011; Peytchev, 2009; Steinbrecher, Roßmann, and Blumenstiel, 2015), one hypothesis is that participants in this experiment took the treatment email as a survey form and were thus more likely to break off from the experiment as the length of the treatment message increased. Second, in Figure 3.8, we document the way our treatment messages were rendered on mobile devices. If a recipient used a large font-size for the email client on the mobile phone, only the first paragraph of the treatment email would have been shown, and the emails would read identically across all three treatment conditions. With the default font-size, participants who received the control message might have seen actionable items to the bottom of the screen. For the other two treatment conditions, the paragraph implementing the treatment took up all the space on the screen, leaving the participants with nothing to "click on" upon the first glance at the email. We speculate that this may have been one driving factor for the ineffectiveness of the treatment message. Note, however, that this is mainly speculation, as we do not have a record of the type of devices participants used to access our treatment emails.

Finally, in this study, we adopted an innovative treatment assignment method to eliminate possible spillovers of treatments and achieved a high level of precision where we delivered individually customized treatment emails to the participants. As documented in Appendix 3.D, the whole infrastructure can be of value for future studies if reused by other researchers or data deposits, as it facilitates the collection of metadata at large scale.

Bibliography

Ahmed, Syed Ishtiaque et al. (2016). "Peer-to-peer in the Workplace: A View from the Road". In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. CHI '16. San Jose, California, USA: ACM, pp. 5063–5075. DOI: 10.1145/2858036.2858393.

Ai, Chunrong and Edward C Norton (2003). "Interaction terms in logit and probit models". In: *Economics letters* 80.1, pp. 123–129.

Akerlof, George A. and Rachel E. Kranton (2000). "Economics and Identity". In: *The Quarterly Journal of Economics* 115.3, pp. 715–753. DOI: 10.1162/003355300554881.

Andreoni, James and B Douglas Bernheim (2009). "Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects". In: *Econometrica* 77.5, pp. 1607–1636.

Ariely, Dan et al. (2009). "Doing good or doing well? Image motivation and monetary incentives in behaving prosocially". In: *The American Economic Review* 99.1, pp. 544–555.

Autor, David and Susan Houseman (2005). *Do Temporary Help Jobs Improve Labor Market Outcomes for Low-skilled Workers? Evidence from Random Assignments*. National Bureau of Economic Research.

Babcock, Linda et al. (2017). "Gender Differences in Accepting and Receiving Requests for Tasks with Low Promotability". In: *American Economic Review* 107.3, pp. 714–47. DOI: 10.1257/aer.20141734.

Bahler, Kristen (2017). "Unemployment Is Really Low. So Why Can't These People Find Jobs?" In: ed. by Money. [Accessed: 2019-08-15].

Bear, Julia B. and Benjamin Collier (Mar. 2016). "Where Are the Women in Wikipedia? Understanding the Different Psychological Experiences of Men and Women in Wikipedia". In: *Sex Roles* 74.5, pp. 254–265. DOI: 10.1007/s11199-015-0573-y.

Bénabou, Roland and Jean Tirole (2006). "Incentives and Prosocial Behavior". In: *American Economic Review* 96.5, pp. 1652–1678. DOI: 10.1257/aer.96.5.1652.

Benz, Matthias and Stephan Meier (Feb. 2008). “Do people behave in experiments as in the field?—evidence from donations”. In: *Experimental Economics* 11.3, pp. 268–281. DOI: 10.1007/s10683-007-9192-y.

Bergstrom, Theodore et al. (1986). “On the private provision of public goods”. In: *Journal of Public Economics* 29.1, pp. 25–49.

Bernanke, Ben S. (2004). “Editorial Statement”. In: *The American Economic Review* 94.1, pp. 404–404.

Bertrand, Marianne and Esther Duflo (2016). *Field Experiments on Discrimination*. Working Paper 22014. National Bureau of Economic Research. DOI: 10.3386/w22014.

Bertrand, Marianne and Sendhil Mullainathan (2004). “Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination”. In: *American Economic Review* 94.4, pp. 991–1013.

Bielby, William T and James N Baron (1986). “Men and women at work: Sex segregation and statistical discrimination”. In: *American Journal of Sociology* 91.4, pp. 759–799.

Bishop, Bradley Wade et al. (2019). “Scientists’ Data Discovery and Reuse Behavior: (Meta)Data Fitness for Use and the FAIR Data Principles”. In: *Proceedings of the Association for Information Science and Technology* 56.1, pp. 21–31. DOI: 10.1002/pra2.4.

Blondel, Vincent D. et al. (Oct. 9, 2008). “Fast Unfolding of Communities in Large Networks”. In: *Journal of Statistical Mechanics: Theory and Experiment* 2008.10, P10008. DOI: 10.1088/1742-5468/2008/10/P10008. arXiv: 0803.0476.

Borgman, Christine L. et al. (Feb. 2018). “Digital Data Archives as Knowledge Infrastructures: Mediating Data Sharing and Reuse”. In.

Bowers, John et al. (1995). “Workflow from within and without: Technology and cooperative work on the print industry shopfloor”. In: *Proceedings of the Fourth European Conference on Computer-supported Cooperative Work ECSCW’95*. Springer, pp. 51–66.

Buser, Thomas et al. (May 2014). “Gender, Competitiveness, and Career Choices ∗”. In: *The Quarterly Journal of Economics* 129.3, pp. 1409–1447. DOI: 10.1093/qje/qju009.

Carlsson, Magnus (2011). “Does hiring discrimination cause gender segregation in the Swedish labor market?” In: *Feminist Economics* 17.3, pp. 71–102.

Chapman, Adriane et al. (Jan. 2020). “Dataset Search: A Survey”. In: *The VLDB Journal* 29.1, pp. 251–272. DOI: 10.1007/s00778-019-00564-x.

Chen, Daniel L et al. (2016a). “oTree—An open-source platform for laboratory, online, and field experiments”. In: *Journal of Behavioral and Experimental Finance* 9, pp. 88–97.

Chen, Daniel L. et al. (Mar. 2016b). “oTree—An Open-Source Platform for Laboratory, Online, and Field Experiments”. In: *Journal of Behavioral and Experimental Finance* 9, pp. 88–97. DOI: 10.1016/j.jbef.2015.12.001.

Chen, M. Keith et al. (2017). *The value of flexible work: Evidence from Uber drivers*. Tech. rep. National Bureau of Economic Research.

Chen, Roy and Yan Chen (2011). “The Potential of Social Identity for Equilibrium Selection”. In: *The American Economic Review* 101.6, pp. 2562–2589.

Chen, Yan (2008). “Incentive-Compatible Mechanisms for Pure Public Goods: A Survey of Experimental Research”. In: *The Handbook of Experimental Economics Results*. Ed. by Charles Plott and Vernon Smith. Vol. 1. Amsterdam: North-Holland, pp. 625 –643.

Chen, Yan et al. (Apr. 2020a). *Motivating Experts to Contribute to Digital Public Goods: A Personalized Field Experiment on Wikipedia*. SSRN Scholarly Paper ID 3588132. Rochester, NY: Social Science Research Network. DOI: 10.2139/ssrn.3588132.

Chen, Yan et al. (2020b). “Motivating Metadata Contributions for Data Re-Use and Reproducibility”. In: *AEA RCT Registry*. DOI: 10.1257/rct.6159-1.0.

— (2020c). *Motivating Metadata Contributions for Data Re-Use and Reproducibility*. DOI: 10.1257/rct.6159-1.0.

Christensen, Garret et al. (2019). *Transparent and Reproducible Social Science Research: How to Do Open Science*. University of California Press. DOI: doi:10.1525/9780520969230.

Coffman, Katherine Baldiga (Nov. 1, 2014). “Evidence on Self-Stereotyping and the Contribution of Ideas”. In: *The Quarterly Journal of Economics* 129.4, pp. 1625–1660. DOI: 10.1093/qje/qju023.

Coffman, Lucas C. and Muriel Niederle (Sept. 2015). “Pre-Analysis Plans Have Limited Upside, Especially Where Replications Are Feasible”. In: *Journal of Economic Perspectives* 29.3, pp. 81–98. DOI: 10.1257/jep.29.3.81.

Comenetz, Joshua (2016). “Frequently occurring surnames in the 2010 Census”. In: *United States Census Bureau*.

Cook, Cody et al. (2018). *The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers*. Tech. rep. National Bureau of Economic Research.

Cosley, Dan et al. (2007). “SuggestBot: using intelligent task routing to help people find work in wikipedia”. In: *Proceedings of the 12th international conference on Intelligent user interfaces*. Downloaded on February 23, 2003 at http://www.communitytechnology.org/nsf_ci_report/, pp. 32 –41.

Cox, D. R. (1958). *Planning of Experiments*. A Wiley Publication in Applied Statistics. John Wiley & Sons.

Daniels, Morgan et al. (2012). “Managing Fixity and Fluidity in Data Repositories”. In: *Proceedings of the 2012 iConference on - iConference ’12*. Toronto, Ontario, Canada: ACM Press, pp. 279–286. DOI: 10.1145/2132176.2132212.

Dillahunt, Tawanna R. (2014). “Fostering social capital in economically distressed communities”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, pp. 531–540.

Dillahunt, Tawanna R. and Amelia R. Malone (2015). “The Promise of the Sharing Economy among Disadvantaged Communities”. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. CHI ’15. Seoul, Republic of Korea: ACM, pp. 2285–2294. DOI: 10.1145/2702123.2702189.

Dillahunt, Tawanna R. et al. (2016a). “Designing for Disadvantaged Job Seekers: Insights from Early Investigations”. In: *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*. DIS ’16. Brisbane, QLD, Australia: ACM, pp. 905–910. DOI: 10.1145/2901790.2901865.

Dillahunt, Tawanna R. et al. (2016b). “Do Massive Open Online Course Platforms Support Employability?” In: *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. CSCW ’16. San Francisco, California, USA: ACM, pp. 233–244. DOI: 10.1145/2818048.2819924.

Dillahunt, Tawanna R. et al. (Dec. 2017). “The Sharing Economy in Computing: A Systematic Literature Review”. In: *Proceedings of the ACM on Human-Computer Interaction* 1.CSCW, 38:1–38:26. DOI: 10.1145/3134673.

Dillahunt, Tawanna R. et al. (2018). “Designing Future Employment Applications for Underserved Job Seekers: A Speed Dating Study”. In: *Proceedings of the 2018 Designing Interactive Systems Conference*. ACM, pp. 33–44.

Dokko, Jane et al. (2015). “Workers and the Online Gig Economy. A Hamilton Project Framing Paper”. In: *The Hamilton Project: Advancing Opportunity, Prosperity, and Growth*.

Dombrowski, Lynn et al. (2017). “Low-Wage Precarious Workers’ Sociotechnical Practices Working Towards Addressing Wage Theft”. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. CHI ’17. Denver, Colorado, USA: ACM, pp. 4585–4598. DOI: 10.1145/3025453.3025633.

Eckel, Catherine C. and Philip J. Grossman (2005). “Managing Diversity by Creating Team Identity”. In: *Journal of Economic Behavior & Organization* 58.3, pp. 371–392.

Farrell, Diana et al. (2018). “The Online Platform Economy in 2018: Drivers, Workers, Sellers, and Lessors”. In: *JPMorgan Chase Institute*.

Fischbacher, Urs and Simon Gachter (Mar. 2010). “Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Goods Experiments”. In: *American Economic Review* 100.1, pp. 541–556. DOI: 10.1257/aer.100.1.541.

Ford, Heather and Judy Wajcman (Aug. 1, 2017). “‘Anyone Can Edit’, Not Everyone Does: Wikipedia and the Gender Gap”. In: *Social Studies of Science*.

Gaddis, S. Michael (2018). *Audit studies: Behind the scenes with theory, method, and nuance*. Vol. 14. Springer.

Gebel, Michael (2013). “Is a temporary job better than unemployment? A cross-country comparison based on British, German, and Swiss panel data”. In: *SOEP paper* 543.

Glott, Ruediger et al. (2010). *Wikipedia Survey–Overview of Results*. Tech. rep.

Granovetter, Mark S. (1995). *Getting a job: A study of contacts and careers*. Chicago: University of Chicago Press.

Greenberg, Jane et al. (Oct. 24, 2001). “Author-Generated Dublin Core Metadata for Web Resources: A Baseline Study in an Organization”. In: *International Conference on Dublin Core and Metadata Applications* 0.0 (0), pp. 38–45.

Gregory, Kathleen (July 13, 2020). “A Dataset Describing Data Discovery and Reuse Practices in Research”. In: *Scientific Data* 7.1 (1), p. 232. DOI: 10.1038/s41597-020-0569-5.

Groves, Theodore and John O. Ledyard (1987). “Incentive Compatibility since 1972”. In: *Information, Incentives and Economic Mechanisms: Essays in Honor of Leonid Hurwicz*. Ed. by Theodore Groves et al. Minneapolis: University of Minnesota Press, pp. 48–111.

Hargittai, Eszter and Aaron Shaw (Apr. 3, 2015). “Mind the Skills Gap: The Role of Internet Know-How and Gender in Differentiated Contributions to Wikipedia”. In: *Information, Communication & Society* 18.4, pp. 424–442. DOI: 10.1080/1369118X.2014.957711.

Heerwegh, Dirk and Geert Loosveldt (2006). “An Experimental Study on the Effects of Personalization, Survey Length Statements, Progress Indicators, and Survey Sponsor Logos in Web Surveys”. In: *Journal of Official Statistics* 22.2, p. 191.

Hemphill, Libby et al. (Mar. 2022). “How Do Properties of Data, Their Curation, and Their Funding Relate to Reuse?” In: *Journal of the Association for Information Science and Technology*, asi.24646. DOI: 10.1002/asi.24646.

Hendry, David G. et al. (2017). “Homeless Young People, Jobs, and a Future Vision: Community Members’ Perceptions of the Job Co-op”. In: *Proceedings of the 8th International Conference on Communities and Technologies*. C&T ’17. Troyes, France: ACM, pp. 22–31. DOI: 10.1145/3083671.3083680.

Hendry, David G. et al. (2017). "U-District Job Co-op: constructing a future vision for homeless young people and employment". In: *Information Technology & People* 30, pp. 602–628.

Holt, Charles A. and Susan K. Laury (2002). "Risk Aversion and Incentive Effects". English. In: *The American Economic Review* 92.5, pp. 1644–1655.

Hui, Julie et al. (Nov. 2018). "Making a Living My Way: Necessity-driven Entrepreneurship in Resource-Constrained Communities". In: *Proceedings of the ACM on Human-Computer Interaction* 2.CSCW, 71:1–71:24. DOI: 10.1145/3274340.

Hui, Julie S. et al. (2018). "IntroAssist: A Tool to Support Writing Introductory Help Requests". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, p. 22.

Ichino, Andrea et al. (2005). "Temporary work agencies in Italy: A springboard toward permanent employment?" In: *Giornale degli economisti e annali di economia*, pp. 1–27.

Johansson, Louise (2013). "A Study of the Motivation Behind Collaborative Knowledge Production and the Formation of Community in Web 2.0, Using the Case Study of wikiHow.Com". MA thesis.

Jørgensen, Bent (1987). "Exponential Dispersion Models". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 49.2, pp. 127–145. DOI: 10.1111/j.2517-6161.1987.tb01685.x.

Juanchich, Marie et al. (May 2015). "Cognitive Reflection Predicts Real-Life Decision Outcomes, but Not Over and Above Personality and Decision-Making Styles". In: *Journal of Behavioral Decision Making* 29.1, pp. 52–59. DOI: 10.1002/bdm.1875.

Karlan, Dean (2005). "Using Experimental Economics to Measure Social Capital and Predict Financial Decision". In: *American Economic Review* 95.5, pp. 1688–1699.

Krueger, Alan B. et al. (2014). "Are the long-term unemployed on the margins of the labor market?" In: *Brookings Papers on Economic Activity* 2014.1, pp. 229–299.

Labor Statistics, Bureau of (2019a). "Economic News Release: Employment Situation Summary". In: [Accessed: 2019-04-03].

— (2019b). "Occupational Employment and Wages Summary". In: [Accessed: 2019-04-03].

Lahno, Amrei M. et al. (Oct. 2015). "Conflicting risk attitudes". In: *Journal of Economic Behavior & Organization* 118, pp. 136–149. DOI: 10.1016/j.jebo.2015.03.003.

Laury, Susan and Laura Taylor (2008). "Altruism spillovers: Are behaviors in context-free experiments predictive of altruism toward a naturally occurring public good". In: *Journal of Economic Behavior Organization* 65.1, pp. 9–29.

Ledyard, John (1995). "Public goods: A survey of experimental research". In: *The Handbook of Experimental Economics*. Ed. by John H. Kagel and Alvin E. Roth. Vol. 1. Princeton, New Jersey: Princeton University Press.

Leider, Stephen et al. (2009). "Directed Altruism and Enforced Reciprocity in Social Networks". In: *The Quarterly Journal of Economics* 124.4, pp. 1815–1851.

Li, Linfeng and Yan Chen (2020). "Gender Inequality in Contributions to Wikipedia". In: *AEA RCT Registry*. DOI: 10.1257/rct.6753.

Mavletova, Aigul and Mick P. Couper (2015). "A Meta-Analysis of Breakoff Rates in Mobile Web Surveys". In: *Mobile research methods: Opportunities and challenges of mobile research methodologies*, pp. 81–98.

Mesgari, Mostafa et al. (2015). ""The Sum of All Human Knowledge": A Systematic Review of Scholarly Research on the Content of Wikipedia". In: *Journal of the Association for Information Science and Technology* 66.2, pp. 219–245. DOI: 10.1002/asi.23172.

Mittereder, Felicitas and Brady T West (June 2021). "A DYNAMIC SURVIVAL MODELING APPROACH TO THE PREDICTION OF WEB SURVEY BREAKOFF". In: *Journal of Survey Statistics and Methodology*, smab015. DOI: 10.1093/jssam/smab015.

Murawski, Carsten and Peter Bossaerts (Oct. 2016). "How Humans Solve Complex Problems: The Case of the Knapsack Problem". In: *Scientific Reports* 6.1. DOI: 10.1038/srep34851.

Newman, Mark (Sept. 2017). *Networks: An Introduction*. 2 edition. Oxford ; New York: Oxford University Press.

Newman, Mark EJ and Michelle Girvan (2004). "Finding and Evaluating Community Structure in Networks". In: *Physical review E* 69.2, p. 026113.

Niederle, M. and L. Vesterlund (Aug. 2007). "Do Women Shy Away From Competition? Do Men Compete Too Much?" In: *The Quarterly Journal of Economics* 122.3, pp. 1067–1101. DOI: 10.1162/qjec.122.3.1067.

Peytchev, Andy (2009). "Survey Breakoff". In: *Public Opinion Quarterly* 73.1, pp. 74–97.

— (2011). "Breakoff and Unit Nonresponse across Web Surveys". In: *Journal of Official Statistics* 27.1, p. 33.

Pienta, Amy M. et al. (Nov. 2010). "The Enduring Value of Social Science Research: The Use and Reuse of Primary Research Data". In.

Piwowar, Heather A. et al. (Mar. 21, 2007). "Sharing Detailed Research Data Is Associated with Increased Citation Rate". In: *PLoS ONE* 2.3. Ed. by John Ioannidis, e308. DOI: 10.1371/journal.pone.0000308.

Pons, Pascal and Matthieu Latapy (2006). "Computing Communities in Large Networks Using Random Walks". In: p. 28.

Rege, Mari and Kjetil Telle (2004). “The Impact of Social Approval and Framing on Cooperation in Public Good Situations”. In: *Journal of Public Economics* 88.7, pp. 1625–1644.

Reuben, Ernesto et al. (Nov. 2015). *Taste for Competition and the Gender Gap Among Young Business Professionals*. Tech. rep. DOI: 10.3386/w21695.

Saccardo, Silvia et al. (Apr. 2018). “On the Size of the Gender Difference in Competitiveness”. In: *Management Science* 64.4, pp. 1541–1554. DOI: 10.1287/mnsc.2016.2673.

Salehi, Niloufar and Michael S. Bernstein (2018). “Ink: Increasing Worker Agency to Reduce Friction in Hiring Crowd Workers”. In: *ACM Transactions on Computer-Human Interaction (TOCHI)* 25.2, p. 10.

Samuelson, Paul A. (1954). “The Pure Theory of Public Expenditure”. In: *Review of Economics and Statistics* 36.4, pp. 387–389.

Santos, Luiz et al. (Sept. 2016). “FAIR Data Points Supporting Big Data Interoperability”. In: p. 10.

Sariisik, Merve (2018). *Identity Discrimination in the Sharing Economy: A Field Experiment*. University of Michigan Working Paper.

Schnusenberg, Oliver and Andrés Gallo (2011). “On Cognitive Ability and Learning in a Beauty Contest”. In: *Journal for Economic Educators* 11.1, pp. 13–24.

Shaw, Aaron and Eszter Hargittai (Feb. 1, 2018). “The Pipeline of Online Participation Inequalities: The Case of Wikipedia Editing”. In: *Journal of Communication* 68.1, pp. 143–168. DOI: 10.1093/joc/jqx003.

Steinbrecher, Markus et al. (June 2015). “Why Do Respondents Break Off Web Surveys and Does It Matter? Results From Four Follow-up Surveys”. In: *International Journal of Public Opinion Research* 27.2, pp. 289–302. DOI: 10.1093/ijpor/edu025.

Suchman, Lucy A. (1987). *Plans and situated actions: The problem of human-machine communication*. Cambridge University Press.

Suzuki, Ryo et al. (2016). “Atelier: Repurposing expert crowdsourcing tasks as micro-internships”. In: *Proceedings of the 2016 CHI conference on human factors in computing systems*. ACM, pp. 2645–2656.

Thomer, Andrea K. et al. (Feb. 2022). “The Craft and Coordination of Data Curation: Complicating ”Workflow” Views of Data Science [PREPRINT]”. In: DOI: 10.7302/4017.

Traag, V. A. et al. (Mar. 26, 2019). “From Louvain to Leiden: Guaranteeing Well-Connected Communities”. In: *Scientific Reports* 9.1 (1), p. 5233. DOI: 10.1038/s41598-019-41695-z.

Van Belle, Eva et al. (2017). "Why is unemployment duration a sorting criterion in hiring?" In: *IZA Discussion Paper No. 10876. Available at SSRN: https://ssrn.com/abstract=2998986*.

Vesterlund, Lise (2015). "Using experimental methods to understand why and how we give to charity". In: *The Handbook of Experimental Economics*. Ed. by John H. Kagel and Alvin E. Roth. Vol. 2. Princeton, New Jersey: Princeton University Press.

Vilhuber, Lars (May 2019). "Report by the AEA Data Editor". In: *AEA Papers and Proceedings* 109, pp. 718–29. DOI: 10.1257/pandp.109.718.

Voors, Maarten J et al. (Apr. 2012). "Violent Conflict and Behavior: A Field Experiment in Burundi". In: *American Economic Review* 102.2, pp. 941–964. DOI: 10.1257/aer.102.2.941.

Wakita, Ken and Toshiyuki Tsurumi (May 8, 2007). "Finding Community Structure in Mega-Scale Social Networks: [Extended Abstract]". In: *Proceedings of the 16th International Conference on World Wide Web*. WWW '07. Banff, Alberta, Canada: Association for Computing Machinery, pp. 1275–1276. DOI: 10.1145/1242572.1242805.

Ward, Patrick S. and Vartika Singh (June 2015). "Using Field Experiments to Elicit Risk and Ambiguity Preferences: Behavioural Factors and the Adoption of New Agricultural Technologies in Rural India". In: *The Journal of Development Studies* 51.6, pp. 707–724. DOI: 10.1080/00220388.2014.989996.

Wheeler, Earnest and Tawanna R. Dillahunt (2018). "Navigating the Job Search as a Low-Resourced Job Seeker". In: *Proceedings of the 36th Annual ACM Conference on Human Factors in Computing Systems*. CHI '18. Montreal, QC, Canada: ACM. DOI: 10.1145/3173574.3173622.

Wilkinson, Mark D. et al. (Mar. 2016). "The FAIR Guiding Principles for Scientific Data Management and Stewardship". In: *Scientific Data* 3.1, p. 160018. DOI: 10.1038/sdata.2016.18.

Yaraghi, Niam and Shamika Ravi (2017). "The current and future state of the sharing economy". In: *Available at SSRN: https://ssrn.com/abstract=3041207*.

Zhang, Xiaoquan Michael and Feng Zhu (2011). "Group Size and Incentives to Contribute: A Natural Experiment at Chinese Wikipedia". In: *American Economic Review* 101.4, pp. 1601–15.

Appendices

3.A Experiment Interface

In this section, we provide screenshots of the experiment interface. Each author entera the experiment interface with an individualized link that is delivered to the author's inbox through the treatment email. Immediately after viewing the consent page, authors are presented with a view of the data deposit corresponding to the article they have published. Metadata fields are collected on the second page of the experiment interface (Figure 3.4).

3.A.1 Metadata Fields

The metadata fields we collect are designed to take into account two relevant sources, namely, the AEA guidance[7] and the current set of available fields on openICPSR.

1. Subject Terms (e.g., "Machine Learning", "Randomized Control Trial", "Nudges", etc.).
2. Geographic coverage (e.g, "United States", "Florida, U.S.", "Indonesia", ...)
3. Time period(s) (e.g., "1982-2008")
4. Universe (text field, e.g. "Adult non-institutionalized population of the United States living in households.")
5. Data Type(s) (a drop-down menu, include experimental data, observational data, survey data . . .)
6. Collection Notes (A description of technical details and other characteristics of the data collection (such as unique authoring, dissemination, or processing information) that cannot be recorded in the other metadata fields but constitute important information for the user.)

[7] See `https://aeadataeditor.github.io/aea-deguidance/data-deposit-aea-guidance.html#checklist`, Guidance on how to deposit data at the AEA Data and Code Repository.

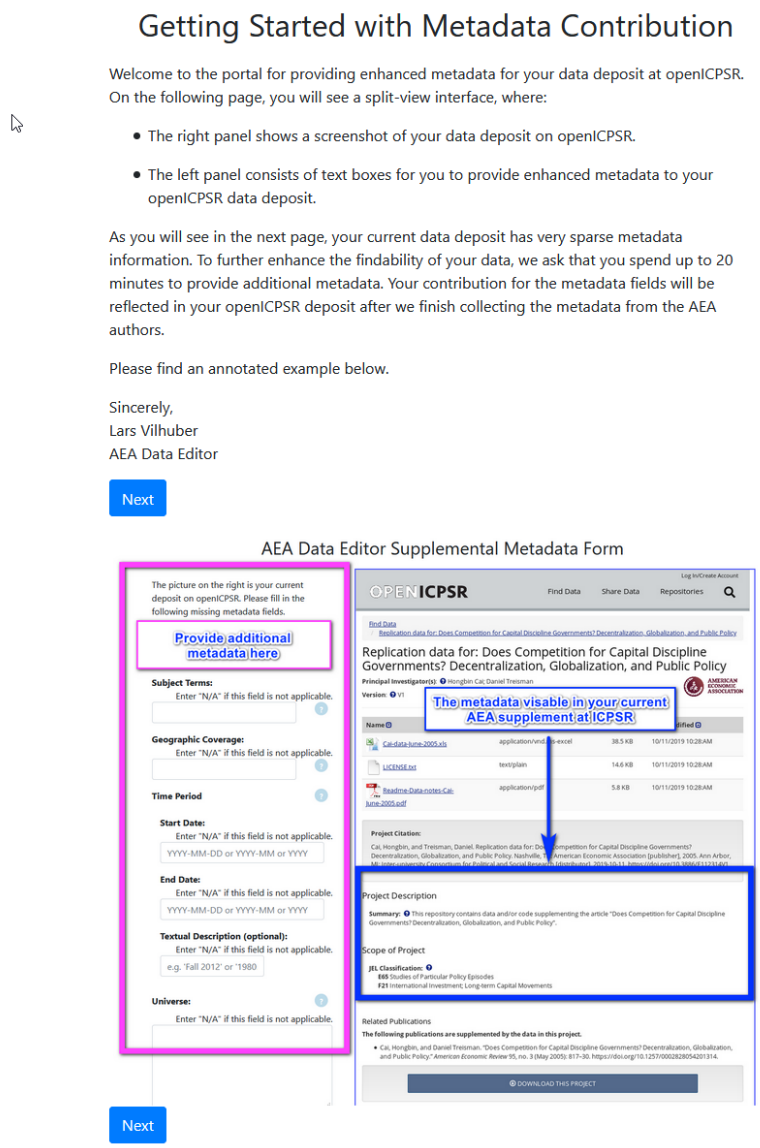


Figure 3.3: Experiment interface - Page 1: Introduction

7. Data Source

8. Unit(s) of Observation

For each metadata field, we provide a form field on the experiment interface to collect the input. Since not all metadata fields are applicable for a given study, we instruct the authors to write "N/A" if the metadata field does not apply. Additionally, each metadata field is accompanied by a "help tip", a blue icon with question marks that provide further explanation

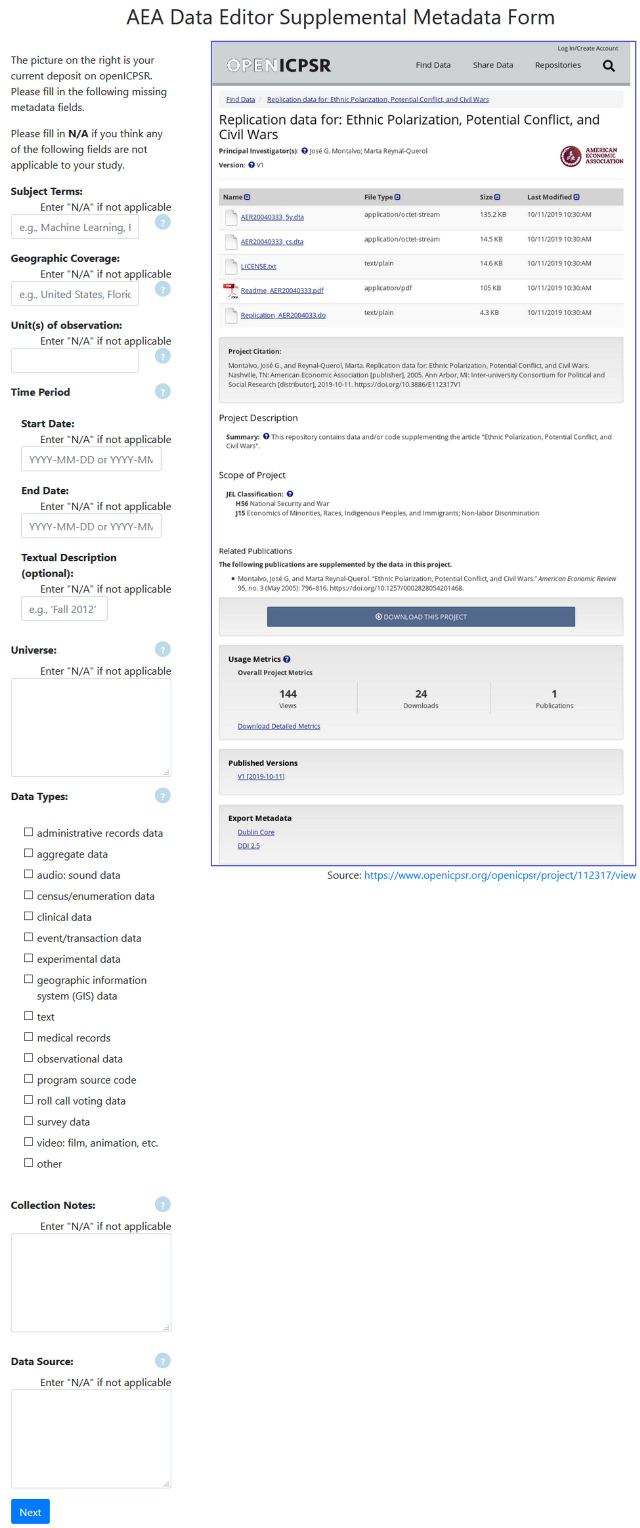
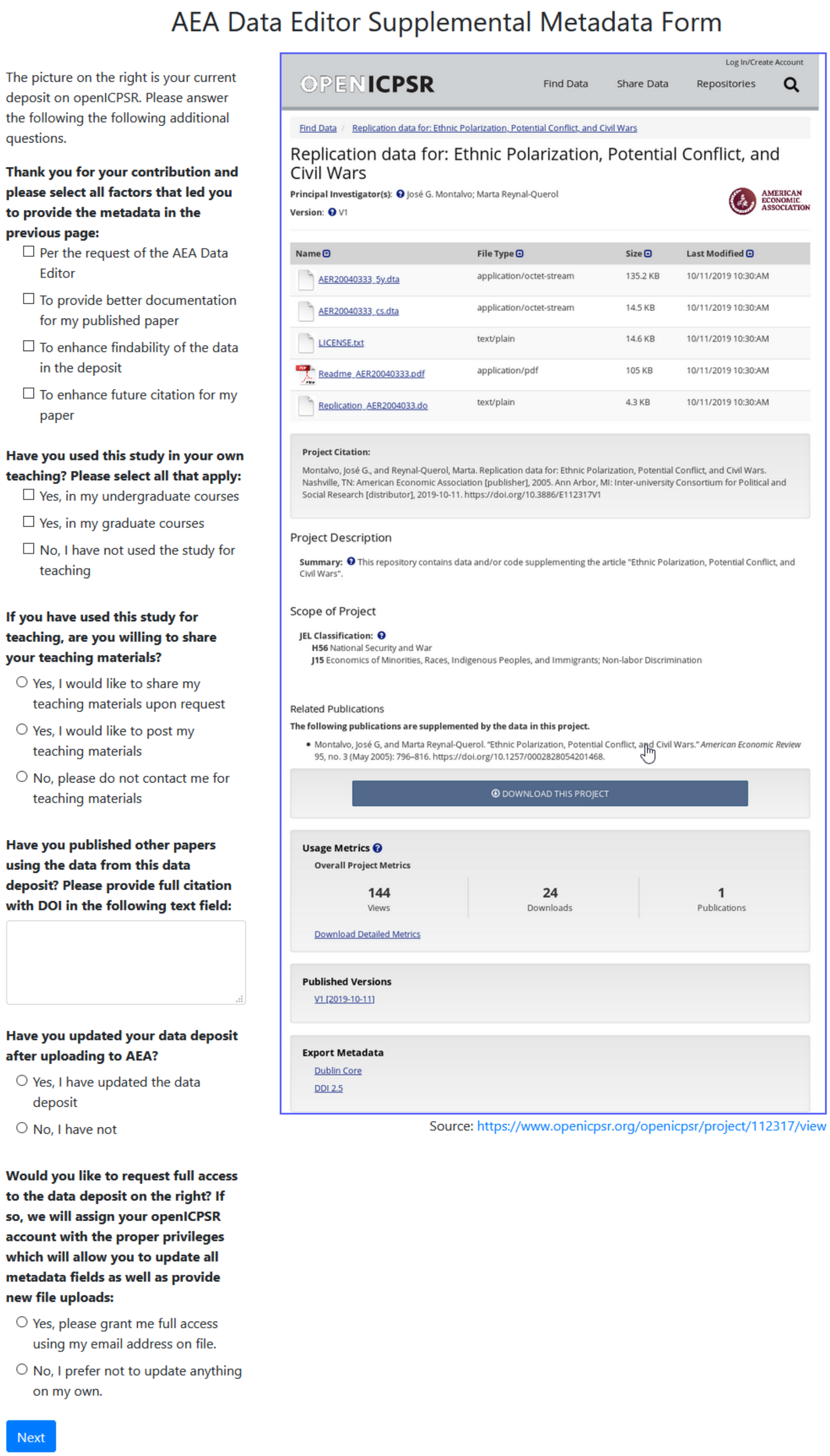


Figure 3.4: Experiment interface - Page 2: Collect metadata contribution



AEA Data Editor Supplemental Metadata Form

The picture on the right is your current deposit on openICPSR. Please answer the following the following additional questions.

Thank you for your contribution and please select all factors that led you to provide the metadata in the previous page:

- ☐ Per the request of the AEA Data Editor
- ☐ To provide better documentation for my published paper
- ☐ To enhance findability of the data in the deposit
- ☐ To enhance future citation for my paper

Have you used this study in your own teaching? Please select all that apply:

- ☐ Yes, in my undergraduate courses
- ☐ Yes, in my graduate courses
- ☐ No, I have not used the study for teaching

If you have used this study for teaching, are you willing to share your teaching materials?

- ○ Yes, I would like to share my teaching materials upon request
- ○ Yes, I would like to post my teaching materials
- ○ No, please do not contact me for teaching materials

Have you published other papers using the data from this data deposit? Please provide full citation with DOI in the following text field:

Have you updated your data deposit after uploading to AEA?

- ○ Yes, I have updated the data deposit
- ○ No, I have not

Would you like to request full access to the data deposit on the right? If so, we will assign your openICPSR account with the proper privileges which will allow you to update all metadata fields as well as provide new file uploads:

- ○ Yes, please grant me full access using my email address on file.
- ○ No, I prefer not to update anything on my own.

Next

Log In/Create Account

OPENICPSR Find Data Share Data Repositories

Find Data / Replication data for: Ethnic Polarization, Potential Conflict, and Civil Wars

Replication data for: Ethnic Polarization, Potential Conflict, and Civil Wars

Principal Investigator(s): José G. Montalvo; Marta Reynal-Querol

Version: V1

AMERICAN ECONOMIC ASSOCIATION

Name	File Type	Size	Last Modified
AER20040333_5y.dta	application/octet-stream	135.2 KB	10/11/2019 10:30:AM
AER20040333_cs.dta	application/octet-stream	14.5 KB	10/11/2019 10:30:AM
LICENSE.txt	text/plain	14.6 KB	10/11/2019 10:30:AM
Readme_AER20040333.pdf	application/pdf	105 KB	10/11/2019 10:30:AM
Replication_AER2004033.do	text/plain	4.3 KB	10/11/2019 10:30:AM

Project Citation:

Montalvo, José G., and Reynal-Querol, Marta. Replication data for: Ethnic Polarization, Potential Conflict, and Civil Wars. Nashville, TN: American Economic Association [publisher], 2005. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2019-10-11. https://doi.org/10.3886/E112317V1

Project Description

Summary: This repository contains data and/or code supplementing the article "Ethnic Polarization, Potential Conflict, and Civil Wars".

Scope of Project

JEL Classification:
H56 National Security and War
J15 Economics of Minorities, Races, Indigenous Peoples, and Immigrants; Non-labor Discrimination

Related Publications

The following publications are supplemented by the data in this project.

- Montalvo, José G, and Marta Reynal-Querol. "Ethnic Polarization, Potential Conflict, and Civil Wars." *American Economic Review* 95, no. 3 (May 2005): 796–816. https://doi.org/10.1257/0002828054201468.

DOWNLOAD THIS PROJECT

Usage Metrics
Overall Project Metrics

144 Views | 24 Downloads | 1 Publications

Download Detailed Metrics

Published Versions
V1 [2019-10-11]

Export Metadata
Dublin Core
DDI 2.5

Source: https://www.openicpsr.org/openicpsr/project/112317/view

Figure 3.5: Experiment interface - Page 3: Collect additional information

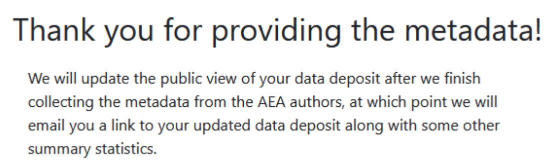


Figure 3.6: Experiment interface - Page 4: Finish page

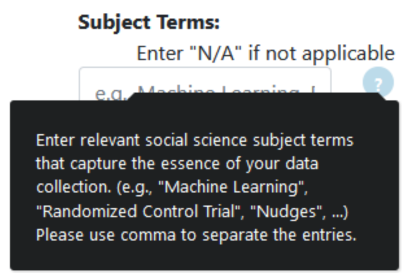


Figure 3.7: Mouse-over clarification text

of the field when activated.

3.A.1.1 Mouse-over Clarification Text

For all the metadata fields that we collect, we provide a mouse-over "help-tip" that clarifies what is expected for the corresponding metadata field. The clarification text is adopted from the formal openICPSR metadata editing interface as well as the AEA Data and Code Guidance document. Please refer to Figure 3.7 for an example of the "help-tip".

3.B Email Template and Data Collection

3.B.1 Email Templates

In Table 3.12, we provide an overview of the email templates used for each experimental condition. In the following email templates, <variable-text> was replaced with the proper value. The blue and underlined text denotes hyperlinked text, which is clickable[8]. The paragraph in italics has been added as an emphasis in this paper. The corresponding paragraphs

Table 3.12: Email Templates for the experiment

Control	Private Benefit	Private + Social Benefit
Subject line: Your AEA data set migration Dear Dr. <Lastname>, Since July 16, 2019, the American Economic Association has used the AEA Data and Code Repository at openICPSR as the default archive for its supplements. The migration increases the findability of your data through a variety of federated search interfaces such as Google Dataset Search, the openICPSR search interface, and the general ICPSR search interface.	Subject line: Your AEA data set migration Dear Dr. <Lastname>, Since July 16, 2019, the American Economic Association has used the AEA Data and Code Repository at openICPSR as the default archive for its supplements. The migration increases the findability of your data through a variety of federated search interfaces such as Google Dataset Search, the openICPSR search interface, and the general ICPSR search interface.	Subject line: Your AEA data set migration Dear Dr. <Lastname>, Since July 16, 2019, the American Economic Association has used the AEA Data and Code Repository at openICPSR as the default archive for its supplements. The migration increases the findability of your data through a variety of federated search interfaces such as Google Dataset Search, the openICPSR search interface, and the general ICPSR search interface.
	Analyses of search and usage of ICPSR's data catalog indicate that most datasets are discovered because searches pick up metadata that includes citation to published articles and key concepts (geography, methods). Enhancing the metadata for your dataset will increase the likelihood that your publication and data are found and cited.	*Analyses of search and usage of ICPSR's data catalog indicate that most datasets are discovered because searches pick up metadata that includes citation to published articles and key concepts (geography, methods). Enhancing the metadata for your dataset will increase the likelihood that your publication and data are found and cited, making it more useful to graduate students and others.*
To further enhance the findability of your data, we ask that you spend up to 20 minutes to provide additional metadata for your AEA data deposit through a user-friendly web interface. The information will be batch-imported back to the original openICPSR deposit.	To further enhance the findability of your data, we ask that you spend up to 20 minutes to provide additional metadata for your AEA data deposit through a user-friendly web interface. The information will be batch-imported back to the original openICPSR deposit.	To further enhance the findability of your data, we ask that you spend up to 20 minutes to provide additional metadata for your AEA data deposit through a user-friendly web interface. The information will be batch-imported back to the original openICPSR deposit.
• If you are interested to proceed, please click here to provide additional metadata for your study titled "<Article 1:Title>." • Please click here if you think your co-authors are better suited for providing metadata. We will opt you out of future communications. • Please click here if you are not interested in providing metadata and would like to opt out of future communications altogether.	• If you are interested to proceed, please click here to provide additional metadata for your study titled "<Article 1:Title>." • Please click here if you think your co-authors are better suited for providing metadata. We will opt you out of future communications. • Please click here if you are not interested in providing metadata and would like to opt out of future communications altogether.	• If you are interested to proceed, please click here to provide additional metadata for your study titled "<Article 1:Title>." • Please click here if you think your co-authors are better suited for providing metadata. We will opt you out of future communications. • Please click here if you are not interested in providing metadata and would like to opt out of future communications altogether.
For articles with more than one author, each co-author is receiving an identical email with an individualized link. Thank you for your effort! Sincerely, Lars Vilhuber AEA Data Editor	For articles with more than one author, each co-author is receiving an identical email with an individualized link. Thank you for your effort! Sincerely, Lars Vilhuber AEA Data Editor	For articles with more than one author, each co-author is receiving an identical email with an individualized link. Thank you for your effort! Sincerely, Lars Vilhuber AEA Data Editor

* Note, additional vertical spaces are inserted here to assist comparison across experimental conditions.

were not italicized in the email message we sent.

In the Control condition (C), we first provide the background information that data deposits previously hosted by AEA are now being migrated to openICPSR. After this paragraph, the authors are invited to provide metadata for the data deposit associated with the selected AEA publication through our experiment interface.

Compared to the control message, for the private benefit condition (T_p), we add a new paragraph and introduce evidence that explains and supports the "findability" implication of enhanced metadata. This addition reduces the uncertainty of the future value of the metadata. The paragraph concludes with an emphasis on the private benefits of enhanced metadata.

In the private and social benefit condition (T_{p+s}), in addition to mentioning the private benefit introduced through enhanced finability, we add the social benefit of enhanced metadata for graduate students and others.

At the end of the email, we provide three links for authors to choose from; they can opt to do one of the following: (1) provide the metadata for their study; (2) let their coauthor(s) provide the metadata; or (3) opt out.

In the last paragraph, we tell each author that their coauthor(s) will receive an identical message. This reduces the likelihood that the authors will communicate with each other before responding to the message.

In Figure 3.8, we provide a few sample screenshots taken on two mobile phones. When our

[8] There are two types of links. In the first bullet point, the link to provide additional metadata is populated with an author-specific URL that takes the author to the experiment interface introduced in Section 3.A. The remaining two links are populated using MailChimp's survey tools provide a binary outcome variable.

treatment email is scaled down to fit the small screen, not all treatment texts and clickable items are immediately visible.

3.B.2 Experiment Procedure

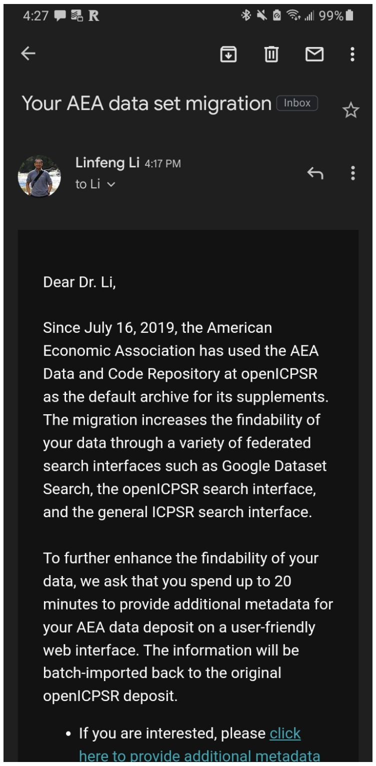
In this experiment, each author receives a customized email using one of the three templates. The links in the emails for the customized web interface are individual specific; the article we selected is displayed as the authors fill in the missing metadata. For screenshots of the experiment interface, please refer to Appendix 3.A.

Outcome variables We collect outcome variables through both the treatment email and the survey interface. First, we record the intention to contribute through the hyperlinks in the body of the treatment email. A positive response is recorded when an author clicks on the link to contribute metadata. A negative response is recorded when an author chooses either "coauthors are better suited for providing metadata" or "not interested in providing metadata." We use MailChimp to administered the email delivery and tracking. Per our testing, the email-opening and link-clicking events are tracked even when our email reaches an old email address and gets forwarded to another email address belonging to the same researcher. For authors who proceed to the experiment interface, we collect individual responses for metadata fields and aggregate the responses at the study level for analysis. Appendix 3.A.1 contains the complete list of metadata fields we collect.

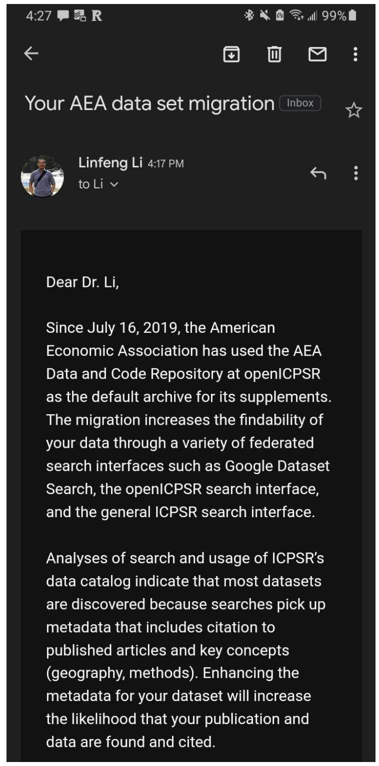
After collecting the metadata contents in the customized interface, the authors answer a few survey questions about their data reuse activities and their willingness to update the data deposit. For those interested in further updating the data deposits on openICPSR and providing additional metadata, we grant them write access to their data deposits directly on openICPSR. With write-access, authors can upload new files to the deposit and provide additional metadata. Last, we ask the authors about their motivation for providing the metadata.

Accommodation for sequential inputs For each article with multiple authors, our interface is designed to accommodate sequential inputs of metadata by coauthors. In the event that multiple coauthors contribute metadata sequentially, a new contributor can see who has contributed what and add content accordingly. In the unlikely event that multiple coauthors edit at the same time, our interface is not able to let them see the location of each other's cursors.

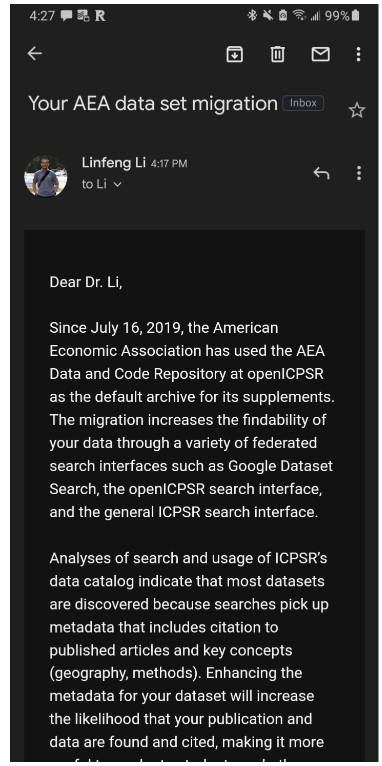
As a result of sequential inputs, when accounting for individual contributions to a given metadata field, we consider both the overall contribution of all coauthors, as well as a binary variable denoting *individual improvement*. The latest edit is counted as the overall contribution.



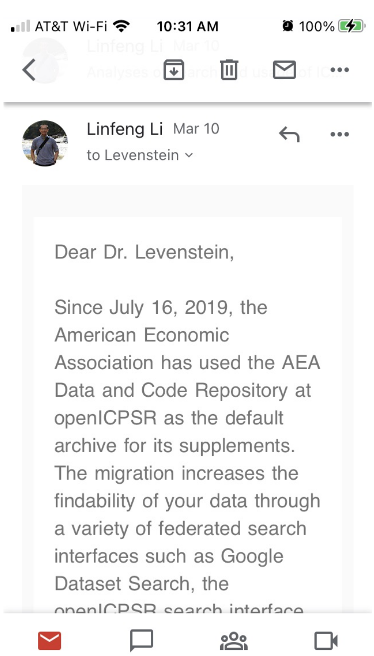
(a) Control condition



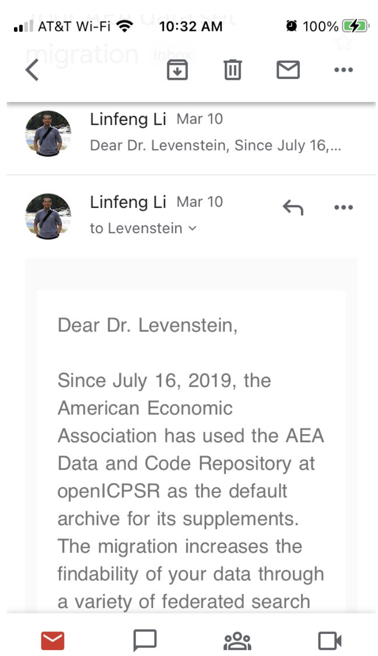
(b) Private benefit condition



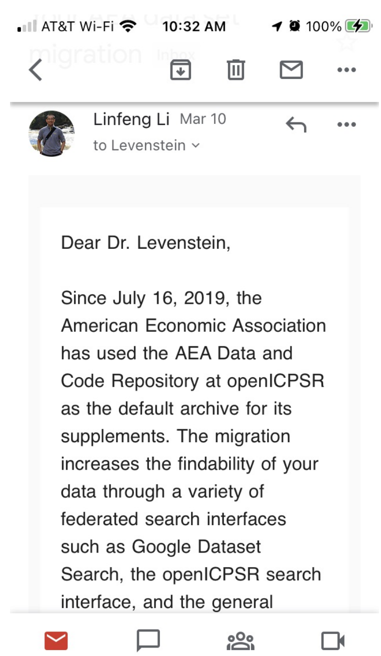
(c) Private+Social benefit condition



(d) Control condition



(e) Private benefit condition



(f) Private+Social benefit condition

Figure 3.8: Contribution rate by metadata fields

To capture the individual improvement for a given metadata field, we compare the contributions provided by authors who edited consecutively. If there is a difference, we document it as an improvement. Otherwise, we set the improvement to zero.

In accordance with our registration on the AEA RCT Registry (Chen et al., 2020c), we launched the experiment on Monday, August 3, 2020, and finished the data collection process on September 30, 2020.

3.B.3 Random Assignment and Balancedness Check

Experimental Conditions:	Control	T_p	T_{p+s}
Number of articles	487	487	486
Number of authors	1007	1013	1003

Table 3.13: Number of Observations after Random Assignment

We randomly assign experimental conditions at the article level and block level according to three basic characteristics of the article: the network position of the article[9], the number of authors in the article, and the articles year of publication. Based on our selection algorithm, articles are included from simple components or randomly chosen from a complex component in the reduced graph. Furthermore, simple components can be either naturally occurring or generated through the network-trimming exercise. In practice, we sort the list of articles sequentially by (network position, number of authors, year of publication) and reshuffle the ordering within each group. To complete random assignment, we enumerate through the sorted and reshuffled list and iteratively assign C, T_p, T_{p+s} for all articles we include in the experiment. For more detailed random assignment procedures, please refer to Appendix 3.D.

3.C Data Preparation

In this section, we describe the sources we used to prepare for the experiment. First, we collect identifying information about each data deposit and subsequently about all authors associated with the publication. Coupled with the steps we took to address potential ambiguous author names, we then document the steps we took to collect the contact emails for all authors invited in this study.

[9]As detailed in Appendix 3.D, we draw articles from "independent" components in the network. The random assignment is summarized in Table 3.13. The network position of an article is defined by its generating component, which can be a simple natural component, a simple component that belongs to a complex natural component or a complex component in the trimmed graph.

3.C.1 Collecting Metadata for the Migrated Data Deposits

For this experiment, we consider all data deposits in the AEA Data and Code Repository that have been migrated through the AEA Repository migration[10]. Since the official migration record on openICPSR is incomplete[11], we first scrape the search page for all AEA deposits on the openICPSR website and collect the URL for each deposit. From the individual openICPSR webpage for each data deposit, we are able to extract the full set of authors together with the AEA_DOI and openICPSR_ID. Migrated studies are identified by their release date and are included if the release date falls within the following set of dates: {2019-10-11, 2019-10-12, 2019-10-13, 2019-12-06, 2019-12-07}.

We generate the following set of records from various sources using the master list of all migrated data deposits.

1. From the openICPSR website, we collect all publicly available information on the study-page, including the following:
   - All existing metadata fields;
   - openICPSR_DOI for the data deposit; and
   - AEA_DOI for the original publication.
2. From Crossref, we query using AEA_DOI and obtain the following:
   - Structured names that are parsed into multiple fields (given name, family name, suffix ...); and
   - Institutional affiliation (parsed originally from the footnote field according to Crossref documentation);
3. From raw PDF versions of the paper, we look for texts on the first page in the footnote area for the following fields:
   - Last name;
   - Institutional affiliation; and
   - Email address (usually in parentheses).

Auxiliary records include emails for corresponding authors from the publisher, which does not cover the full population of participants and contains outdated emails for those who have since changed their institution affiliation.

[10] `https://aeadataeditor.github.io/aea-supplement-migration/programs/aea201910-migration.html`

[11] Updated on 2020-02-24, the deposit titled "Data files for AEA Repository migration" on openICPSR has 2,562 studies, while the total number of migrated studies is 3,073.

3.C.2 Disambiguation of Author Names

Despite having access to both the official AEA publication records and the deposited list of data deposits on openICPSR and Crossref, it remains a laborious task to disambiguate the author names. For example, two authors who have the exact same name in two publication records may be distinct individuals who happen to share the same name. One motivation for this exercise is that we would like to administer treatment to the author who is actually involved with the data deposit to which we invite him/her to contribute.

We engage in two iterations to generate a minimal set of "SubjectIDs" for the participants in our experiment. We use records gathered in Iteration 1 as a benchmark and results from Iteration 2 for production. We are able to identify 4,320 unique authors from a total of 4,503 different names.

3.C.2.1 The Origin of Duplicates

The origin of duplicate names is clear. The source page on the AEA website may contain variants of the name for the same economist. For example, see Figure 3.9, where "John M. Roberts" and "John Roberts" were both listed. The variants of such names were then used to propagate records on openICPSR and Crossref. intended meaning.

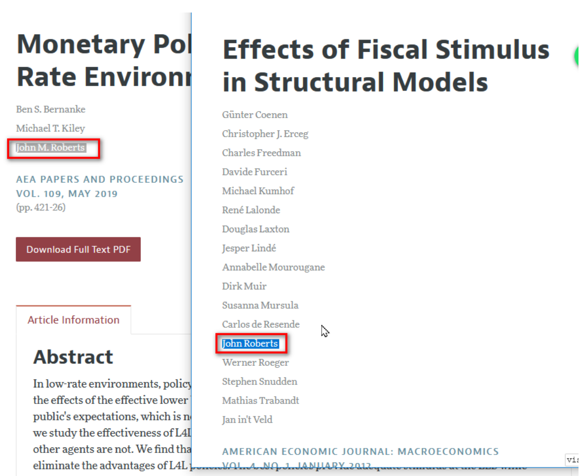


Figure 3.9: Source of the Closed Duplicates from Original Records

3.C.2.2 Iteration 1: Disambiguation Based on String Values

Since all the authors' names are drawn from the official publication record, we assume that the first and last names are spelled correctly[12]. One common approach for two authors who share the same first and last names to distinguish themselves is to include their middle name as well. However, since middle names were not consistently spelled out in full in multiple publications by the same author, we match similarly spelled author names with their respective institutional affiliation to consolidate the identity of these same authors.

In our disambiguation practices, we encounter the following edge cases:

- Foreign names with special characters are mistaken at times under different character encoding, where, for example, "ş" and "s" are used interchangeably;
- Authors with misspelled first names that have been "shortened"; and
- Authors postfixing/extending their last name with "Jr. ", for example.

In practice, we consider all the cases mentioned above and assign all the occurrences of similar name spellings with a unique SubjectID. This attempt reduced the number of unique authors from 4,503 to 4,409.

3.C.2.3 Iteration 2: Name Matching Using openICPSR & Crossref Records

This iteration leverages the fact that our publication record contains a list of digital object identifiers (DOIs) for the published papers. In particular, we assign a SubjectID to records of author-name + DOI, and we resolve inconsistencies along the way. In total, from the 3,070 publications, there are 7,251 author-name + DOI records. At the highest level, we enumerate through the set of names and assign an ID to each unique author-name + DOI record. Practically, we attempt four independent ID assignment rounds and perform two rounds of consolidation of the SubjectID field.

Sources of SubjectID labeling First, for both openICPSR records and Crossref records, we assign two sets of SubjectIDs, as detailed below.

1. For names in openICPSR records, we assign a unique ID to each unique combination of the "firstname" and "lastname" fields (these fields do not exist in the raw dataset and are

[12]To the best of our knowledge, among the 3,070 publications we consider in our experiment, only one author has her last name spelled wrong in the original AEA record. This author has three publications in total; we assign her a unique ID as we did with all other close variations of names that represent the same author.

parsed using simple rules[13]);

2. For names in the openICPSR records, we assign a unique ID to each unique spelling of the full name, and we unify the IDs belonging to the set of known "pairs of name-variants" that refer to the same individual. (This list is composed manually in Attempt 2);

3. For names in Crossref records, we assign unique ID to each unique combination of the "firstname" and "lastname" fields (these fields are provided by Crossref per a DOI query, yet the middle name may enter either the "firstname" or the "lastname" field);

4. For names in the Crossref records, we manually assign unique IDs based on first name, last name and institutional affiliation.

Toward the end of these steps, we obtain ID-V1 from openICPSR and ID-V2 from Crossref.

Consolidation of SubjectID For the consolidation of the SubjectID field, we adopt the following steps. We first identify a set of observations with a buggy SubjectID from *a pair of SubjectID* and perform the fix by following the steps below:

1. For each set of SubjectIDs (ID-V1 and ID-V2), we use the associated DOI field to compose the publication list under the "name" of the SubjectID.

2. For each observation, if the publication list under ID-V1 is different from the publication list under ID-V2, we extract both records for further inspection.

3. We manually inspect the discrepancies between ID-V1 and ID-V2, and "fix" the wrongly labeled SubjectID when appropriate[14];

4. Upon performing the fix (relabeling of the ID) for both ID-V1 and ID-V2, we end up with a one-to-one mapping between ID-V1 and ID-V2, where the lists of publications are identical regardless of how they are extracted.

In practice, we perform two rounds of consolidation exercises. First, we consolidate two independently generated openICPSR records by themselves. This process is repeated for two independently generated Crossref records, which yields openICPSR_ID and Crossref_ID. In the

[13]From the raw_full_name, we parse by white spaces and keep the first word as the first name; then, from the rest of the "bag of words", we take the longest as the last name. Of course, this approach is not perfect, as "FFF M LLLL" and "FFF LLLL" will be assigned two distinct SubjectIDs, for example.

[14]For example, ID-V1 may assign different SubjectIDs for FFF LLLL and FFF M LLLL, while ID-V2 may assign identical IDs to these two records. In this case, we fix ID-V1 by keeping one ID. ID-V2 can be fixed in an identical manner. In rare cases, we need to fix both ID fields.

second round, we consolidate the two IDs by first merging the records on (fuzzy) author-name & DOI [15] and then performing the consolidation steps detailed above.

Results At the end of the process, we obtain 4,320 unique SubjectIDs. This is an improvement over the total of 4,503 unique spellings of names in the original record.

3.C.2.4 Exceptions

In our disambiguation process, we uncover a list of distinct authors with similar names. These pairs are generated throughout multiple rounds of iteratively assigning and correcting the `SubjectID` field, as mentioned in Appendix 3.C.2.3. Here are 6 pairs of names that we have identified that belong to different individuals:

1. Michael P. Devereux and Michael B Devereux
2. Robert Gordon and Robert J. Gordon
3. Benjamin B Lockwood and Ben Lockwood
4. Benjamin M. Marx and Benjamin Marx
5. A. Banerji and Abhijit Banerjee
6. Michal Bauer and Michael D. Bauer

3.C.3 Email Collection Notes

We have combined multiple sources to ensure that our email database is up to date. In early March 2020, we scraped all available studies deposited in AEA deposits and conducted a Google search for all the authors that we found. We obtained a total of 4,503 names from the web-scraping exercise. From AEA, we obtained a separate set of contact emails for the corresponding authors in the 5 major publication outlets of AEA, which were updated in 2018. To generate a comprehensive set of contact emails for all authors whose AEA publications have been migrated to openICPSR AEA Data Deposit, we adopt the following data cleaning strategies.

[15] Merging by strings of names in both records can only cover 3,069 records, while we have 7,251 in total. Thus, we pair the openICPSR record with the Crossref record by merging each name in the openICPSR record with the Crossref author-name in the same DOI that matches the openICPSR name the best.

1. For those emails that coincide between the web scrapes and the contact list of corresponding authors, we mark these emails as "up to date". Furthermore, we ask a research assistant (RA) to determine the authors Gender and Year of PhD using the credible sources where the valid emails were scraped from. This is a relatively fast task at the rate of 30 records an hour for a total of 1,023 records.

2. For the rest of the emails, we employ two RAs to manually verify the emails and generate additional labels for the following:

    - Gender
    - Year of PhD
    - Primary website

    Since our RAs need to verify the emails by cross-referencing various sources (CV $\succ$ Personal website $\succ$ Departmental/Workplace/Organization Website $\succ$ NBER website), it takes a considerable amount of time to verify each record. In total, it took 211 hours to finish 3,696 rows (17 rows/hour). Two RAs worked full-time on this task in April 2020.

All emails and demographic information are collected on web pages with public access. Due to the total volume of records, each email record is verified by only one RA. Without the benefit of double entry, approximately 5% of the emails we previously collected were found to be no longer up to date by the time that we planned to launch the study[16]. To ensure that our contact emails are valid for the participants that we include in the experiment, we conduct another round of email validation for the 3,023 participants that we choose to include based on the random assignment procedure. The production rate is 50 rows/hour, and we finish this final round of email validation a week before launching the experiment.

3.D Random Assignment in the Network

In this section, we document the set of "network-trimming" routines we adopt to generate the set of *independent* data deposits we include in the experiment.

[16]In preparation for launch, we perform a spot check of email qualities by revisiting the primary website and verifying our email record against the listed contact emails on an authors CV. If no CV is found, we use the contact email provided within the latest working paper by the author. Among a random sample of 1,124 author records checked in early August 2020, 54 needed an update, which amounts to a total of 4.8%. Among those emails that needed to be updated, less than 1% were due to human error (attaching a completely wrong email) and the remaining majority was due to a change of institutional affiliations.

3.D.1 The Coauthorship Network

We build the network of authors based on coauthorship in the papers with data and code supplements, where all authors are uniquely identified by the SubjectID produced from the disambiguation exercise (see Appendix 3.C.2). All data deposits are identified through the DOI of the original paper. Technically, we build a multigraph that permits multiple edges between two nodes, which also allows for self-loops. Respectively, this representation of the network keeps the record of multiple coauthored papers between two authors and allows for multiple single-authored papers.

To build the network, we scrape the AEA Data and Code Repository hosted on openICPSR and obtain a complete list of authors for each study in the AEA Data Deposit. We consider only those studies that were *migrated* in a series of data dumps[17]. Table 3.14 offers a demonstration of the raw datafile.

Article ID	Author 1	Author 2	Author 3
$\mathbb{A}_1$	A_1	A_2	A_4
$\mathbb{A}_2$	A_2	A_4	
$\mathbb{A}_3$	A_1	A_3	
$\mathbb{A}_4$	A_5		

Table 3.14: Original Format of the Scraped Data

Given our raw data structure, we build such a multigraph by enumerating the list of all articles and creating an edge with the `ArticleID` between any possible combination of two authors that belongs to the same article. In essence, we transform Table 3.14 into the edge list shown in Table 3.15. Through the multigraph built through the edge list, we are able to track

ArticleID	Author1	Author 2	Note
$\mathbb{A}_1$	A_1	A_2	
$\mathbb{A}_1$	A_1	A_4	
$\mathbb{A}_1$	A_2	A_4	parallel edge
$\mathbb{A}_2$	A_2	A_4	parallel edge
$\mathbb{A}_3$	A_1	A_3	
$\mathbb{A}_4$	A_5	A_5	Self-loop

Table 3.15: Edge list view

how an edge is introduced into the graph through the article labeling of the edges.

[17]The migration had two waves; the first wave took three days (Oct 11 - Oct 13, 2019), and the second wave took two days (Dec 7 - Dec 8, 2019).

3.D.1.1 Motivation for the Network-trimming Exercise

One major threat for measuring treatment effects in a network setting is "spillover effect", where the treatment received by one author may spill over to another author. Worded more broadly as "interference", the concern is that the outcome variable is influenced by not only the treatment but also other treated participants (Cox, 1958).

Here, we provide one way to cut off the channels for potential "spillovers" by isolating each treated data deposit to be a connected *component*[18]in a trimmed network.

Throughout this section, we assume the unit of randomization is at the deposit (article) level. Nevertheless, there are fairly few other feasible randomization units at our disposal. In the rest of this section, we introduce how we perform the trimming exercise.

3.D.1.2 Community-detection Algorithms

For a given connected graph, community-detection algorithms generate the "best partition" that "maximizes" the *modularity* measure. In principle, these algorithms identify the communities composed of closely connected individuals. In our setting, we extract the unit of randomization within an element of the "best partition" and establish independence among all selected units.

Let g_i and g_j denote the "group" to which i and j belong, respectively, with $g_i \in 1, \cdots, N$, where N is the total number of "groups" or communities/partitions. Then, Q, i.e., the *modularity*, measures "the extent to which like is connected to like in a network" (Newman, 2017).

$$Q = \frac{1}{2}\sum_{ij} A_{ij}\delta_{g_i g_j} - \frac{1}{2}\sum_{ij}\frac{k_i k_j}{2m}\delta_{g_i g_j} = \frac{1}{2m}\sum_{ij}\left[A_{ij} - \frac{k_i k_j}{2m}\right]\delta_{g_i,g_j}$$

where A_{ij} is the adjacency matrix for the graph, δ denotes the Kroncker delta[19], k_i, k_j denotes the degree for node i and j, and m is the total number of edges.

In practice, we choose from two well-adopted community-detection algorithms: The Louvain algorithm (Blondel et al., 2008) and the Leiden algorithm (Traag, Waltman, and van Eck, 2019). The Louvain algorithm identifies a partition of the graph that maximizes the *modularity*. As reported in Table 1 in Blondel et al. (2008), the Louvain algorithm achieves a higher modularity score than and outperforms standard algorithms Newman and Girvan (2004), Pons and Latapy (2006) and Wakita and Tsurumi (2007). The Leiden algorithm, as introduced

[18]In our settings, a connected component is a subgraph in which any two vertices are connected to each other by at least one edge, which is connected to no additional vertices in the supergraph.

[19]Kroncker delta is defined as $\delta_{g_i,g_j} = \begin{cases} 0, & \text{if } g_i \neq g_j \\ 1, & \text{if } g_i = g_j \end{cases}$

in Traag, Waltman, and van Eck (2019), prevents the further partitioning of well-connected components that can be taken apart under the Louvain algorithm. In Section 3.D.3, we compare the performance of the Louvain algorithm against that of the Leiden algorithm.

3.D.2 Composition of the Data-deposit Network of AEA Authors

With the full network, we have more than 887 *components*, where the giant component contains more than half of all the authors ($2,008$ out of $4,320$). We also have a total of 666 *simple components* that are composed of a set of authors who have published one and only one article in AEA outlets.

3.D.2.1 Simple vs Complex Components

For a given component, we enumerate through all its edges and account for the number of unique "data deposits" that introduced the edges. Then,

- If all the edges in a component are introduced by a unique data deposit, we call it a simple component; and
- If the edges in a component are introduced by a collection of data deposits, we call it a complex component.

Note that this dichotomy is defined by assuming a certain underlying network to begin with. In our "trimming exercise" that follows, we update this underlying network by *dropping* certain edges.

For our experiment, we aim to include all *simple components* in a "trimmed graph". For *complex components* with multiple articles, we choose one article at random to include in the experiment. The following section details the operation relevant to constructing such a "trimmed graph".

3.D.2.2 Partition the Network and Trim its Edges

A partition of a network is a partition over the set of nodes in a network. A typical use of the partitioning algorithm is to identify "communities" in a network, where the algorithm generates the partition and researchers try to make sense of the composition of the components. For practical purposes, we use the out-of-the-box Louvain algorithm to generate the partition.

Notes on getting "better" partitions from the community-detection algorithm For any given undirected network, the Louvain algorithm can generate a partition based on a seeded random search; the solution tries to maximize the modularity of the partition for a given graph, and this is an NP-hard problem. Although we cannot obtain the optimal solution, we can still improve the "quality" of the generated partition by choosing which subgraph to feed to the Louvain algorithm. It is noteworthy that we started the investigation with the Louvain algorithm. All comparisons of performance hold true for the Leiden algorithm, which we eventually choose for production.

Given that our original network has 887 independent components, we perform the following two tests for an arbitrary component:

- We generate a partition of the graph using the full graph and extract the partitions that belong to the component that we picked;
- We take the subgraph of the component and generate a partition for the subgraph.
- We compute the modularity for the subgraph using the following two partitions: M_{full} is obtained by partitioning the full graph, and M_{sub} is obtained by partitioning the subgraph.

We repeat the computation 1,000 times, with a randomly generated "seed" that guarantees reproducibility. We confirm that the Louvain algorithm will perform better with the subgraph, with $M_{full} < M_{sub}$ across all the trials.

Which deposit to keep and which deposit to drop? From a given partition of a component, we construct a "quotient graph", whose vertices denote the communities identified by the partition. Then, by construction, the edges in the quotient graph are those that are "bridging" communities in the network, thus termed intercommunity edges. Removing these intercommunity edges from the original component generates several connected components from the previously connected subgraph. Last, with the intercommunity edges removed, the remaining edges from the deposits that generated intercommunity edges no longer provide much help for our randomization scheme. To conclude, we drop all edges introduced by the deposits that generated intercommunity edges, and we keep all the deposits that were nested within the communities.

3.D.2.3 Iteratively Identify and Remove the Intercommunity Edges

In practice, we adhere to the *best practice* [20]for applying the community-detection algorithm and iteratively build a "drop list" until the community detection algorithm fails. In Algorithm 1, we provide the pseudo code for the operation.

Require: G is the full graph and L_{drop} is a set of deposits to drop

▷ Main Function

function TRIM_GEN_DROP_LIST(G, L_{drop})

  trimmed_graph $=G$ without any edges introduced by the deposits in L_{drop}

  complex_component = complex components in **trimmed_graph**

  ▷ Iterate through all complex components in the trimmed graph

  for $component$ in complex_component **do**

    Generate partitions using a ***community-detection algorithm***

    Generate quotient graph using the subgraph of the $component$ and the partition

    l_{drop} = deposits that introduced the **intercommunity edges**

    $L_{drop} = L_{drop} \cup l_{drop}$

  end for

  return L_{drop}

end function

▷ Initialize the initial list of dropped deposits with no edge removed

L_{drop} = TRIM_GEN_DROP_LIST(G, $\emptyset$)

▷ Iterate until L_{drop} is stable

while TRIM_GEN_DROP_LIST$(G, L_{drop}) \setminus L_{drop} \neq \emptyset$ **do**

  $L_{drop} =$ TRIM_GEN_DROP_LIST(G, L_{drop})

end while

In summary, for a given network, we first classify the (connected) components into simple vs complex components. For each complex component, we attempt to apply the community-detection algorithm and collect a "drop list" from articles that introduced the inter-community edges in the quotient graph. We repeat the algorithm over rounds of edge-trimming, until all complex components in the reduced graph shall withstand the community-detection algorithm of choice. We will discuss our choice of community detection algorithm in Section 3.D.3.

[20]Note that community detection is an NP-hard problem, and all we have are approximation algorithms that are sensitive to the initial input. One would intuitively predict that the modularity score of a partition for a given component is higher when the partitioning algorithm is only told about the precise component. This is explained in full in Section 3.D.2.2

3.D.3 Comparison of Community-detection Algorithms

The core step in Algorithm 1, as highlighted, is to partition a given complex component using a community detection algorithm. We consider two candidates: the Leiden algorithm (Traag, Waltman, and van Eck, 2019) and the Louvain algorithm (Blondel et al., 2008). In this section, we summarize the numerical exercises we conduct to decide which community-detection algorithm to employ.

Metrics	Louvain	Leiden
Part 1: Review of dropped DOI and remaining components		
UniqueDOI_Dropped	1075.63 (0.21)	929.76 (0.17)
AuthorsDropped	1003.66 (0.34)	923.24 (0.35)
# Simple Compo	1208.20 (0.14)	1058.07 (0.15)
# Complex Compo	313.85 (0.10)	399.60 (0.07)
sum(Simple, Complex)	1522.05 (0.09)	1457.67 (0.10)
Part 2: Max and random set of included participants		
Max participants*	3125.35 (0.31)	3150.52 (0.37)
Rand participants*	3072.25 (0.42)	3011.80 (0.63)

Table 3.16: Compare Trimming Output from 500 Distinctive Random Seeds
Mean value is reported, with standard error in parenthesis.
* Max participants is collected from articles with the largest coauthor count in each complex component.
Rand participants are collected from a randomly chosen article in each complex component.

For production, we choose to use the Leiden algorithm for multiple reasons. Overall, fewer articles (DOIs) were dropped during the trimming exercise with the Leiden algorithm, and fewer authors were completely dropped from the experiment. In terms of total units (components) for random assignment, the Leiden algorithm gave a comparable result. Aside from what is shown through the comparison in Table 3.16, when applied to individual components, the Leiden algorithm outperforms the Louvain algorithm in 155 out of 222 complex components in terms of the modularity of the derived partition of the components. Furthermore, there are 2,653 authors in the 155 components compared to the 395 authors in the remaining 67 components where the Louvain algorithm gave a "slightly better partition". To conclude, we employ the Leiden algorithm for our network-trimming exercise.

3.D.4 Random Assignment Procedures

Since the Leiden algorithm is an approximation method, we run the proposed network-trimming algorithm 1,000 times with distinct initial seeds to look for the realization where we drop the fewest number of authors from the trimming exercise[21]. After obtaining the reduced graph without intercommunity edges, we follow the procedure below to choose who to include in the experiment and randomly assign authors into treatment conditions.

1. Choose articles from components in the reduced graph so that each component has only one article chosen; this is trivial for simple components where only one article is involved. For complex components that remain after iterations of trim-and-drop, they are well-connected; thus, we pick the article with *most* number of authors. For the complex components that remain, we randomly pick an article from each component and include all authors in the experiment.

2. We assign articles/components into experimental conditions, where we block by the network position of the components and by the number of authors in the chosen articles.

3. Last, since a very small proportion of authors have missing emails or are deceased, we remove them *after* the random assignment procedure.

We end up with a trimmed graph of $1,460$ components and $3,023$ authors. In our population of authors, 11 authors are deceased and 6 have missing email addresses despite our best efforts to find them. We contact all the rest of the authors with valid email addresses. In total, there are $1,459$ articles with at least one author who has a valid email address, according to our records.

[21] According to our network-trimming algorithm, we drop all articles that introduced intercommunity edges. For a given author, he/she is dropped completely if all his/her articles are dropped. Among the 1,000 repetitions, the mean number of authors dropped is 922.665, with a standard error of 0.253. The realization with the fewest dropped authors drops 893 authors.

CHAPTER 4

Audit Study in a Digital Era: Gig-work Experience Does Not Harm

4.1 Introduction

As of February 2019, the U.S. Bureau of Labor Statistics reported a total of 6.2 million unemployed individuals in the U.S., an unemployment rate of 3.8% (Labor Statistics, 2019a). Approximately 1.3 million of these individuals had been unemployed for more than 6 months and 3 weeks (i.e., 27 weeks), or were experiencing long-term unemployment. Those facing long-term unemployment are often perceived as having fewer social and intellectual skills, being less trainable, and being less up-to-date with technological changes (Van Belle et al., 2017). As a result, some economists theorize that those with long career gaps experience intrinsic bias and often face discrimination from employers (Bahler, 2017). In addition, new automated applicant tracking systems that sort through high numbers of resumes can negatively impact the people with long-term unemployment (Bahler, 2017). Research consistently finds that those who face long-term unemployment are more than twice as likely to have left the market altogether than to have settled into stable, full-time work (Krueger, Cramer, and Cho, 2014).

The rise of the digital on-demand economy, or the gig economy, provides nontraditional and contingent employment opportunities that are akin to independent contract work. The study of such platforms in HCI and CSCW as well as the use of technology for employment and entrepreneurship has been steadily increasing (e.g., Ahmed et al., 2016; Dillahunt et al., 2017; Dillahunt et al., 2018; Dillahunt et al., 2016a; Dillahunt and Malone, 2015; Dombrowski, Alvarado Garcia, and Despard, 2017; Hui et al., 2018; Hui, Gergle, and Gerber, 2018; Salehi and Bernstein, 2018; Suzuki et al., 2016). On-demand work has been said to aid in supporting household incomes and job growth, which benefit both workers and employers. These jobs also offer little to no training costs, low cost of entry, and flexible hours Dokko, Mumford, and

Schanzenbach, 2015. The gig economy is particularly attractive for people who value the flexibility often unavailable in traditional jobs (Chen et al., 2017).

Broadly, we would like to investigate whether gig work helps to mitigate the negative effects of long-term unemployment for low-skilled job seekers with employment gaps. However, given the variety of types of gig work available and geographic locations, we focused our study on a specific type of gig work—driving. We began our investigation with driving because real-time ridesharing platforms like Lyft and Uber have transformed the essence of traditional workspaces (Yaraghi and Ravi, 2017). With fewer requirements to work for the ridesharing platforms, the unemployed job seekers can acquire such work experience more easily compared with overcoming the entry barriers to traditional driving jobs (Cook et al., 2018). On the demand side, based on our preliminary work, driving jobs are among the top three most frequent job listings categories across major metropolitan areas.

We asked the research question: Does driving for Uber (a form of non-traditional work) help low-skilled job seekers [1] fill resume gaps? In other words, does this form of work lead to more, fewer, or as many callbacks, as low-skilled job seeker resumes with employment gaps? To answer this question, we conducted a field experiment to uncover the differences in employer responses to understand whether this form of gig work can be used to lessen the negative effects of long-term unemployment for low-skilled job seekers with employment gaps. We hypothesized that on-demand driving gigs could help these job seekers fill in their employment gaps, which would lead to more callbacks than those with unfilled gaps. We also speculated that the hypothetical skills gained from performing these gigs could lead to more callbacks than those without these skills listed. Drawing on the economics, sociology, and HCI literature, we contribute:

- Methodological insights from the use of an audit study to identify the effect of resume content on initial employer interest;

- Our early findings, which suggest that driving for a real-time ridesharing service does not substitute for traditional driving jobs in bridging employment gaps;

- Results showing unequal callback rates between men and women; and,

- A call to CSCW to investigate methods that help to understand *why* real-time ridesharing services do not substitute for traditional jobs in bridging employment gaps, and solutions on how to overcome it.

[1] We define *low-skilled job seekers* as job seekers who don't have a high level of education. In our study, all fictitious applicants are high school graduates.

4.2 Related Work

The effect of temporary work on long-term employment has often been discussed in economics and sociology literature (e.g., Autor and Houseman, 2005; Gebel, 2013; Ichino, Mealli, and Nannicini, 2005). However, most of this literature was written at a time when the sharing economy and the concept of gig work were not prevalent. In this section, we first discuss the literature pertaining to the effects of temporary work on long-term employment pre-dating the rise of the sharing economy. We then provide an overview of more recent HCI and CSCW literature surrounding employment in the sharing economy.

4.2.1 The effects of Temporary Employment on Long-term Employment

There are discrepancies related to the effects of temporary employment on long-term employment. Economics and sociology literature suggest that temporary employment is associated with disadvantages when compared to permanent employment (Gebel, 2013). Yet temporary work provides job seekers with an opportunity to acquire human capital, expand contacts with potential employers, possibly transition to more stable employment, and increase employment earnings (Autor and Houseman, 2005). Understanding the effect of temporary employment for those who are unemployed is also an opportunity for future research(Gebel, 2013).

To address this opportunity, Gebel (2013) compared the potential integrative power of working a temporary job for unemployed workers to the counterfactual situation of searching for another job while remaining unemployed. He found that working a job temporarily increases the employment chances during the following 5 years; however, these results were limited to Germany and the United Kingdom. This research confirmed findings in the Italian context (Ichino, Mealli, and Nannicini, 2005); however, neither long-run advantages nor disadvantages of working a temporary job were found in the flexible Swiss labor market, which demonstrate region-to-region variation.

Additional research aimed to resolve the differences between U.S. and European data because the majority of these studies (1) made use of European data, (2) used non-experimental data, and (3) assumed that temporary job selection was driven by observable characteristics up to a random factor (Ichino, Mealli, and Nannicini, 2005). At the time of this work, only one study showed that temporary work had a negative effect on employment outcomes (Autor and Houseman, 2005), and these findings were based on U.S. data. Ichino et al. concluded that the cost of firing employees was cheaper in the U.S. than in all European cities where the effect of temporary work on employment had been evaluated (Ichino, Mealli, and Nannicini, 2005). If firing costs are higher, then employers place greater importance on worker quality before hiring, which could explain the reason for positive outcomes in countries with higher firing costs than in the U.S. The authors concluded that the effects of temporary employment will vary in countries with different employment protection regimes. Another investigation to understand whether temporary jobs increased a person's chance of finding permanent employment, among low-skilled U.S. workers who participated in Detroit's welfare-to-work program, confirmed these results (Autor and Houseman, 2005). Autor and Houseman found that temporary work did not increase and could actually decrease the chance of finding permanent employment for low-skilled workers. However, job placements with direct-hire employers[2] significantly increased employment and earnings over a seven-quarter follow-up period. Our research further explores these discrepancies but in the new gig economy context.

4.2.2 Employment Opportunities in the Sharing Economy

A growing body of HCI and CSCW employment research aims to support job seekers by providing platforms to support skill development (Salehi and Bernstein, 2018; Suzuki et al., 2016) and education and learning (Dillahunt et al., 2016b; Hui, Gergle, and Gerber, 2018). The use of and opportunities for technology platforms like the sharing economy to support resource-constrained individuals (Dillahunt, 2014; Dillahunt and Malone, 2015) and entrepreneurs has also been investigated (Hui et al., 2018). Efforts to prevent wage theft among low-wage precarious workers also exist (Dombrowski, Alvarado Garcia, and Despard, 2017).

Participation in non-traditional work, like that of Uber and Lyft, has increased over the last decade (Chen et al., 2017). Such opportunities provide a flexible work environment and opportunities to earn supplemental income and are a viable option for unemployed job seekers or individuals who are resource-constrained (Cook et al., 2018). While some have speculated that such benefits favor women, an investigation of earnings and labor supply choices among more than 1 million U.S. Uber drivers revealed about a 7% gender earnings gap among drivers (Cook et al., 2018). This suggests that the gig economy does not close gender gaps. A Swedish investigation of gender discrimination in hiring found that women had a slightly higher callback rate to interview in female-dominated occupations but no difference in callback rates for male-dominated occupations such as driving (Carlsson, 2011). Decomposing the positive and negative signaling effects of gig work and measuring the net effect by gender has important practical implications because it provides lessons for job applicants who might be unsure whether to disclose gig work on their formal resume.

Past HCI and CSCW work has investigated the impact of new technologies on work practices (Bowers, Button, and Sharrock, 1995; Suchman, 1987). More recently, the field has

[2]In labor economics, "direct-hire employers" refer to the companies that hire their employees directly. Direct-hire jobs are to be contrasted with temporary jobs, which are mainly filled by contract employees.

begun exploring the opportunities for work (Dillahunt, 2014; Dillahunt and Malone, 2015) as well as the working conditions in the sharing economy (Dillahunt et al., 2017; Dombrowski, Alvarado Garcia, and Despard, 2017).

Ahmed et al. (2016) found that Ola [3] drivers were often burdened with locating passengers in a timely manner, rarely reported earnings or reduced hours, and remained unstable in terms of having regular work. While past studies have investigated the working conditions, it's unclear whether such opportunities could lead to more stable employment.

Dillahunt et al. (2016b) asked a similar question but in the context of Massive Open Online Courses (MOOCs), another instance of the sharing economy. In 22 interviews with MOOC learners, these researchers investigated whether MOOCs served as a platform for employability. They found that while some learners found them beneficial, taking these courses did not *land* them a job. As it is unclear whether MOOC engagement actually affects learners' job prospects; it is also unclear whether those who turn to gig work during periods of unemployment are more likely to be hired than those who do not. We raise this question and draw attention to opportunities in this space to explore similar questions and to consider methods of answering such questions quantitatively.

4.3 Research Methods

We conducted an *audit study*, an approach that has been widely used in social sciences and popularized by Bertrand and Mullainathan's seminal investigation of the difference in labor market outcomes between resumes with names that sounded White versus African-American(Bertrand and Mullainathan, 2004). In earlier audit studies, auditors were matched on many observable characteristics except for the variable of interest (gender, sexual orientation, ethnicity) and then initiated face-to-face transactions with businesses (i.e. applied for rental housing, negotiated a price of a car, etc.). These studies were typically criticized based on the quality of their matches. The use of fictitious resumes used by Bertrand and Mullainathan removed the need for human actors by conducting the entire job application process through mail or email. This method is often referred to as *correspondence audit studies* (Bertrand and Duflo, 2016). The researcher creates a balanced set of resumes consisting of *control* resumes and *treatment resumes* that only differ in the variables of interest (in our case gig employment experience and gender). Audit studies have been widely used in the social sciences because of their lower cost; however, these methods have rarely been used in the sharing economy context (Dillahunt et al., 2017). In addition, changes to digitalize the employment process have increased the overall cost of executing these studies. We first provide an overview of our study

[3] Ola is an Uber-like service that exists in India.

before giving the details of our implementation.

4.3.1 Study Focus and Overview

In preparation for our audit study, we conducted a preliminary investigation that consisted of job searches in eight major U.S. metropolitan areas. Based on our preliminary investigation results (Appendix 4.A), we decided to focus our study on one geographic location, one job type, and one job platform, which we address in the next subsection. We then provide details of our resume profiles and our study hypotheses.

4.3.1.1 Addressing study limitations

First, we limited our audit study area to one major metropolitan area that is closest to our institution. As part of the audit study, we needed to compose application profiles that were indistinguishable when compared against real applications from the local area. This required us to control for home address, education records and phone numbers. Choosing a nearby metropolitan area helped us to create more realistic resumes when producing this information: with knowledge about the local neighborhoods, we generated fictitious addresses from lower-income neighborhoods and picked nearby public high schools. Further, we acquired telephone numbers with the local area code. Given our preliminary investigation results, past work in HCI and CSCW, the popularity of employment opportunities as drivers in Uber and Lyft, and because transportation jobs are typically in high demand (Labor Statistics, 2019b), we focused on driving-related job-postings.

We also decided to conduct the experiment solely on Indeed.com. [4] As described in our preliminary investigation, we reviewed job postings available on Monster.com and found them comparable to Indeed.com for low-skilled jobs. This platform facilitates the ability for employers to post their job announcements and for job seekers to apply to these jobs directly by submitting their resumes. We created a set of artificial resumes for low-skilled male and female workers who all had long-term unemployment gaps. We then created two versions of each resume: a baseline resume without gig work experience and an "enhanced" profile with gig work. We applied for jobs using these artificial resumes on the job search platform and tracked callback rates.

[4] http://www.indeed.com; Noted as "the world's # 1 Job Site" per its default landing page found via Google search.

4.3.1.2 Application profiles

We narrowly define gig work in this study as activities that are mediated by technological platforms. We created eight fictitious profiles of low-skilled job applicants with generic names (4 male and 4 female) and associated work histories. To construct credible and naturalistic resumes, we scraped a total of 1,808 publicly available Indeed.com resumes to get typical employer names, job titles, descriptions, employment durations, and associated skills, which we used to create experience entries on the resumes. Our institutional review board reviewed and granted our study exempt status. We independently considered ethical implications associated with creating a very small number of fictitious profiles on a large job search platform and decided on a research protocol to minimize potential harm to employers such as promptly declining requests for further information and interviews.

In our study, we implemented the gig work experience by creating a resume entry that reflected the experience of driving for a major ride-sharing platform. This is a well-perceived form of gig work, and transportation has been highlighted as a key sector of the gig economy (Farrell, Greig, and Hamoudi, 2018). To match with our key treatment, we composed our intend-to-treat group as employers who posted "Driver" as part of their recruitment criteria (i.e., driver-related jobs). We investigated whether employment experience filled with gig work affects callbacks differently from having a gap on the resume (gap vs. gig work) and whether employer responses differ by gender. The main hypotheses that we tested are listed below.

H1: Having gig work to fill the unemployment gap increases the likelihood of an applicant to receive a callback .

H2: Callback rates do not differ by gender.

While prior research found differences in pay related to gender in Uber (Cook et al., 2018), we had not seen prior literature to suggest gender differences in hiring in this specific context. Therefore, we formulated our second hypothesis based on the documentation of no difference in callbacks by gender in Sweden for driving jobs (Carlsson, 2011). Acknowledging the different social norms in gender equality in Sweden versus the U.S., our goal was to help decompose the positive and negative signaling effects of gig work and measure the net effect by gender.

4.3.2 Implementation

To reflect the most up-to-date change in the labor market, we maintained an hourly scraper to continuously download job postings that matched our search criteria: ("Driver jobs," "within

25 miles," "full-time") in our study area. Among all of the scraped job postings, we dropped those postings that simply pointed to external company websites and kept only those that were directly hosted on Indeed. This step dropped roughly one-half of the total postings. Given our knowledge of the local businesses that pointed to external company websites (e.g. size of business, number of postings), we did not identify any systematic differences between the two sets of job postings.

With the combination of the web-scraping tools and a job posting validity checker, we maintained a real-time repository of job-postings: because of the dynamic nature of the work environment, new jobs were posted every day and old jobs would expire. The web-scraping tools accumulated the new jobs and the validity checker visited the job-posting URL afresh daily to verify whether jobs in our raw list were still accepting applications. In total, we had an average of 500 jobs available for application at any given date.

During the job application study period, we randomly sampled from the valid list of job-postings each day. We assigned these jobs to application profiles in the application schedule list and sent two resumes to each job-posting. In total, we created eight application profiles of identical age and comparable education backgrounds on Indeed.com. Each employer received two resumes that were generated based off the two templates, where we added gig work experience to one of the resumes and left the most recent work experience empty for the other resume. Except for resumes that had gig work experience, the most recent work experience ended in March 2018. This generated an 8- to 12-months unemployment gap. To implement the gender treatment, we picked male-sounding or female-sounding first names and added White-sounding surnames based on the Frequently Occurring Surnames from the 2010 Census (Comenetz, 2016). See Appendix 4.B for sample resumes.

To enhance the validity of the application profiles and to collect callbacks, we acquired unique phone numbers with local area codes from Twilio and created dedicated email accounts using Gmail for each profile. All phone calls were first forwarded to voicemail and then transcribed. All incoming emails for each application profile, together with the transcription of the voice recordings, were forwarded to a master email account. We retrieved all responses through the Gmail API and labeled the callbacks for each application profile. Note that callbacks can be of different types. We accounted for all positive responses, including email and voicemail responses that offered phone interviews, email and voicemail responses that arranged for onsite interviews and individualized email responses that asked for specific work experience. We dropped all automated responses that were either email confirmation of the application or a simple pointer to external assessment platforms. We did not systematically differentiate between these different types of callbacks. We classified all "No" responses and "Null" responses as "No Callback."

4.4 Results

We submitted two waves of applications using the same set of applications profiles. We submitted the first batch of 144 applications in November 2018 and started the second wave of applications in February 2019.[5] By March 28, 2019, we submitted an additional 862 applications. Overall, our callback rate was 11.63%, which is relatively high compared to other audit studies (Bertrand and Duflo, 2016). Table 4.1 breaks out the raw callback data by gender and treatment. Table 4.2 shows the corresponding callback rates. Because of the simplicity of our design we conducted our empirical analysis by comparing sample means (callback rates).

Table 4.1: Summary of callbacks by gender and treatment

	Gap		Gig		All
	Female	Male	Female	Male	
Callback	31	35	21	30	117
No Callback	221	216	234	218	889
All	252	251	255	248	1006

Table 4.2: Callback rates by gender and treatment

	Gap		Gig		All
	Female	Male	Female	Male	
Callback rate	12.30%	13.94%	8.24%	12.10%	11.63%

Note: resumes in the "Gap" condition had an unemployment history of 8—12 months because the most recent work experience terminated in March 2018 on these resumes.

We first note that the callback rates for both men and women were lower for the gig-enhanced resumes compared to the baseline resumes with a gap. The overall callback rate for gig-enhanced resumes was 10.13% across all applications (men and women) and 13.12% for baseline resumes. A simple t-test for equality in callback rates rejected the null hypothesis with borderline significance (p-value of 0.14). This was mostly driven by women (p-value of 0.13).

Another way to look at the difference between our main treatments is to consider the 95% confidence interval for the difference in callback rates between gig-enhanced resumes and

[5]While logistical restrictions were the main reason for introducing a 2-month gap between the two waves of applications, the gap also allowed us to avoid the Christmas and holiday season, during which the demand for driver jobs would increase temporarily. Reassuringly, according to the "Economy At A Glance" table provided by the Bureau of Labor Statistics, the unemployment rate in our study area remained stable during the months that we were actively sending applications.

baseline resumes, which is $(-6.95\%, 0.98\%)$. This implies that *at best* the gig-enhanced resumes raised callback rates by about 1% from a baseline rate of 13.52% for the "Gap" condition. This maximal effect size was less than 10% of the baseline rate and small compared to audit studies that looked at discrimination in labor markets (Bertrand and Duflo, 2016).

Looking at gender, we did not see a statistically significant difference between callback rates for male and female applicants without gig-enhanced resumes (p-value of 0.58). We did reject the null hypothesis, with borderline significance equality in callback rates for gig-enhanced resumes (p-value of 0.15). In other words, the callback rates between men and women were unequal.

4.5 Discussion

To summarize, we did not find that gig-enhanced resumes—to mitigate the negative effects of long-term unemployment for low-skilled job seekers—increase callback rates in economically significant ways. This is true even if we take the upper bound of our confidence interval in the difference between callback rates for enhanced and baseline resumes. To the contrary, we found some evidence that callback rates are *lower* for gig-enhanced resumes, especially for women. One possible explanation for observing lower callback rates for gig-enhanced resumes is that gig work not only signals worker quality to the employer (which should raise callback rates) but also signals that the worker has the outside option of taking up gig work if she is unhappy with her new job. The very flexibility of gig work might therefore bolster one's resume and provide a low-friction "escape" from a regular job. To address this question, one might have to go beyond the audit-study methodology and survey employers directly. Next, we speculate on our results and reflect on the use of audit studies.

4.5.1 Recognizing gender differences

Occupations such as drivers are male-driven—according to Data USA,[6] 83.9% of taxi drivers and chauffeurs are male (Bielby and Baron, 1986). Our results showed that male profiles had a higher callback rate than female profiles, regardless of whether the resumes were "enhanced" with gig experience. Employers might apply gender stereotypes and attribute a greater demand for flexibility to female applicants with gig experience than male applicants (for example, they might assume that a female applicant needs the flexibility of gig work to take care of her children). Similarly, employers might infer that female applicants with gig experience

[6]Data USA provides the most comprehensive visualization of U.S. public data: https://datausa.do/profile/soc/533041/#employment

are less willing to accept driving assignments. Past research found that women drivers appeared to avoid unsafe locations, because of either crime or the likelihood of encountering more intoxicated drivers (Cook et al., 2018). Avoiding unsafe locations could lead to route inefficiencies and increased cost in terms of gas and vehicle wear. It would be an interesting extension to better understand these gender differences. In other words, are women hurt more in other domains by disclosing gig work on their resumes? It would require looking at job categories other than "Drivers" to understand how robust this finding is.

4.5.2 Methodological reflections

In this section, we reflect on the methodological contribution of our audit study and discuss the challenges that we experienced. "The audit study is a specific type of field experiment that permits researchers to examine difficult to detect behaviors, such as racial and gender discrimination, and decision-making in real-world scenarios"(Gaddis, 2018, p. 3). In a labor market setting, we examined employers' interview decisions by presenting our application profiles to them as naturally as possible. We did this by conforming to the industry standard of crafting and submitting applications, and responding to the callbacks in a timely manner. However, today's application process has been fully digitalized and highly integrated, which posed a number of challenges. In our reflection, we highlight some of the specific changes and the resulting impact on our study.

4.5.2.1 Digitalization of employment and its impact on audit studies

Executing large-scale audit studies was relatively easy in the early 2000s when employers still posted job openings in local newspapers and accepted mail applications. No technology platform vetted applications and ensured unique emails and phone numbers, for example. Early studies could therefore create thousands of resumes with multiple treatment arms by simply printing and mailing customized resumes (Bertrand and Mullainathan, 2004).

However, the rise of job listing sites has fundamentally changed the nature of audit studies. These sites require that applicants spend a considerable amount of time creating a profile and completing sometimes lengthy surveys. Now, employers on these sites can search applicants' profiles and send invitations for interviews, which transforms the traditional job-search process. While digitization is likely to decrease *average* combined search cost (the sum of frictional costs that are borne by employers and workers) by making the matching process more efficient, past research suggests that it might do little to help under-served job seekers, who sometimes lack experience in completing these online onboarding processes (Wheeler and Dillahunt, 2018).

4.5.2.2 Challenges in the new digital era

Perhaps ironically, job sites make automated creation of profiles and applications for audit studies far more labor-intensive than previously. There are two new types of barriers not present in early audit studies: (1) creating a profile requires great care to not be screened out as a spammer; and (2) some employers demand additional interactions in addition to the standard application form.

First, consider profile creation. In our setting, a unique email address and telephone number identifies an application profile. We needed to prepare and register eight sets of these identifiers on Indeed to conduct our audit study. We customized the home address, telephone number, and education records to the local area, which would need to be replicated for any new metropolitan areas.

Because job applicants typically use free consumer email providers, we created personal Gmail accounts for our fictitious applicants. At the time of our Gmail account setups (October 2018), Google required a *valid* phone number to receive a verification code during the account setup process. The same phone number could be used for at most two accounts, which required us to purchase four prepaid AT&T SIM cards. We then rented local phone numbers on Twilio ($1 per month) in the study area and set up a call-forwarding and transcription service to Google Voice. We could not use Google Voice directly because no local area codes were available in our study area. To finish the account setup, we needed to populate all required fields in Indeed, which took more than an hour for each application profile.

Although we used a program to specify which profiles applied to which openings, completing the application was difficult, if not impossible to automate. The steps required to complete an application differ by employer and could take anywhere from 20 seconds to 10 minutes. In some cases, employers asked for full employment history, which needed to be carefully copied from the resume and pasted to the HTML form on Indeed.com. However, other questions and tasks were often required such as leadership aptitude surveys and work-preference-related questions such as preferred shifts and earliest start date. Employers also asked about specific work experience such as whether an applicant had experience driving a forklift or working as a manual laborer. Such heterogeneity in the application process posed automation challenges.

Our experiment design required us to commit to finish all the jobs that were assigned with an application profile. In total, we logged a minimum of 60 hours to manually send out all of the applications for one metropolitan city, a minimum of 1 hour for each valid digital profile, and 10 hours to create a basic research infrastructure for a new metropolitan area.[7] Moreover,

[7] List of infrastructures include: web-scrapers for jobs and resumes, email forwarding from personal Gmail accounts to a master email account, call-forwarding from the local phone number to the voicemail, and an auto

to submit 1,000 applications in a new city, a minimum of 60-person hours will be needed at a rate of 16 applications an hour. Lower average search costs could move more of the hiring process to these sites over time compared to alternative channels such as word-of-mouth and social networks, which were traditionally used to fill more than half of all job openings Granovetter, 1995.

Our work also identified key execution steps for automating and replicating these new types of audit studies across geographic regions. Studying digital job searching requires innovative HCI solutions and opens several future work opportunities.

4.5.2.3 Opportunities for audit studies in CSCW and HCI

There has been a growing body of HCI and CSCW research around employment and the sharing-economy (Dillahunt et al., 2017). To strengthen this literature, we reflect on how audit studies could lead to new contributions to the field. We envision future work focused on measuring the extent of labor market discrimination by creating fictitious profiles that vary in numerous aspects of an applicant's identity. For example, an applicant's gender or ethnicity can be easily primed by using appropriate names. Political attitudes or sexual orientation can be indirectly inferred through entries in "hobbies," "volunteer experience," or "work experience," and homeless status can be revealed via addresses, or the lack thereof (Hendry, Woelfer, and Duong, 2017; Hendry et al., 2017). The audit study method can be fruitfully applied to other sharing-economy applications where individual profile characteristics might indirectly affect subsequent outcomes. For example, Sariisik (2018) used the audit study method to document discrimination against users with Arab/Muslim-sounding names on Airbnb. Because user profiles on some platforms have a rich set of identity-based features (such as profile photos, which can reveal gender, physical attractiveness, ethnicity or social class), audit studies can help researchers identify the impact of those characteristics on job search or product market outcomes. Informed by such research, online platforms have the potential to overcome documented discrimination by changing their platform design. Platform users, such as job seekers, can be coached on how to minimize possible bias by curating the information they indirectly reveal through their profiles.

4.6 Limitations

We acknowledge the methodological limitations of our study. Because we only submitted resumes to one U.S. city, we limited generalizability across the U.S. because of the strong

script to refresh the application schedule list daily .

cultural and economic variances. Transportation is the largest gig-economy sector in the country, and transportation jobs, in particular, are in high demand (Labor Statistics, 2019b). Nevertheless, it would be interesting to consider other industries, especially those employing more highly skilled workers such as programmers. In addition, drivers tend to be dominated by men (Cook et al., 2018). The addition of more gender-balanced gig work (e.g., freelance in Upwork, services offered in TaskRabbit) or more women-dominated fields (e.g., babysitting in Sittercity.com[8]) could help to strengthen our results, especially as they relate to understanding gender-based differences in gig work. For example, would experience with Sittercity.com help with applications to traditional employment such as childcare centers or nursing homes?

In the job search context, audit studies showcase *which* resumes employers respond to—not *why* they respond to those resumes. Understanding *why* women seem to benefit the least (and are potentially hurt) by disclosing gig work on their resumes is an area worth exploring further. This signals another limitation of our study: a discussion with employers who responded to our study would add a rich and complementary perspective to our results; however, obtaining honest insights on potentially discriminatory behaviors could be challenging. Further investigations are needed to understand how to streamline audit studies in this context and how to effectively complement this method with a qualitative approach. Finally, the U.S. unemployment rate was historically very low at the time of the study, even though we selected a U.S. city with a higher unemployment rate than the national average. This could dampen the positive signaling effect of gig work because employers could compete even for lower-ability workers and amplify the negative signaling effect of workers having better outside options.

4.7 Conclusion

Our work contributes to the growing gig-economy literature, especially to the few quantitative studies that exist (Dillahunt et al., 2017). Specifically, audit studies such as ours provide a clean methodology to identify the effect of resume content on initial employer interest. However, as mentioned as a limitation, audit studies in the digital employment domain are becoming more challenging to implement as advances in artificial intelligence make the initial application process more interactive. Further investigations within HCI and CSCW are needed to understand how to ease the burden of executing these studies.

Our early results suggest that driving for a real-time ridesharing service does not substitute for traditional driving jobs in bridging employment gaps. This has important implications for both job seekers and policy makers. Job applicants might not want to emphasize such gig work

[8] `Sittercity.com` is an online marketplace that supports families, individuals and corporate employees who wish to hire local in-home care such as babysitting, senior care, pet care and even housekeeping.

on their resumes if the net signaling effect is negative. Policy makers should be encouraged to gain a better understanding of the direct *and* indirect costs and benefits of working for certain gig work platforms. Finally, if we are to support job seekers in making informed decisions about what to include or not include on their resumes, we must understand, going forward, *why* driving for a gig platform had the impact it did.

Bibliography

Ahmed, Syed Ishtiaque et al. (2016). "Peer-to-peer in the Workplace: A View from the Road". In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. CHI '16. San Jose, California, USA: ACM, pp. 5063–5075. DOI: 10.1145/2858036.2858393.

Ai, Chunrong and Edward C Norton (2003). "Interaction terms in logit and probit models". In: *Economics letters* 80.1, pp. 123–129.

Akerlof, George A. and Rachel E. Kranton (2000). "Economics and Identity". In: *The Quarterly Journal of Economics* 115.3, pp. 715–753. DOI: 10.1162/003355300554881.

Andreoni, James and B Douglas Bernheim (2009). "Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects". In: *Econometrica* 77.5, pp. 1607–1636.

Ariely, Dan et al. (2009). "Doing good or doing well? Image motivation and monetary incentives in behaving prosocially". In: *The American Economic Review* 99.1, pp. 544–555.

Autor, David and Susan Houseman (2005). *Do Temporary Help Jobs Improve Labor Market Outcomes for Low-skilled Workers? Evidence from Random Assignments*. National Bureau of Economic Research.

Babcock, Linda et al. (2017). "Gender Differences in Accepting and Receiving Requests for Tasks with Low Promotability". In: *American Economic Review* 107.3, pp. 714–47. DOI: 10.1257/aer.20141734.

Bahler, Kristen (2017). "Unemployment Is Really Low. So Why Can't These People Find Jobs?" In: ed. by Money. [Accessed: 2019-08-15].

Bear, Julia B. and Benjamin Collier (Mar. 2016). "Where Are the Women in Wikipedia? Understanding the Different Psychological Experiences of Men and Women in Wikipedia". In: *Sex Roles* 74.5, pp. 254–265. DOI: 10.1007/s11199-015-0573-y.

Bénabou, Roland and Jean Tirole (2006). "Incentives and Prosocial Behavior". In: *American Economic Review* 96.5, pp. 1652–1678. DOI: 10.1257/aer.96.5.1652.

Benz, Matthias and Stephan Meier (Feb. 2008). "Do people behave in experiments as in the field?—evidence from donations". In: *Experimental Economics* 11.3, pp. 268–281. DOI: 10.1007/s10683-007-9192-y.

Bergstrom, Theodore et al. (1986). "On the private provision of public goods". In: *Journal of Public Economics* 29.1, pp. 25–49.

Bernanke, Ben S. (2004). "Editorial Statement". In: *The American Economic Review* 94.1, pp. 404–404.

Bertrand, Marianne and Esther Duflo (2016). *Field Experiments on Discrimination*. Working Paper 22014. National Bureau of Economic Research. DOI: 10.3386/w22014.

Bertrand, Marianne and Sendhil Mullainathan (2004). "Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination". In: *American Economic Review* 94.4, pp. 991–1013.

Bielby, William T and James N Baron (1986). "Men and women at work: Sex segregation and statistical discrimination". In: *American Journal of Sociology* 91.4, pp. 759–799.

Bishop, Bradley Wade et al. (2019). "Scientists' Data Discovery and Reuse Behavior: (Meta)Data Fitness for Use and the FAIR Data Principles". In: *Proceedings of the Association for Information Science and Technology* 56.1, pp. 21–31. DOI: 10.1002/pra2.4.

Blondel, Vincent D. et al. (Oct. 9, 2008). "Fast Unfolding of Communities in Large Networks". In: *Journal of Statistical Mechanics: Theory and Experiment* 2008.10, P10008. DOI: 10.1088/1742-5468/2008/10/P10008. arXiv: 0803.0476.

Borgman, Christine L. et al. (Feb. 2018). "Digital Data Archives as Knowledge Infrastructures: Mediating Data Sharing and Reuse". In.

Bowers, John et al. (1995). "Workflow from within and without: Technology and cooperative work on the print industry shopfloor". In: *Proceedings of the Fourth European Conference on Computer-supported Cooperative Work ECSCW'95*. Springer, pp. 51–66.

Buser, Thomas et al. (May 2014). "Gender, Competitiveness, and Career Choices ∗". In: *The Quarterly Journal of Economics* 129.3, pp. 1409–1447. DOI: 10.1093/qje/qju009.

Carlsson, Magnus (2011). "Does hiring discrimination cause gender segregation in the Swedish labor market?" In: *Feminist Economics* 17.3, pp. 71–102.

Chapman, Adriane et al. (Jan. 2020). "Dataset Search: A Survey". In: *The VLDB Journal* 29.1, pp. 251–272. DOI: 10.1007/s00778-019-00564-x.

Chen, Daniel L et al. (2016a). "oTree—An open-source platform for laboratory, online, and field experiments". In: *Journal of Behavioral and Experimental Finance* 9, pp. 88–97.

Chen, Daniel L. et al. (Mar. 2016b). "oTree—An Open-Source Platform for Laboratory, Online, and Field Experiments". In: *Journal of Behavioral and Experimental Finance* 9, pp. 88–97. DOI: 10.1016/j.jbef.2015.12.001.

Chen, M. Keith et al. (2017). *The value of flexible work: Evidence from Uber drivers*. Tech. rep. National Bureau of Economic Research.

Chen, Roy and Yan Chen (2011). "The Potential of Social Identity for Equilibrium Selection". In: *The American Economic Review* 101.6, pp. 2562–2589.

Chen, Yan (2008). "Incentive-Compatible Mechanisms for Pure Public Goods: A Survey of Experimental Research". In: *The Handbook of Experimental Economics Results*. Ed. by Charles Plott and Vernon Smith. Vol. 1. Amsterdam: North-Holland, pp. 625 –643.

Chen, Yan et al. (Apr. 2020a). *Motivating Experts to Contribute to Digital Public Goods: A Personalized Field Experiment on Wikipedia*. SSRN Scholarly Paper ID 3588132. Rochester, NY: Social Science Research Network. DOI: 10.2139/ssrn.3588132.

Chen, Yan et al. (2020b). "Motivating Metadata Contributions for Data Re-Use and Reproducibility". In: *AEA RCT Registry*. DOI: 10.1257/rct.6159-1.0.

— (2020c). *Motivating Metadata Contributions for Data Re-Use and Reproducibility*. DOI: 10.1257/rct.6159-1.0.

Christensen, Garret et al. (2019). *Transparent and Reproducible Social Science Research: How to Do Open Science*. University of California Press. DOI: doi:10.1525/9780520969230.

Coffman, Katherine Baldiga (Nov. 1, 2014). "Evidence on Self-Stereotyping and the Contribution of Ideas". In: *The Quarterly Journal of Economics* 129.4, pp. 1625–1660. DOI: 10.1093/qje/qju023.

Coffman, Lucas C. and Muriel Niederle (Sept. 2015). "Pre-Analysis Plans Have Limited Upside, Especially Where Replications Are Feasible". In: *Journal of Economic Perspectives* 29.3, pp. 81–98. DOI: 10.1257/jep.29.3.81.

Comenetz, Joshua (2016). "Frequently occurring surnames in the 2010 Census". In: *United States Census Bureau*.

Cook, Cody et al. (2018). *The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers*. Tech. rep. National Bureau of Economic Research.

Cosley, Dan et al. (2007). "SuggestBot: using intelligent task routing to help people find work in wikipedia". In: *Proceedings of the 12th international conference on Intelligent user interfaces*. Downloaded on February 23, 2003 at http://www.communitytechnology.org/nsf_ci_report/, pp. 32 –41.

Cox, D. R. (1958). *Planning of Experiments*. A Wiley Publication in Applied Statistics. John Wiley & Sons.

Daniels, Morgan et al. (2012). "Managing Fixity and Fluidity in Data Repositories". In: *Proceedings of the 2012 iConference on - iConference '12*. Toronto, Ontario, Canada: ACM Press, pp. 279–286. DOI: 10.1145/2132176.2132212.

Dillahunt, Tawanna R. (2014). "Fostering social capital in economically distressed communities". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, pp. 531–540.

Dillahunt, Tawanna R. and Amelia R. Malone (2015). "The Promise of the Sharing Economy among Disadvantaged Communities". In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. CHI '15. Seoul, Republic of Korea: ACM, pp. 2285–2294. DOI: 10.1145/2702123.2702189.

Dillahunt, Tawanna R. et al. (2016a). "Designing for Disadvantaged Job Seekers: Insights from Early Investigations". In: *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*. DIS '16. Brisbane, QLD, Australia: ACM, pp. 905–910. DOI: 10.1145/2901790.2901865.

Dillahunt, Tawanna R. et al. (2016b). "Do Massive Open Online Course Platforms Support Employability?" In: *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. CSCW '16. San Francisco, California, USA: ACM, pp. 233–244. DOI: 10.1145/2818048.2819924.

Dillahunt, Tawanna R. et al. (Dec. 2017). "The Sharing Economy in Computing: A Systematic Literature Review". In: *Proceedings of the ACM on Human-Computer Interaction* 1.CSCW, 38:1–38:26. DOI: 10.1145/3134673.

Dillahunt, Tawanna R. et al. (2018). "Designing Future Employment Applications for Underserved Job Seekers: A Speed Dating Study". In: *Proceedings of the 2018 Designing Interactive Systems Conference*. ACM, pp. 33–44.

Dokko, Jane et al. (2015). "Workers and the Online Gig Economy. A Hamilton Project Framing Paper". In: *The Hamilton Project: Advancing Opportunity, Prosperity, and Growth*.

Dombrowski, Lynn et al. (2017). "Low-Wage Precarious Workers' Sociotechnical Practices Working Towards Addressing Wage Theft". In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. CHI '17. Denver, Colorado, USA: ACM, pp. 4585–4598. DOI: 10.1145/3025453.3025633.

Eckel, Catherine C. and Philip J. Grossman (2005). "Managing Diversity by Creating Team Identity". In: *Journal of Economic Behavior & Organization* 58.3, pp. 371–392.

Farrell, Diana et al. (2018). "The Online Platform Economy in 2018: Drivers, Workers, Sellers, and Lessors". In: *JPMorgan Chase Institute*.

Fischbacher, Urs and Simon Gachter (Mar. 2010). “Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Goods Experiments”. In: *American Economic Review* 100.1, pp. 541–556. DOI: 10.1257/aer.100.1.541.

Ford, Heather and Judy Wajcman (Aug. 1, 2017). “‘Anyone Can Edit’, Not Everyone Does: Wikipedia and the Gender Gap”. In: *Social Studies of Science*.

Gaddis, S. Michael (2018). *Audit studies: Behind the scenes with theory, method, and nuance*. Vol. 14. Springer.

Gebel, Michael (2013). “Is a temporary job better than unemployment? A cross-country comparison based on British, German, and Swiss panel data”. In: *SOEP paper* 543.

Glott, Ruediger et al. (2010). *Wikipedia Survey–Overview of Results*. Tech. rep.

Granovetter, Mark S. (1995). *Getting a job: A study of contacts and careers*. Chicago: University of Chicago Press.

Greenberg, Jane et al. (Oct. 24, 2001). “Author-Generated Dublin Core Metadata for Web Resources: A Baseline Study in an Organization”. In: *International Conference on Dublin Core and Metadata Applications* 0.0 (0), pp. 38–45.

Gregory, Kathleen (July 13, 2020). “A Dataset Describing Data Discovery and Reuse Practices in Research”. In: *Scientific Data* 7.1 (1), p. 232. DOI: 10.1038/s41597-020-0569-5.

Groves, Theodore and John O. Ledyard (1987). “Incentive Compatibility since 1972”. In: *Information, Incentives and Economic Mechanisms: Essays in Honor of Leonid Hurwicz*. Ed. by Theodore Groves et al. Minneapolis: University of Minnesota Press, pp. 48–111.

Hargittai, Eszter and Aaron Shaw (Apr. 3, 2015). “Mind the Skills Gap: The Role of Internet Know-How and Gender in Differentiated Contributions to Wikipedia”. In: *Information, Communication & Society* 18.4, pp. 424–442. DOI: 10.1080/1369118X.2014.957711.

Heerwegh, Dirk and Geert Loosveldt (2006). “An Experimental Study on the Effects of Personalization, Survey Length Statements, Progress Indicators, and Survey Sponsor Logos in Web Surveys”. In: *Journal of Official Statistics* 22.2, p. 191.

Hemphill, Libby et al. (Mar. 2022). “How Do Properties of Data, Their Curation, and Their Funding Relate to Reuse?” In: *Journal of the Association for Information Science and Technology*, asi.24646. DOI: 10.1002/asi.24646.

Hendry, David G. et al. (2017). “Homeless Young People, Jobs, and a Future Vision: Community Members’ Perceptions of the Job Co-op”. In: *Proceedings of the 8th International Conference on Communities and Technologies*. C&T ’17. Troyes, France: ACM, pp. 22–31. DOI: 10.1145/3083671.3083680.

Hendry, David G. et al. (2017). “U-District Job Co-op: constructing a future vision for homeless young people and employment”. In: *Information Technology & People* 30, pp. 602–628.

Holt, Charles A. and Susan K. Laury (2002). “Risk Aversion and Incentive Effects”. English. In: *The American Economic Review* 92.5, pp. 1644–1655.

Hui, Julie et al. (Nov. 2018). “Making a Living My Way: Necessity-driven Entrepreneurship in Resource-Constrained Communities”. In: *Proceedings of the ACM on Human-Computer Interaction* 2.CSCW, 71:1–71:24. DOI: 10.1145/3274340.

Hui, Julie S. et al. (2018). “IntroAssist: A Tool to Support Writing Introductory Help Requests”. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, p. 22.

Ichino, Andrea et al. (2005). “Temporary work agencies in Italy: A springboard toward permanent employment?” In: *Giornale degli economisti e annali di economia*, pp. 1–27.

Johansson, Louise (2013). “A Study of the Motivation Behind Collaborative Knowledge Production and the Formation of Community in Web 2.0, Using the Case Study of wikiHow.Com”. MA thesis.

Jørgensen, Bent (1987). “Exponential Dispersion Models”. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 49.2, pp. 127–145. DOI: 10.1111/j.2517-6161.1987.tb01685.x.

Juanchich, Marie et al. (May 2015). “Cognitive Reflection Predicts Real-Life Decision Outcomes, but Not Over and Above Personality and Decision-Making Styles”. In: *Journal of Behavioral Decision Making* 29.1, pp. 52–59. DOI: 10.1002/bdm.1875.

Karlan, Dean (2005). “Using Experimental Economics to Measure Social Capital and Predict Financial Decision”. In: *American Economic Review* 95.5, pp. 1688–1699.

Krueger, Alan B. et al. (2014). “Are the long-term unemployed on the margins of the labor market?” In: *Brookings Papers on Economic Activity* 2014.1, pp. 229–299.

Labor Statistics, Bureau of (2019a). “Economic News Release: Employment Situation Summary”. In: [Accessed: 2019-04-03].

— (2019b). “Occupational Employment and Wages Summary”. In: [Accessed: 2019-04-03].

Lahno, Amrei M. et al. (Oct. 2015). “Conflicting risk attitudes”. In: *Journal of Economic Behavior & Organization* 118, pp. 136–149. DOI: 10.1016/j.jebo.2015.03.003.

Laury, Susan and Laura Taylor (2008). “Altruism spillovers: Are behaviors in context-free experiments predictive of altruism toward a naturally occurring public good”. In: *Journal of Economic Behavior Organization* 65.1, pp. 9–29.

Ledyard, John (1995). “Public goods: A survey of experimental research”. In: *The Handbook of Experimental Economics*. Ed. by John H. Kagel and Alvin E. Roth. Vol. 1. Princeton, New Jersey: Princeton University Press.

Leider, Stephen et al. (2009). “Directed Altruism and Enforced Reciprocity in Social Networks”. In: *The Quarterly Journal of Economics* 124.4, pp. 1815–1851.

Li, Linfeng and Yan Chen (2020). “Gender Inequality in Contributions to Wikipedia”. In: *AEA RCT Registry*. DOI: 10.1257/rct.6753.

Mavletova, Aigul and Mick P. Couper (2015). “A Meta-Analysis of Breakoff Rates in Mobile Web Surveys”. In: *Mobile research methods: Opportunities and challenges of mobile research methodologies*, pp. 81–98.

Mesgari, Mostafa et al. (2015). ““The Sum of All Human Knowledge”: A Systematic Review of Scholarly Research on the Content of Wikipedia”. In: *Journal of the Association for Information Science and Technology* 66.2, pp. 219–245. DOI: 10.1002/asi.23172.

Mittereder, Felicitas and Brady T West (June 2021). “A DYNAMIC SURVIVAL MODELING APPROACH TO THE PREDICTION OF WEB SURVEY BREAKOFF”. In: *Journal of Survey Statistics and Methodology*, smab015. DOI: 10.1093/jssam/smab015.

Murawski, Carsten and Peter Bossaerts (Oct. 2016). “How Humans Solve Complex Problems: The Case of the Knapsack Problem”. In: *Scientific Reports* 6.1. DOI: 10.1038/srep34851.

Newman, Mark (Sept. 2017). *Networks: An Introduction*. 2 edition. Oxford ; New York: Oxford University Press.

Newman, Mark EJ and Michelle Girvan (2004). “Finding and Evaluating Community Structure in Networks”. In: *Physical review E* 69.2, p. 026113.

Niederle, M. and L. Vesterlund (Aug. 2007). “Do Women Shy Away From Competition? Do Men Compete Too Much?” In: *The Quarterly Journal of Economics* 122.3, pp. 1067–1101. DOI: 10.1162/qjec.122.3.1067.

Peytchev, Andy (2009). “Survey Breakoff”. In: *Public Opinion Quarterly* 73.1, pp. 74–97.

— (2011). “Breakoff and Unit Nonresponse across Web Surveys”. In: *Journal of Official Statistics* 27.1, p. 33.

Pienta, Amy M. et al. (Nov. 2010). “The Enduring Value of Social Science Research: The Use and Reuse of Primary Research Data”. In.

Piwowar, Heather A. et al. (Mar. 21, 2007). “Sharing Detailed Research Data Is Associated with Increased Citation Rate”. In: *PLoS ONE* 2.3. Ed. by John Ioannidis, e308. DOI: 10.1371/journal.pone.0000308.

Pons, Pascal and Matthieu Latapy (2006). “Computing Communities in Large Networks Using Random Walks”. In: p. 28.

Rege, Mari and Kjetil Telle (2004). “The Impact of Social Approval and Framing on Cooperation in Public Good Situations”. In: *Journal of Public Economics* 88.7, pp. 1625–1644.

Reuben, Ernesto et al. (Nov. 2015). *Taste for Competition and the Gender Gap Among Young Business Professionals*. Tech. rep. DOI: 10.3386/w21695.

Saccardo, Silvia et al. (Apr. 2018). “On the Size of the Gender Difference in Competitiveness”. In: *Management Science* 64.4, pp. 1541–1554. DOI: 10.1287/mnsc.2016.2673.

Salehi, Niloufar and Michael S. Bernstein (2018). “Ink: Increasing Worker Agency to Reduce Friction in Hiring Crowd Workers”. In: *ACM Transactions on Computer-Human Interaction (TOCHI)* 25.2, p. 10.

Samuelson, Paul A. (1954). “The Pure Theory of Public Expenditure”. In: *Review of Economics and Statistics* 36.4, pp. 387–389.

Santos, Luiz et al. (Sept. 2016). “FAIR Data Points Supporting Big Data Interoperability”. In: p. 10.

Sariisik, Merve (2018). *Identity Discrimination in the Sharing Economy: A Field Experiment*. University of Michigan Working Paper.

Schnusenberg, Oliver and Andrés Gallo (2011). “On Cognitive Ability and Learning in a Beauty Contest”. In: *Journal for Economic Educators* 11.1, pp. 13–24.

Shaw, Aaron and Eszter Hargittai (Feb. 1, 2018). “The Pipeline of Online Participation Inequalities: The Case of Wikipedia Editing”. In: *Journal of Communication* 68.1, pp. 143–168. DOI: 10.1093/joc/jqx003.

Steinbrecher, Markus et al. (June 2015). “Why Do Respondents Break Off Web Surveys and Does It Matter? Results From Four Follow-up Surveys”. In: *International Journal of Public Opinion Research* 27.2, pp. 289–302. DOI: 10.1093/ijpor/edu025.

Suchman, Lucy A. (1987). *Plans and situated actions: The problem of human-machine communication*. Cambridge University Press.

Suzuki, Ryo et al. (2016). “Atelier: Repurposing expert crowdsourcing tasks as micro-internships”. In: *Proceedings of the 2016 CHI conference on human factors in computing systems*. ACM, pp. 2645–2656.

Thomer, Andrea K. et al. (Feb. 2022). “The Craft and Coordination of Data Curation: Complicating ”Workflow” Views of Data Science [PREPRINT]”. In: DOI: 10.7302/4017.

Traag, V. A. et al. (Mar. 26, 2019). “From Louvain to Leiden: Guaranteeing Well-Connected Communities”. In: *Scientific Reports* 9.1 (1), p. 5233. DOI: 10.1038/s41598-019-41695-z.

Van Belle, Eva et al. (2017). “Why is unemployment duration a sorting criterion in hiring?” In: *IZA Discussion Paper No. 10876. Available at SSRN:* *https://ssrn.com/abstract=2998986*.

Vesterlund, Lise (2015). “Using experimental methods to understand why and how we give to charity”. In: *The Handbook of Experimental Economics*. Ed. by John H. Kagel and Alvin E. Roth. Vol. 2. Princeton, New Jersey: Princeton University Press.

Vilhuber, Lars (May 2019). “Report by the AEA Data Editor”. In: *AEA Papers and Proceedings* 109, pp. 718–29. DOI: 10.1257/pandp.109.718.

Voors, Maarten J et al. (Apr. 2012). “Violent Conflict and Behavior: A Field Experiment in Burundi”. In: *American Economic Review* 102.2, pp. 941–964. DOI: 10.1257/aer.102.2.941.

Wakita, Ken and Toshiyuki Tsurumi (May 8, 2007). “Finding Community Structure in Mega-Scale Social Networks: [Extended Abstract]”. In: *Proceedings of the 16th International Conference on World Wide Web*. WWW ’07. Banff, Alberta, Canada: Association for Computing Machinery, pp. 1275–1276. DOI: 10.1145/1242572.1242805.

Ward, Patrick S. and Vartika Singh (June 2015). “Using Field Experiments to Elicit Risk and Ambiguity Preferences: Behavioural Factors and the Adoption of New Agricultural Technologies in Rural India”. In: *The Journal of Development Studies* 51.6, pp. 707–724. DOI: 10.1080/00220388.2014.989996.

Wheeler, Earnest and Tawanna R. Dillahunt (2018). “Navigating the Job Search as a Low-Resourced Job Seeker”. In: *Proceedings of the 36th Annual ACM Conference on Human Factors in Computing Systems*. CHI ’18. Montreal, QC, Canada: ACM. DOI: 10.1145/3173574.3173622.

Wilkinson, Mark D. et al. (Mar. 2016). “The FAIR Guiding Principles for Scientific Data Management and Stewardship”. In: *Scientific Data* 3.1, p. 160018. DOI: 10.1038/sdata.2016.18.

Yaraghi, Niam and Shamika Ravi (2017). “The current and future state of the sharing economy”. In: *Available at SSRN:* *https://ssrn.com/abstract=3041207*.

Zhang, Xiaoquan Michael and Feng Zhu (2011). “Group Size and Incentives to Contribute: A Natural Experiment at Chinese Wikipedia”. In: *American Economic Review* 101.4, pp. 1601–15.

Appendices

4.A Details of our preliminary investigation

In preparation for our audit study, we conducted preliminary job searches in eight major metropolitan areas in the U.S., including Chicago, Detroit, New York, Miami, Los Angeles, Atlanta, Seattle and the Dallas–Fort Worth. For each metropolitan area, we searched Indeed.com for full-time jobs within a 25-mile radius. Further, to target lower-skilled jobs, we restricted the searches by salary ranges ($20K$–$30K$, $30K$–$40K$, $40K$–$50K$ and $50K$–$60K$). For each set of search results, we categorized the first 200 jobs into sensible bins where each bin had at least 10 jobs (rare jobs were collected into the "Others" bin) and aggregated the categories across salary ranges. The search-and-categorize exercise generated a distribution of jobs for each of the eight metropolitan areas. To enhance the external validity in our job search, we repeated the exercise on Monster.com, which is the head-to-head competitor of Indeed. Based on a set of comparisons of the distributions of jobs, both across the two job sites and across the eight metropolitan areas, we observed that: (1) The types of jobs on Indeed are not systematically different from the type of jobs on Monster. (2) The distributions of jobs within the metropolitan areas are also indistinguishable from one another; and (3) After combining all the jobs that we searched, driving jobs, in particular, made up a stable fraction in all metropolitan areas.

4.B Sample resumes

Note that the city, state, ZIP code, and phone numbers were included in the original study.

Sample Resume, version 1, with gig work.

Firstname Lastname

City, State, XXXXX
xxxxxxxxxxxxxx@gmail.com
(xxx) xxx-xxxx

Authorized to work in the US for any employer

Work Experience

Driver
Gig_Company - City, State
March 2018 to Present

- Drive clients to and from local locations through the Metro area
- Follow all traffic laws and regulations
- Provide excellent customer service

Valet Driver
Hotel XXXXXX - City, State
January 2016 to March 2018

- Greet customers in a pleasant manner and inquire into their car parking needs.
- Drive customers' vehicles to designated parking locations in a safe manner.
- Provide customers with a receipt in exchange of handed-over keys.
- Ascertain that vehicles are properly parked in designated lots and locked before being left alone.
- Use hand signals, batons and lights to direct customers' vehicles in available parking spots.
- Take receipt tags from customers, locate their cars and drive them to the waiting areas.

Parking Attendant
XXXX Parking - City, State
June 2014 to December 2015

- Maintain great customer service
- Accept debit and cash transactions
- Answer any questions needed
- Maintain parking lot area
- Greet and give directions (if needed)

Pizza Delivery Driver
XXX XXXXX's Pizza - City, State
October 2011 to May 2014

- Received and delivered quality products to customers
- Communicated with kitchen staff
- Maintained kitchen work areas, equipment, and utensils in clean and orderly condition Answered telephone calls and responded to inquiries.
- Performed all transactions in a cordial, efficient and professional manner

Sample Resume, version 1, with gig work, continued.

- Took food orders and relayed orders to kitchens or serving counters
- Washed dishes, glassware, flatware, pots, and/or pans using dishwashers or by hand
- Cleaned and sterilized equipment and facilities

Team Member
XXX King - City, State
July 2010 to October 2011

- Works as food cashier
- Meet and greet customers
- Take food and drink orders
- Prepared food and drink orders
- Answer questions about menu items, policies, and services
- Provide excellent customer care
- Maintain a clean work environment

Education

BBBBB High School - City, State
August 2007 to May 2010

Sample Resume, version 2, without gig work

Firstname Lastname

City, State XXXXX
xxxxxxxxxxxxxxxxx@gmail.com
(xxx) xxx-xxxx

Authorized to work in the US for any employer

Work Experience

Driver
XXX Landscape Supply - City, State
May 2016 to March 2018

- Deliver landscape supplies and consumer goods to customers
- Process customer orders
- Load & unload products

Hand Lawn Mower and Leaf Blower
XXXXXXXX Services, INC - City, State
August 2015 to April 2016

- Completed range of landscaping duties including raking, shoveling
- Operating weed-whacking equipment, hand lawn mower, leaf blower, riding mower

Truck Driver
XXXXXX, LLC - City, State
March 2012 to August 2015

- Pre and post vehicle inspections. Maintain all vehicle fluid levels.
- Recommend vehicle service when required. Update daily logs and documents.
- Pick up and deliver truckload automotive parts/products between various suppliers.

Warehouse Associates
XXXX - City, State
September 2010 to February 2012

- Perform customer service, meet and greet customers, organize store products, meet customer needs with desired products, and housekeeping duties as well
- Unload trucks for inventory

Education

AAAAAAA High School - City, State
August 2007 to May 2010